# 【企画セッション】欠測のあるデータに対する各種解析手法と欠測メカニズムに対する感度分析

## (3)解析手法の解説2

高橋文博[1)2)] 藤原正和[1)3)] 大浦智紀[1)4)] 横山雄一[1)5)]
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 欠測のあるデータに対する解析方法論・SASプログラム検討チーム
2) 田辺三菱株式会社 3) 塩野義製薬株式会社
4) 日本イーライリリー株式会社 5) 持田製薬株式会社

# The review of analytical approach 2 (MI, PMM, SPM)

Fumihiro Takahashi[1)2)], Masakazu Fujiwara[1)3)],

Tomonori Oura[1)4)] , Yuichi Yokoyama[1)5)],
1) The team for statistical methodologies and SAS programming of data analysis with missing data, task force 4, data science expert committee, drug evaluation committee, Japan Pharmaceutical Manufacturers Association.

2) Mitsubishi Tanabe Pharma Co., Ltd., 3) Shionogi & Co., Ltd. ,

4) Eli Lilly Japan K. K., 5) Mochida Pharmaceutical Co., Ltd.,

# 要旨:

多重補完法（Multiple Imputation），制約条件を用いたPattern Mixture model，及びShared Parameter Modelの理論的解説，並びにシミュレーションデータに対するSASの実行結果を提示する.

キーワード:

PMM, Identifying restrictions, CCMV, ACMV, NCMV, NFMV, Multiple Imputation, Shared Parameter Model, MNAR, NLMIXED procedure

# 発表の流れ

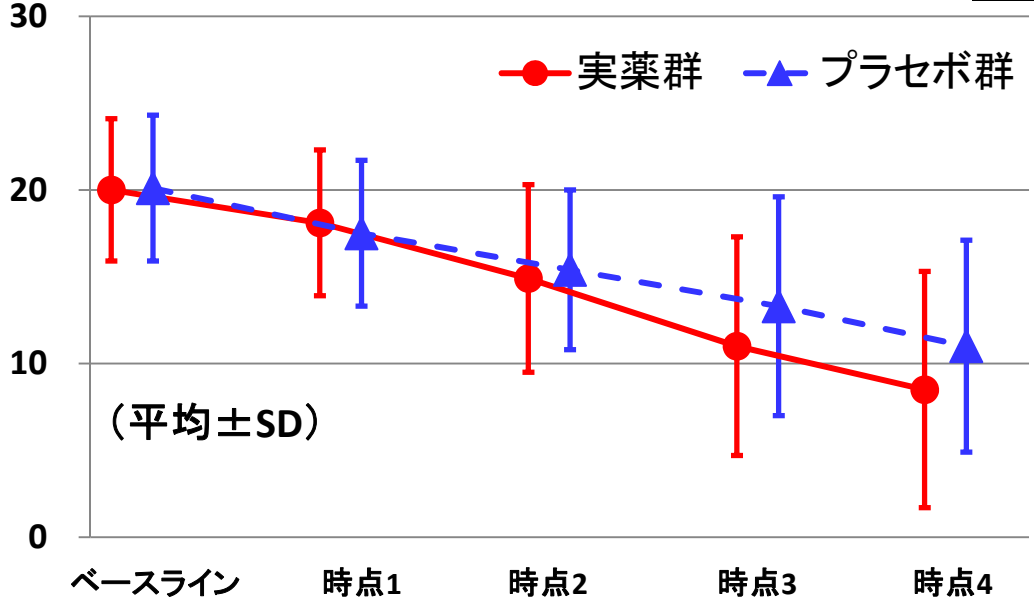
1. Multiple Imputation
2. Pattern Mixture Model

＊欠測メカニズムに対する感度分析セッションの前準備

   1. Pattern Mixture Modelとは
   2. 制約条件を仮定したPattern Mixture Model
      1. CCMV, NCMV, NFMV
      2. 実行方法, Multiple Imputation, マクロ
3. Shared Parameter Model
   1. Shared Parameter Modelとは
   2. 位置づけ
   3. マクロ
   4. シミュレーションデータを用いた解析例

# 解析対象データ

## ◎うつ病の第III相試験を想定したシミュレーションデータ




●主要評価項目：HAM-D
→ スコア低下：改善
（解析には変化量使用）

●実薬群 vs プラセボ群
・1群100例（ベースライン時）

◎単調な欠測のみ

| | ベースライン | | 時点1 | | 時点2 | | 時点3 | | 時点4 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 例数 | 平均 (SD) | 例数 | 平均 (SD) | 例数 | 平均 (SD) | 例数 | 平均 (SD) | 例数 | 平均 (SD) |
| 実薬群 | 100 | 20.0 (4.1) | 93 | 18.1 (4.2) | 89 | 14.9 (5.4) | 84 | 11.0 (6.3) | 83 | 8.5 (6.8) |
| プラセボ群 | 100 | 20.1 (4.2) | 90 | 17.5 (4.2) | 87 | 15.4 (4.6) | 85 | 13.3 (6.3) | 80 | 11.0 (6.1) |

