2015年度　計量生物セミナー
臨床試験における estimand と感度分析

欠測のあるデータに対する解析手法の基礎 ～ (2) 主解析の検討～

製薬協

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TF4
欠測のあるデータの解析チーム(JPMA 欠測チーム)

塩野義製薬株式会社 藤原 正和
田辺三菱製薬株式会社 高橋 文博

発表構成

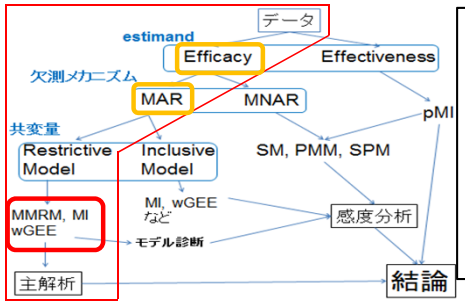
製薬協

1. Analytic Road Mapにおける主解析の候補
2. Mixed Models for Repeated Measures (MMRM)
3. Multiple Imputation（MI）
   – MARを仮定した (単調回帰の) MI
4. 最後に

Analytic Road Map (Mallinckrodt, 2013)

製薬協

- 欠測のあるデータの解析における主解析の選択



・Estimand：efficacy
・欠測メカニズム：MAR
のとき、
Mallinckrodt(2013)の
主解析の候補は
・**MMRM**
・**MI**
・**wGEE**
の3つである

SASユーザー総会2014, 2015：土居ら(2014)，藤原ら(2015)【JPMAシンポジウム】：横山ら(2015)

主解析の候補

製薬協

- **MMRM (Mixed Models for Repeated Measures)**
  - **尤度の考え方に基づく方法．**
- **MI (Multiple Imputation)**
  - **多重補完を行い，補完後の完全データに対して適切な解析を実施する方法．**
- **wGEE (weighted Generalized Estimating Equation)**
  - **確率の逆数の重みづけによるGEE法．セミパラメトリックな解析方法．**

wGEEについては本セッションでは取り上げない．wGEEの詳細はSASユーザー総会2015の駒嵜ら(2015)参照

記号の定義

製薬協

対象となるデータ：経時データ（連続値）

$$\mathbf{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_i^o \\ \mathbf{Y}_i^m \end{pmatrix} \quad (i = 1, \cdots, N)$$

n ：被験者 i の（計画された）測定時点

N ：被験者数

$\mathbf{Y}_i^o$ ：観測データ

$\mathbf{Y}_i^m$ ：欠測データ

欠測識別変数

$$R_{ij} = \begin{cases} 1 & \text{被験者 } i \text{ の } j \text{ 時点でのデータが観測} \\ 0 & \text{被験者 } i \text{ の } j \text{ 時点でのデータが欠測} \end{cases} \qquad \mathbf{R}_i = \begin{pmatrix} R_{i1} \\ \vdots \\ R_{in} \end{pmatrix}$$

$$D_i = \sum_{j=1}^{n} R_{ij} + 1 \rightarrow$$

単調な欠測の場合，
被験者 i は時点 D_i で脱落，
完了例は $D_i = n+1$ ，

Mixed Models for Repeated Measures (MMRM)

製薬協

尤度を用いた方法

製薬協

- **欠測のあるデータの尤度**
  - **応答変数** $\mathbf{Y}_i = (\mathbf{Y}_i^o, \mathbf{Y}_i^m)'$
  - **欠測識別変数** $\mathbf{R}_i$

  } 両方の尤度の寄与を考えなくてはならない

- **完全データの尤度（Full Data Likelihood）**

$$\tilde{L}(\boldsymbol{\theta}, \boldsymbol{\psi}) \propto \prod_{i=1}^{N} f(\mathbf{Y}_i, \mathbf{R}_i | \boldsymbol{\theta}, \boldsymbol{\psi})$$

  - **完全データの尤度は，欠測データ $\mathbf{Y}_i^m$ も含む．**

- **観測データの尤度（Observed Data Likelihood）**

$$L(\boldsymbol{\theta}, \boldsymbol{\psi}) = \prod_{i=1}^{N} f(\mathbf{Y}_i^o, \mathbf{R}_i | \boldsymbol{\theta}, \boldsymbol{\psi})$$

$$= \prod_{i=1}^{N} \int f(\mathbf{Y}_i^o, \mathbf{Y}_i^m, \mathbf{R}_i | \boldsymbol{\theta}, \boldsymbol{\psi}) d\mathbf{Y}_i^m$$

$\boldsymbol{\theta}$ ：応答変数の分布のパラメータ

$\boldsymbol{\psi}$ ：欠測識別変数の分布のパラメータ

  - **尤度を用いた方法では，観測データの尤度に基づいて推測を行う．**

SM（Selection Model）

製薬協

- **SM（Selection Model）とは**
  - **完全データの尤度が，以下のように分解されることを想定．**

$$f(\mathbf{Y}_i, \mathbf{R}_i | \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{Y}_i | \boldsymbol{\theta}) \cdot f(\mathbf{R}_i | \mathbf{Y}_i, \boldsymbol{\psi})$$

$$= f(\mathbf{Y}_i^o, \mathbf{Y}_i^m | \boldsymbol{\theta}) \cdot f(\mathbf{R}_i | \mathbf{Y}_i^o, \mathbf{Y}_i^m, \boldsymbol{\psi})$$

応答変数の分布のパラメータ

  - **第2項が，「観測された集団」または「欠測した集団」への個人の選択をモデル化していると解釈できるため，“Selection Model”と呼ばれる．**

- **SMにおける観測データの尤度**

$$L(\boldsymbol{\theta}, \boldsymbol{\psi}) = \prod_{i=1}^{N} \int f(\mathbf{Y}_i^o, \mathbf{Y}_i^m | \boldsymbol{\theta}) \cdot f(\mathbf{R}_i | \mathbf{Y}_i^o, \mathbf{Y}_i^m, \boldsymbol{\psi}) d\mathbf{Y}_i^m$$

MMRMの位置づけ

製薬協

- SMの一つの形式
  - MARを仮定したもとで，観測データの尤度に基づいて推測を行うことができる．
- SMの中での位置づけ

$$f(\mathbf{Y}_i, \mathbf{R}_i|\mathbf{X}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{Y}_i|\mathbf{X}_i, \boldsymbol{\theta}) f(\mathbf{R}_i|\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\psi})$$

欠測過程を含まない観測データの尤度

$f(\mathbf{Y}_i^o|\mathbf{X}_i, \boldsymbol{\theta})$

MCAR $= f(\mathbf{R}_i|\mathbf{X}_i, \psi)$

MAR $= f(\mathbf{R}_i|\mathbf{Y}_i^o, \mathbf{X}_i, \psi)$

MNARの場合は，そのまま

最大化により，$\boldsymbol{\theta}$ の一致推定量を得ることができる．

LMM（Linear Mixed Model）を想定する

MMRM

■MMRMという呼び名に関しては，様々な意見があるため，Mallinckrodt (2013) 等を参照

MMRMで特定が必要なもの Mallinckrodt (2013)

製薬協

- 平均構造
  - 共変量の選択，変量効果を組み込むかどうか
- 推定方法
  - 制限付き最尤法（REML）が第一選択
- （周辺モデルの）分散共分散構造
  - 収束しなかった場合の対応（構造の変更順など）
  - 例：UN→Toeplitz⇒HCS⇒AR（1）⇒CS⇒VC
- 自由度の計算方法
  - Kenward-Roger法が一般的

PROC MIXEDを利用して，実装することが可能

Multiple Imputation （MI）

製薬協

Multiple Imputation（MI）

製薬協

- MIとは何か
  - Rubin（1978, 1987）によって提案された方法

欠測値を含むデータに対して，

1. 複数回の補完を行い，
2. 補完後のそれぞれの完全データに対して解析を行い，
3. その結果を1つの最終結果に統合する

方法である．

  - 補完の方法には様々な種類がある．
  - 完全データに対する解析はANCOVAやMMRMが利用されることが多い．
- 本発表では，MARを仮定したMIに注目する．

Multiple Imputation（MI）

製薬協

- **複数回の補完を行うことで，欠測値の補完に対して不確実性を考慮することができる．**
- **MIにおける補完モデルと解析モデルが尤度ベースのモデルと同じならば，MIの結果は尤度ベースの結果と類似する．（Mallinckrodt，2013）**
  - **補完モデル：欠測値を補完するための統計モデル**
  - **解析モデル：多重補完された完全データを用いて解析するための統計モデル**

補完モデルの3つのアプローチ (Carpenter (Chapter 3), 2012)

製薬協

欠測パターン	補完モデル内の変数	方法
単調のみ	連続値のみ	• Sequential regression imputation (単調回帰と呼ぶ) ➢ 欠測値のある変数に対して，その時点までに得られている観測値から予測する回帰モデルを構築する．欠測値のある被験者に対しては，この回帰モデルの予測値を代入する．
非単調も可	連続値のみ	• Joint modeling approach ➢ 観測値が与えられたもとで，多変量分布から計算される欠測値に対する条件付分布から補完値を生成する．
	カテゴリカル，連続値を含む場合でも可	• Full conditional specification ➢ 欠測値のある多変量データの補完を欠測値のある変数ごとに実施する．各々の欠測値のある変数に対して，補完モデルを構築し，それぞれの変数に対して補完値を繰り返し作成する．