# Multiple Imputation
# （MI）

MARに基づく
MIの解析
（主解析）

欠測値の
データに対する
MI

## ➢Multiple Imputation（MI）

- Rubin（1978,1987）によって提案された方法
- 欠測メカニズムがMARであれば，MARに基づくMIの解析(主解析)は妥当
- 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
- MIにおける解析モデルと補完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．（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

- MIが有用な状況（Dmitrienko et al, 2005）：
  共変量の欠測を補完
  PMM（MNAR）の枠組みでも適用可能
  非単調な欠測を補完

## ➢Multiple Imputation（MI）のSASの実行種類

- 単調又は非単調 欠測パターン仮定
- 表1 Proc MIにおける補完方法 (Yang Yuan, 2011)

| Pattern of missingness | Type of imputed variable | Available methods |
|---|---|---|
| Monotone | Continuous | Monotone regression |
| | | Monotone predicted mean matching |
| | | Monotone propensity score |
| Monotone | Classification (ordinal) | Monotone logistic regression |
| Monotone | Classification (nominal) | Monotone discriminant function |
| Arbitrary | Continuous | MCMC full-data imputation |
| | | MCMC monotone-data imputation |

Table 1: Imputation methods in PROC MI.

## ➢MIにおける単調回帰の概略

- 欠測値 $Y_j$を補完するため，観測されたデータ $Y_1, \ldots, Y_j$ を用いて補完モデルを構築
  $$Y_j = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k$$
- 回帰パラメータの推定値 $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$
- 共分散行列 $\hat{\sigma}_j^2 \mathbf{V}_j$ ,$\mathbf{V}_j$ は $(\mathbf{XX})^{-1}$
- パラメータの事後予測分布から下記をサンプル
  $$\beta_* = (\beta_{*0}, \beta_{*1}, \beta_{*2}, \ldots, \beta_{*k}), \sigma_{*j}^2$$
  分散: $\sigma_{*j}^2 = \hat{\sigma}_j^2 (n_j - k - 1) / g, g \sim \chi^2_{n_j - k - 1}$
  $n_j$ は$Y_j$が観測されているデータ数
  回帰係数: $\beta_* = \hat{\beta} + \sigma_{*j} \mathbf{V}_{hj}^T \mathbf{Z}, \mathbf{V} = \mathbf{V}_{hj} \mathbf{V}_{hj}^T$
  Z: 長さ(k+1)の独立の正規変数ベクトル
- 欠測値を下記の式で補完
  $$\beta_{*0} + \beta_{*1} X_1 + \beta_{*2} X_2, \ldots, \beta_{*k} X_k + z_i \sigma_{*j}$$

## ➢単調回帰のMIによるSAS プログラム（例）

①MIの実行

```
proc mi data=XXXX
  seed=シード番号
  nimpute=補完回数（補完されたデータセットの作成個数）
  out=多重補完後の出力データセット名;
  by 薬剤群変数;
  monotone method=reg(単調回帰の場合);
  var v0 val1 val2 val3 val4（解析対象変数）;
run;
```

3回～5回が推奨
(Molenberghs et al 2007)

②多重補完後，多重補完によって得られた複数のデータセットのそれぞれに対して解析モデルを適用.

例)Mixed プロシジャーによるANCOVAモデル

最終時点の値=治療効果+ベースライン値

## ➢単調回帰のMIによるSAS プログラム（例）

③解析モデルで得られた複数の解析結果より
MIANALYZEを用いて解析結果の統合

```
proc mianalyze data=XXX alpha=0.05;
   modeleffects estimate（対象推定値の変数）;
   stderr stderr（標準誤差の変数）;
   ods output ParameterEstimates=統合結果;
run;
```