SASのproc miでいずれも指定が可能である

単調回帰を用いるMIの補完手順

製薬協

- 時点 j の欠測値 Y_j を補完するため，観測されたデータ $Y_1, \ldots, Y_{j-1}$ を用いて予測分布を構築

$$Y_j \sim f\left(Y_j | Y_1, \ldots, Y_{j-1}\right)$$

$j = 3$ の場合

補完したいのはここ

	時点1	時点2	時点3	時点4
パターン2	○	○	×	×
パターン3	○	○	○	×
パターン4	○	○	○	○

予測分布

欠測メカニズムとして，MARを仮定している

○：観測
×：欠測

単調回帰を用いるMIの補完手順

製薬協

- 予測分布を構築する為の回帰モデル

$$Y_j = \beta_0 + \beta_1 Y_1 + \ldots + \beta_{j-1} Y_{j-1}$$

- 回帰パラメータの推定値

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_{j-1})'$$

- 推定値の共分散行列 $\hat{\sigma}_j^2 \mathbf{V}_j$
  - ただし，$\mathbf{V}_j$ は $(\mathbf{X}'\mathbf{X})^{-1}$
    $(\mathbf{X}'\mathbf{X})$ は $Y_1, \ldots, Y_{j-1}$ で構成されるデザイン行列

単調回帰を用いるMIの補完手順

製薬協

- パラメータの事後分布から下記をサンプル

$$\boldsymbol{\beta}_* = (\beta_{*0}, \beta_{*1}, \ldots, \beta_{*j-1})', \sigma^2_{*j}$$

分散： $\sigma^2_{*j} = \hat{\sigma}^2_j (n_j - j)/G, G \sim \chi^2_{n_j - j}$

回帰係数： $\boldsymbol{\beta}_* = \hat{\boldsymbol{\beta}} + \sigma_{*j} \mathbf{V}^T_{hj} \mathbf{Z}$

ただし，$\mathbf{V}^T_{hj}$ はコレスキー分解 $\mathbf{V}_j = \mathbf{V}^T_{hj} \mathbf{V}_{hj}$ により得られる上三角行列

n_j ： Y_j が観測されているデータ数

$\mathbf{Z}$ ： j 個の変数で互いに独立の正規変数ベクトル

単調回帰を用いるMIの補完手順

製薬協

- 欠測値 Y_j を下記の式で生成される値で補完

$$\beta_{*0} + \beta_{*1} Y_1 + \ldots + \beta_{*j-1} Y_{j-1} + z_i \sigma_{*j}$$

$z_i \sim N(0,1)$

- ここまでの過程をM回繰り返す

$j = 3$ の場合

	時点1	時点2	時点3	時点4
パターン2	○	○	×	×
パターン3	○	○	○	×
パターン4	○	○	○	○

Multiple Imputationの概略 (Rubin, 1987)

製薬協

- 前スライドまでのステップで，M個の完全データセットが得られる．
- M個の完全データセットに対して，解析をそれぞれ実施する．
- パラメータベクトル (k x 1) θ がそれぞれ得られる．

例）群間差など

- M個の完全データセットより推定されるパラメータベクトル

$$\hat{\boldsymbol{\theta}}^* = \frac{1}{M} \sum_{m=1}^{M} \hat{\boldsymbol{\theta}}^{(m)} \quad \left(\hat{\theta}^{(m)} - \theta\right) \dot{\approx} N\left(0, U^{(m)}\right)$$

Multiple Imputationの概略 (Rubin, 1987)

製薬協

$$\hat{\boldsymbol{\theta}}^* = \frac{1}{M} \sum_{m=1}^{M} \hat{\boldsymbol{\theta}}^{(m)}$$

- 推測は下記にもとづく

$$\left(\hat{\theta}^* - \theta\right) \sim N(0, V)$$

$$V = \widehat{W} + \left(\frac{M+1}{M}\right) \hat{\mathrm{B}}$$

$$\widehat{\mathrm{W}} = \frac{\sum_{m=1}^{M} \widehat{\mathrm{U}}^{(m)}}{M}$$

補完内分散

$$\hat{\mathrm{B}} = \frac{\sum_{m=1}^{M} \left(\hat{\theta}^{(m)} - \hat{\theta}^*\right)\left(\hat{\theta}^{(m)} - \hat{\theta}^*\right)'}{M-1}$$

補完間分散

Multiple Imputationの概略 (Rubin, 1987)

製薬協

- 漸近正規性により下記の検定統計量を構成

$$\left(\hat{\theta}^{*}-\theta\right)^{T}\mathbf{V}^{-1}\left(\hat{\theta}^{*}-\theta\right)\dot{\approx}\chi_{k}^{2}$$

- Li *et al.* (1991) によるF分布に基づく推測を利用
  $H_0:\theta=\theta_0\ \ H_1:\theta\neq\theta_0$ におけるF統計量とP値

$$F=\frac{\left(\hat{\theta}^{*}-\theta_0\right)^{T}V^{-1}\left(\hat{\theta}^{*}-\theta_0\right)}{k\left(R+1\right)}\dot{\approx}F_{w}^{k}$$

$$w=4+(\tau-4)\left(1+\frac{1-2\tau^{-1}}{R}\right)^{2}$$
$$R=\frac{1}{k}\left(1+\frac{1}{M}\right)tr(\mathbf{B}\mathbf{W}^{-1})$$
$$\tau=k\left(M-1\right)$$
$$\text{P value}=Pr\left(F_{w}^{k}>F\right)$$

- 漸近論の推測は標本数Nや補完回数のMに依存する

補完回数の検討

製薬協

- 推定効率の観点で，3〜5回でも十分としている
  - Rubin(1978, 1987, 1996)
- 最近の文献では，少なくても欠測確率（%）以上
  例えば，20%の欠測データ → 20回以上の補完が必要
  - White *et al.* (2011)

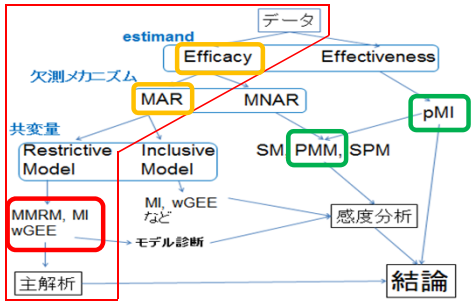
本TFの検討 (大江ら, 2015) では2種類のシミュレーションデータに対して・・・
- 少ない補完回数では，検出力の低下が示唆された
- 補完回数が100回を超えると，検出力が頭打ちしていた
- 補完回数を最大限調整しても，MMRMに比べて，若干劣る検出力であった

補完回数は試験の計画段階で十分に検討することが必要である

最後に

製薬協

- 主解析の方法として，MMRM，MIを紹介した



Estimandとして Efficacy，
欠測メカニズムとして MARを仮定している

感度分析の候補として，PMM，pMIを考える

参考文献

製薬協

- Carpenter, J., & Kenward, M. (2012). ***Multiple imputation and its application.*** John Wiley & Sons.
- Lu, K. & Mehrotra D. V. (2010). Specification of covariance structure in longitudinal data analysis for randomized clinical trials. Statistics in Medicine, 29, 474-488.
- Mallinckrodt, C. H. (2013). Preventing and Treating Missing Data in Longitudinal Clinical Trials. ***Cambridge Press***.
- Molenberghs, G., & Kenward, M. (2007). ***Missing data in clinical studies*** (Vol. 61). John Wiley & Sons.
- National Research Council (2010). ***The Prevention and Treatment of Missing Data in Clinical Trials***. Washington, DC: The National Academies Press.
- Royston, P., & White, I. R. (2011). Multiple imputation by chained equations (MICE): implementation in Stata. ***J Stat Softw***, 45(4), 1-20.
- Rubin, D. B.(1978). Multiple Imputation in sample surveys- Aphenomenological Bayesian approach to nonresponse. Imputation and Editing of Faulty or Missing Survey Data. Washington, DC:***U.S. Department of Commerce***

参考文献

製薬協

- **Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys* (Vol. 81). John Wiley & Sons.**
- **Yuan, Y. (2011). Multiple imputation using SAS software. *Journal of Statistical Software*, *45*(6), 1-25.**
- **White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine*, *30*(4), 377-399.**
- 高橋文博. (2015).【日本製薬工業協会シンポジウム】臨床試験の欠測データの取り扱いに関する最近の展開と今後の課題について ー統計手法・estimandと架空の事例に対する流れの整理ー (3)Pattern-Mixture Modelの解説.
- 土居正明, 大浦智紀, 大江基貴, 駒嵜弘, 髙橋文博, 縄田成毅, 藤原正和, 横溝孝明, 横山雄一. (2014). 欠測のあるデータに対する総合的な感度分析と主解析の選択. SASユーザー総会論文集.
- 土居正明, 鵜飼裕之, 大浦智紀, 大江基貴, 駒嵜弘, 藤原正和, 横山雄一. (2015).欠測のあるデータにおける主解析の検討. SASユーザー総会論文集.

ご清聴ありがとうございました

製薬協