# ➢単調回帰のMIによるSAS 解析結果（例）

## ①及び③で得られるSASデータセット

Proc MI 補完前 SASデータセット

| | trt | id | x0 | val1 | val2 | val3 | val4 |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1001 | 21.056933962 | 2.4655366882 | 0.4432934796 | -6.312199664 | -1.155491255 |
| 2 | 1 | 1002 | 15.064712053 | 0.409474209 | 1.8223774123 | -2.580171928 | 0.42312071 |
| 3 | 1 | 1003 | 19.406783447 | -0.245439827 | -8.394087088 | -11.21844175 | -14.7383048 |
| 4 | 1 | 1004 | 21.095469954 | -3.75371461 | -5.420467503 | -10.94184004 | -11.78001327 |
| 5 | 1 | 1005 | 23.631772118 | | | | |
| 6 | 1 | 1006 | 22.516562352 | 3.2589479676 | -1.319993637 | -4.724164414 | -6.1417252 |
| 7 | 1 | 1007 | 22.979639755 | -6.519147608 | -2.958077682 | -8.664863216 | -18.27490326 |
| 8 | | | | 495 | | | |
| 9 | | | | 388 | -13.61960921 | -17.02113392 | -14.95861723 |
| 10 | | | | 378 | -1.034573751 | -8.47499676 | -3.696205929 |
| 11 | | | | 486 | -8.351983124 | -5.218435528 | -8.894021685 |
| 12 | | | | 201 | -7.811529386 | -5.655680112 | -18.05871233 |
| 13 | | | | 281 | -7.970125036 | -8.618603275 | -16.3583909 |
| 14 | 1 | 1014 | 15.85735517 | 0.4518482447 | -1.517220221 | -7.462345983 | -16.71187236 |
| 15 | 1 | 1015 | 20.224728722 | -8.290224593 | -5.553998849 | | |
| 16 | 1 | 1016 | 19.673618172 | -1.535391886 | -10.14369464 | -7.110106417 | -10.0449832 |
| 17 | 1 | 1017 | 20.673254136 | 3.2651509625 | -1.386241991 | -4.597857718 | -11.14684396 |

Proc MI 補完後 SASデータセット

| | Imputation Number | trt | id | x0 | val1 | val2 | val3 | val4 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1001 | 21.056933962 | 2.4655366882 | 0.4432934796 | -6.312199664 | -1.155491255 |
| 2 | 1 | 1 | 1002 | 15.064712053 | 0.409474209 | 1.8223774123 | -2.580171928 | 0.42312071 |
| 3 | 1 | 1 | 1003 | 19.406783447 | -0.245439827 | -8.394087088 | -11.21844175 | -14.7383048 |
| 4 | 1 | 1 | 1004 | 21.095469954 | -3.75371461 | -5.420467503 | -10.94184004 | -11.78001327 |
| 5 | 1 | 1 | 1005 | 23.631772118 | -4.831790206 | -8.406707041 | -2.616870254 | -8.233387685 |
| 6 | 1 | 1 | 1006 | 22.516562352 | 3.2589479676 | -1.319993637 | -4.724164414 | -6.1417252 |
| 7 | 1 | 1 | 1007 | 22.979639755 | -6.519147608 | -2.958077682 | -8.664863216 | -18.27490326 |
| 8 | | | | | 9361244495 | -3.626844089 | -11.10978516 | -16.02595059 |
| 9 | | | | | 073642388 | -13.61960921 | -17.02113392 | -14.95861723 |
| 10 | | | | | 763665378 | -1.034573751 | -8.47499676 | -3.696205929 |
| 11 | | | | | 699037486 | -8.351983124 | -5.218435528 | -8.894021685 |
| 12 | | | | | 722436201 | -7.811529386 | -5.655680112 | -18.05871233 |
| 13 | | | | | 680890281 | -7.970125036 | -8.618603275 | -16.3583909 |
| 14 | | | | | 0.4518482447 | -1.517220221 | -7.462345983 | -16.71187236 |
| 15 | 1 | 1 | 1015 | 20.224728722 | -8.290224593 | -5.553998849 | -12.1528694 | -21.50380606 |
| 16 | 1 | 1 | 1016 | 19.673618172 | -1.535391886 | -10.14369464 | -7.110106417 | -10.0449832 |
| 17 | 1 | 1 | 1017 | 20.673254136 | 3.2651509625 | -1.386241991 | -4.597857718 | -11.14684396 |

Proc mianalyze 統合前データセット（群間差）

| | Imputation Number | Effect | trt | _trt | | | | | | | | pper |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | trt | 1 | 2 | -2.3011 | 0.9203 | 197 | -2.50 | 0.0132 | 0.05 | -4.1160 | -0.4862 |
| 2 | | | | 2 | -2.4733 | 0.8938 | 197 | -2.77 | 0.0062 | 0.05 | -4.2360 | -0.7106 |
| 3 | | | | 2 | -2.3716 | 0.9378 | 197 | -2.53 | 0.0122 | 0.05 | -4.2210 | -0.5223 |
| 4 | | | | 2 | -2.4758 | 0.9126 | 197 | -2.71 | 0.0073 | 0.05 | -4.2754 | -0.6761 |
| 5 | | | | 2 | -2.9340 | 0.9448 | 197 | -3.11 | 0.0022 | 0.05 | -4.7972 | -1.0707 |

各群の推定値に関する
データセットに対しても同様

Proc mianalyze 統合後データセット（群間差）

| | Parameter | Estimate | Std Error | LCLMean | UCLMean | DF | Minimum | Maximum | Theta0 | t for H0: Parameter=Theta0 | Pr > \|t\| |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | estimate | -2.511149 | 0.961074 | -4.39844 | -0.62386 | 631.4 | -2.933966 | -2.301070 | 0 | -2.61 | 0.0092 |

# ➢MI（単調回帰）+ANCOVA:解析対象データの解析結果

| 解析モデル<br>欠測メカニズム<br>/欠測確率 | 時点4における各群の点推定値(SE) | | 群間差 | 群間差のSE | P値 |
|---|---|---|---|---|---|
| | 実薬群 | プラセボ群 | | | |
| MI<br>MNAR/Low | -11.26(0.68) | -8.75(0.67) | -2.51 | 0.96 | 0.009 |

- 多重補完後の解析モデル: ANCOVA

```
proc mixed data=/*補完後データセット*/;
    by _Imputation_;
    class trt;
    model val4=x0 trt;
    lsmeans trt/pdiff=control("2") cl alpha=0.05;
    ods output diffs=/*群間差のデータセット*/;
    ods output lsmeans=/*各群の推定値のデータセット*/;
run;
```

# Pattern Mixture Model
# （PMM）

## ➢PMMとは

*Little (1993,1994,1995)
Ratitch et al. (2013)

$$Pr\left(\mathbf{Y}^{\mathrm{O}}, \mathbf{Y}^{\mathrm{M}}, \mathbf{R} \setminus \mathbf{X}\right)$$

$$= Pr\left(\mathbf{R} \setminus \mathbf{X}\right) Pr\left(\mathbf{Y}^{\mathrm{O}}, \mathbf{Y}^{\mathrm{M}} \setminus \mathbf{R}, \mathbf{X}\right)$$

$$= \boxed{Pr\left(\mathbf{R} \setminus \mathbf{X}\right) Pr\left(\mathbf{Y}^{\mathrm{O}} \setminus \mathbf{R}, \mathbf{X}\right)} \boxed{Pr\left(\mathbf{Y}^{\mathrm{M}} \setminus \mathbf{Y}^{\mathrm{O}}, \mathbf{R}, \mathbf{X}\right)}$$

Selection model

$$Pr\left(\mathbf{Y}^{\mathrm{O}}, \mathbf{Y}^{\mathrm{M}}, \mathbf{R} \setminus \mathbf{X}\right)$$

$$= Pr\left(\mathbf{R} \setminus \mathbf{Y}^{\mathrm{O}}, \mathbf{Y}^{\mathrm{M}}, \mathbf{X}\right) Pr\left(\mathbf{Y}^{\mathrm{O}}, \mathbf{Y}^{\mathrm{M}} \setminus \mathbf{X}\right)$$

観測データ，欠測データ，欠測識別変数の同時分布の分解方法が異なる

推定不可能な欠測データの確率分布に何らかの制約条件が必要

$\mathbf{Y}^{\mathrm{O}}$: 観測されたアウトカム

$\mathbf{Y}^{\mathrm{M}}$: 観測されなかった（欠測）アウトカム

$\mathbf{R}$ : 欠測識別変数

$\mathbf{X}$ : 観測されている共変量

NRC (2010)
Type1:欠測データの分布に対する検証不可能な仮定
Type2:観測データの分布に対する検証可能な仮定

**観測データから推定可能 Type 2**

**観測データから推定不可能 Type 1**

## ➢欠測パターンの例

欠測メカニズムがMARやMNARの状況を想定

①脱落時点に基づく脱落パターン

例)　観測○　未観測×

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

（ア）パターン1のデータのみで解析した場合，バイアスが入る
（イ）補完してバイアスを減らしたい
（ウ）制約条件を仮定し，×のデータを補完
=>×の分布が不明
=>制約条件（欠測メカニズムと対応）を仮定し，周辺の○のデータを参考に，×のデータを補完

②中止理由に基づく脱落パターン

中止理由1：有害事象により中止

中止理由2：効果不足

中止理由3：被験者の同意撤回　など

他にもパターンの定義は可能

## ➢ 制約条件と欠測メカニズムの関係図

Missing data in clinical trials page 37, Molenberghs et al. (2007)より

$$
\begin{array}{lcccccccc}
\text{SEM} & : & \text{MCAR} & \subset & \text{MAR} & \subset & \text{MNFD} & \subset & \text{general MNAR} \\
 & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow \\
\text{PMM} & : & \text{MCAR} & \subset & \text{ACMV} & \subset & \text{NFMV} & \subset & \text{general MNAR} \\
 & & & & & \subset & \neq & \subset & \\
 & & & & & & \text{interior} & &
\end{array}
$$

**Figure 3.1** Relationship between nested families within the selection model (SEM) and pattern-mixture model (PMM) families. (MCAR, missing completely at random; MAR, missing at random; MNAR, missing not at random; MNFD, missing non-future dependence; ACMV, available case missing values; NFMV, non-future missing values; interior, restrictions based on a combination of the information available for other patterns. The $\subset$ symbol here indicates 'is a special case of'. The $\updownarrow$ symbol indicates correspondence between a class of SEM models and a class of PMM models.)

➢制約条件（CCMV,NCMV,ACMV）を仮定したPMM

Thijs et al. (2002)は上記の制約条件を統一的に示す形でPMMを提案

t時点まで観測

$$f\left(y_1,\ldots,y_T,\boxed{r=t}\right)=f_t\left(y_1,..,y_t\right)\boxed{f_t\left(y_{t+1},y_{t+2},\ldots,y_T|y_1,..,y_t\right)}$$

未観測データの分布

脱落パターン $r=1,\cdots,T$

➢制約条件（CCMV,NCMV,ACMV）

$$f_t\left(y_s|y_1,\ldots,y_{s-1}\right)=\sum_{j=s}^{T}\boxed{\omega_{sj}}f_j\left(y_s|y_1,\ldots,y_{s-1}\right),s=t+1,..,T$$

この重み$\omega_{sj}$（s:補完対象時点，j:どのパターンの分布）の設定により，制約条件CCMV,NCMV,ACMVを表現

以上，3つの制約条件を完全データの分布関数に代入

$$f_t\left(y_1,\ldots,y_T\right)=f_t\left(y_1,..,y_t\right)\prod_{s=0}^{T-t-1}\left(\sum_{j=T-s}^{T}\omega_{T-s,j}f_j\left(y_{T-s}|y_1,\ldots,y_{T-s-1}\right)\right)$$

## ➢制約条件1（CCMV:Complete Case Missing Value）

$$\omega_{T,T} = \omega_{T-1,T} = \omega_{T-2,T} = \ldots = \omega_{t+1,T} = 1$$

かつその他 $\omega_{sj} = 0, j \neq T$ のとき

$$f_t(y_s|y_1, \ldots, y_{s-1}) = f_T(y_s|y_1, \ldots, y_{s-1}), s = t+1, \ldots, T$$

最後まで観測されたCompleter（例，パターン1）の情報から欠測の情報を補完するもの．完全パターンのデータを利用し，多くの被験者がCompleterのパターンの場合を満たしている場合に有用．また，Non-monotoneのときにも利用が容易．

例）　観測〇　未観測×

| | | | | |
|---|---|---|---|---|
| パターン1 | 時点1 〇 | 時点2 〇 | 時点3 〇 | 時点4 〇 |
| パターン2 | 時点1 〇 | 時点2 〇 | 時点3 〇 | 時点4 × |
| パターン3 | 時点1 〇 | 時点2 〇 | 時点3 × | 時点4 × |

## ➢制約条件2（NCMV:Neighboring Case Missing Value）

$$\omega_{T,T} = \omega_{T-1,T-1} = \omega_{T-2,T-2} = \ldots = \omega_{t+1,t+1} = 1$$

かつその他 $\omega_{sj} = 0, j \neq s$ のとき

$$f_t(y_s|y_1, \ldots, y_{s-1}) = f_s(y_s|y_1, \ldots, y_{s-1}), s = t+1, \ldots, T$$

欠測時点で測定された一番近いパターンの情報から欠測の情報を補完するもの．たとえば，下記の例でパターン3の時点3を補完したい場合，一番欠測パターンが近いパターン2のデータの分布から補完

例)　観測〇　未観測×

パターン1　時点1 〇 時点2 〇　時点3 〇　時点4 〇

パターン2　時点1 〇 時点2 〇　時点3 〇　時点4 ×

パターン3　時点1 〇 時点2 〇　時点3 × 時点4 ×

## ➢制約条件3（ACMV: Available Case Missing Value）

$$\omega_{sj} = \frac{\alpha_j f_j(y_1,\ldots,y_{s-1})}{\sum_{l=s}^{T} \alpha_l f_l(y_1,\ldots,y_{s-1})}$$

$\alpha_j$ : 脱落パターンjにおける観測された割合

$$\Rightarrow f_t(y_s \setminus y_1,\ldots,y_{s-1}) = f_{\geq s}(y_s \setminus y_1,\ldots,y_{s-1})$$

$$= \sum_{j=s}^{T} \omega_{sj} f_j(y_s \setminus y_1,\ldots,y_{s-1}), s = t+1,\ldots,T$$

## ➢制約条件3（ACMV: Available Case Missing Value）

欠測時点以降に観測されている全てのパターンの情報から欠測の情報を補完するもの．先ほどの例で，パターン3の時点3の値を補完したい場合，欠測のパターンが一番近いパターン2とCompleteパターン1のデータの分布から補完

例）　観測〇　未観測×

| | | | | |
|---|---|---|---|---|
| パターン1 | 時点1 〇 | 時点2 〇 | 時点3 〇 | 時点4 〇 |
| パターン2 | 時点1 〇 | 時点2 〇 | 時点3 〇 | 時点4 × |
| パターン3 | 時点1 〇 | 時点2 〇 | 時点3 × | 時点4 × |

## ➢制約条件（NFMV）を仮定したPMM

Kenward et al. (2003)は**MNARを仮定したPMM**を提案

ACMVの結果と**NFMV+CCMV又はNFMV+NCMVの結果比較**することは，

MARの欠測メカニズムからの乖離を評価

$$
\begin{aligned}
f(y_1, \ldots, y_T, r = t) &= f(y_1, \ldots, y_T | r = t)\, f(r = t) \\
&= f_t(y_1, \ldots, y_T)\, f(r = t) \\
&= \boxed{f_t(y_1, \ldots, y_t)\, f_t(y_{t+1} | y_1, \ldots, y_t)\, f_t(y_{t+2}, \ldots, y_T | y_1, \ldots, y_{t+1})} \\
&\quad \times \boxed{f(r = t)}
\end{aligned}
$$

観測

$$f_t(y_1, \ldots, y_t) \quad f_t(y_{t+1} | y_1, \ldots, y_t) \quad \prod_{s=t+2}^{T} \boxed{f_t(y_s | y_1, \ldots, y_{s-1})}$$

過去（t時点まで） | 現在（t+1時点） | 未来（t+2時点以降）

観測 | 未観測 | 未観測

制約条件が必要

## ➢制約条件（NFMV: Non-Future Missing Value）

$s \geq t+2$ に対して

$$f\left(y_s | y_1, \ldots, y_{s-1}, r=t\right) = f\left(y_s | y_1, \ldots, y_{s-1}, r \geq s-1\right)$$

- 脱落が今現在観測されていないデータに依存し，未来に観測されていないデータには依存しない
- 欠測メカニズム(Selection ModelのMissing Non-Future Dependent:MNFD)に対応

現在（t+1）の未観測のデータ

$$f\left(r=t | y_1, \ldots, y_T\right) = f\left(r=t | y_1, \ldots, y_{t+1}\right)$$

- 完全にデータの分布が特定できない，すなわち下記の条件付き分布のみが特定されていない
  =>制約条件（CCMVやNCMV）の追加が必要

$$f\left(y_s \setminus y_1, \ldots, y_{s-1}, r=s-1\right)$$

**NFMV+CCMV,NFMV+NCMV => MNARを仮定**

## ➢PMMマクロ（参考：Missingdata.org.uk）

```
%macro %patternmixture(analyset=XXX, /*データセット名*/
Iconstraint=XXX, /*補完モデルに使用, CCMV or NCMV*/
Itype=XXX, /*時点制約としてNFMV or ALL*/
seedgen=XXX, numberimputations=X, /*シード数と補完数*/
YVAR=%STR(XXX), /*目的変数*/
MODEL=%STR(XXXX), /*補完モデル*/
modvars=%STR(xxx),
classvars=%STR(XXX),  id=, /*補完モデルの質的変数*/
MODEL2=%STR(XX), /*解析モデル，補完モデルと同一でなくてよい*/
classvars2=%STR(XX), /*解析モデルの質的変数*/
othervars=%STR(XXX), /*Lsmean算出に必要な連続変数*/
TITLE=%STR(),
FOOTNOTE=%STR( ));
```

# ➢PMMマクロの例

## - NFMV+CCMV

補完モデル：
治療効果，時点効果，ベースライン値の飽和モデル
解析モデル：
ベースライン値，脱落パターン，治療，時点，脱落パターンと治療の交互作用，治療と時点の交互作用，脱落パターンと治療と時点の交互作用
制約条件：NFMV+CCMV
データセット：MNARかつ欠測確率がLow

```
* complete case;
%patternmixture(
   analyset=D7,
   Iconstraint=CCMV,
   Itype=NFMV,
   seedgen=8857954,
   numberimputations=5,
  YVAR=%STR(VAL),
  MODEL=%STR(base treatment visit base*treatment base*visit treatment*visit base*treatment*visit),
  modvars=%STR(subjid treatment base),
   classvars=%STR(treatment visit),  id=,
   MODEL2=%STR(base dgroup treatment visit dgroup*treatment  treatment*visit dgroup*treatment*visit),
   classvars2=%STR(dgroup  treatment visit),
   othervars=%STR(base),
   TITLE=%STR(CCMV Pattern Mixture Model Analysis),
   FOOTNOTE=%STR( )
);
```

# ➢PMMマクロ（データ：MNAR，欠測確率Low）のSAS output

- NFMV+CCMV

CCMV Pattern Mixture Model Analysis
Using CCMV identifiability constraint
By visit LSMEANS using Model with Effects for Dropout Pattern

| Visit (Week) | Treatment | N | LSMEAN Change | SE | LSmean Difference | SE | Lower CL | Upper CL | P-value |
|---|---|---|---|---|---|---|---|---|---|
| Visit 1 | Drug | 93 | -1.97 | 0.44 | . | . | . | . | . |
| | Placebo | 90 | -2.46 | 0.44 | 0.49 | 0.63 | -0.74 | 1.72 | .4329 |
| Visit 2 | Drug | 89 | -5.01 | 0.52 | . | . | . | . | . |
| | Placebo | 87 | -4.58 | 0.53 | -0.42 | 0.73 | -1.86 | 1.01 | .5621 |
| Visit 3 | Drug | 84 | -8.72 | 0.68 | . | . | . | . | . |
| | Placebo | 85 | -6.59 | 0.67 | -2.13 | 0.97 | -4.04 | -0.22 | .0288 |
| Visit 4 | Drug | 83 | -11.28 | 0.74 | . | . | . | . | . |
| | Placebo | 80 | -8.85 | 0.72 | -2.43 | 1.02 | -4.43 | -0.43 | .0171 |
| Overall | Drug | 93 | -6.75 | 0.49 | . | . | . | . | . |
| | Placebo | 90 | -5.62 | 0.49 | -1.12 | 0.69 | -2.48 | 0.23 | .1046 |

---

NFMV observations used subject to identifiable constraint
Model:base dgroup treatment visit dgroup*treatment treatment*visit dgroup*treatment*visit

# Shared Parameter Model
# （SPM）

# Shared Parameter Model

$$f(Y_i, R_i, b_i) = f(Y_i|R_i, b_i)f(R_i|b_i)f(b_i)$$

被験者 $\mathbf{b}_i = (b_{i1}, \ldots, b_{iq})$ $\mathbf{b}_i \sim N(\mathbf{0}, \Sigma)$

- 測定過程に対するモデル，及び脱落過程に対するモデルの両方に影響する潜在変数 (変量効果) を考える.
- 変量効果の条件付きで，測定過程と脱落過程を表す密度関数は分離できると仮定する.

定義

$$f(Y_i, R_i, b_i) = f(Y_i{}^o|b_i)f(Y_i{}^m|b_i)f(R_i|b_i)f(b_i)$$

# Shared Parameter Model

$$f(Y_i, R_i, b_i) = \underline{f({Y_i}^o|b_i)f({Y_i}^m|b_i)}\,\underline{f(R_i|b_i)}f(b_i)$$

- ある時点における脱落が，アウトカムではなく，アウトカムにも関連する個々人の潜在的な特性の影響を受けると考える．(Little, 1995)
- 共通の変量効果が測定過程モデルと脱落過程モデルに含まれると考える場合，欠測メカニズムがMNARの場合に対応する.
- SM（Selection Model)との違い

$$f(Y_i, R_i) = \underline{f({Y_i}^o, {Y_i}^m)}\,\underline{f(R_i|{Y_i}^o, {Y_i}^m)}$$

# 尤度の計算

$$
\begin{aligned}
f({Y_i}^o, R_i) &= \int_{y^m} \int_b f({Y_i}^o, {Y_i}^m, r_i, b_i) db_i d{Y_i}^m \\
&= \int_{y^m} \int_b f({Y_i}^o, {Y_i}^m | b_i) f(r_i | b_i) f(b_i) db_i d{Y_i}^m \\
&= \int_{y^m} \int_b f({Y_i}^o | b_i) f({Y_i}^m | b_i) f(r_i | b_i) f(b_i) db_i d{Y_i}^m \\
&= \int_b f({Y_i}^o | b_i) f(r_i | b_i) f(b_i) \left( \int_{y^m} f({Y_i}^m | b_i) d{Y_i}^m \right) db_i \\
&= \int_b f({Y_i}^o | b_i) f(r_i | b_i) f(b_i) db_i
\end{aligned}
$$

SAS のNLMIXED プロシジャを用いることで実装可能である.

# Shared Parameter Modelの位置づけ

- 感度分析としてのSPM
  - 欠測がMNARの場合において，Selection Model，及びPattern Mixture Modelと並ぶ感度分析の方法である．
    - アウトカムの測定誤差が大きく，脱落するかどうかにアウトカムの値が依存すると考えるより，症状や病態の進行度合いのような個々人の潜在的な効果が関係していると思われる状況で，有用なアプローチと考えられる．(Little, 1995)

SPM $$f(Y_i, R_i, b_i) = f({Y_i}^o|b_i)f({Y_i}^m|b_i)f(R_i|b_i)f(b_i)$$

SM $$f(Y_i, R_i) = f({Y_i}^o, {Y_i}^m)f(R_i|{Y_i}^o, {Y_i}^m)$$

PMM $$f(Y_i, R_i) = f({Y_i}^o, {Y_i}^m|R_i)f(R_i)$$

# マクロ:%shared_parameter

- DIA working group公開マクロを参考にして作成

```
%shared_parameter(
INPUTDS = temp_sim1,
SUBJVAR = id,
TRTVAR = Drug,
TIME = time,
MODL =
   %STR(val = x0 Drug time Drug*time),

LINK=%STR(CLOGLOG)

RANDOM_SLOPE =%STR(LINEAR)
);
```

インプットデータ

被験者

薬剤

時点

測定過程に対するモデルに含まれる固定効果

脱落過程に対するモデル

変量効果の選択

# マクロの仕様

- マクロでは4パターンのモデルが表現可能

モデル1

測定過程のモデルに含めた変量効果を，脱落過程モデルは含まない

モデル2

測定過程のモデルに含めた変量効果を，脱落過程モデルも含む

各モデルに対して・・

RANDOM_SLOPE =%STR(NONE)

変量効果は切片しか含まない

RANDOM_SLOPE =%STR(LINEAR)

変量効果は，切片，及び時点に対する効果も含む

# マクロ:%shared_parameter

- 測定過程に対するモデル

RANDOM_SLOPE =%STR(NONE)

変量効果は切片しか含まない

$$Y_{ij} = \beta_0 + \beta_1 Time_{ij} + \beta_2 Group_i + \beta_3 Time_{ij} \times Group_i + \underline{b_{i0}} + e_{ij}$$

RANDOM_SLOPE =%STR(LINEAR)

変量効果は，切片，及び時点に対する効果も含む

$$Y_{ij} = \beta_0 + \beta_1 Time_{ij} + \beta_2 Group_i + \beta_3 Time_{ij} \times Group_i + \underline{b_{i0} + b_{i1} Time_{ij}} + e_{ij}$$

# マクロ:%shared_parameter

- 脱落過程に対するモデル
  - Complementary log-log linkモデル

RANDOM_SLOPE =%STR(NONE)

変量効果は切片しか含まない

$$Pr(D_i = j | D_i \geq j) = 1 - exp(-exp(\alpha_{0j} + \underline{\gamma_1 b_{i0}}))$$

RANDOM_SLOPE =%STR(LINEAR)

変量効果は，切片，及び時点に対する効果も含む

$$Pr(D_i = j | D_i \geq j) = 1 - exp(-exp(\alpha_{0j} + \underline{\gamma_1 b_{i0} + \gamma_2 b_{i1}}))$$

- ロジットモデルも使用可能

# シミュレーションデータを用いた解析例 (1)

| Parameter | No Linkage ML est | std error | p-value | Intercept linkage (do not vary by treatment) ML est | std error | p-value |
|---|---|---|---|---|---|---|
| b0 | 14.205 | 1.693 | <.001 | 14.185 | 1.685 | <.001 |
| bDrug | 1.532 | 0.887 | .086 | 1.519 | 0.888 | .089 |
| bDrugtime | -1.050 | 0.251 | <.001 | -1.043 | 0.251 | <.001 |
| btime | -2.150 | 0.178 | <.001 | -2.133 | 0.179 | <.001 |
| bx0 | -0.726 | 0.078 | <.001 | -0.725 | 0.078 | <.001 |
| aDrug | -0.015 | 0.447 | .972 | -0.010 | 0.452 | .982 |
| aInt | | | | 0.112 | 0.073 | .129 |
| aDint | | | | | | |
| endpoint trt diff | -2.667 | 0.761 | <.001 | -2.652 | 0.762 | <.001 |
| endpoint trt diff (Mixed Model) | -2.668 | 0.761 | <.001 | | | |

$\beta$

$\gamma$

モデル1

Mixed Model

モデル2

RANDOM_SLOPE =%STR(NONE)

変量効果は切片しか含まない

# シミュレーションデータを用いた解析例 (2)

| Parameter | No Linkage ML est | No Linkage std error | No Linkage p-value | Intercept and Linear slope linkage (do not vary by treatment) ML est | std error | p-value |
|---|---|---|---|---|---|---|
| b0 | 13.764 | 1.489 | <.001 | 12.147 | 1.488 | <.001 |
| bDrug | 1.491 | 0.695 | .033 | 0.943 | 0.688 | .172 |
| bDrugtime | -1.022 | 0.281 | <.001 | -0.875 | 0.291 | .003 |
| btime | -2.135 | 0.200 | <.001 | -2.167 | 0.207 | <.001 |
| bx0 | -0.705 | 0.070 | <.001 | -0.616 | 0.070 | <.001 |
| aDrug | -0.015 | 0.447 | .973 | -2.080 | 3.004 | .490 |
| aInt | | | | -14.117 | 7.039 | .046 |
| aSlp | | | | 22.368 | 10.085 | .028 |
| aDint | | | | | | |
| aDslp | | | | | | |
| endpoint trt diff | -2.596 | 0.939 | .006 | -2.559 | 0.976 | .009 |
| endpoint trt diff (Mixed Model) | -2.596 | 0.939 | .006 | | | |

$\beta$

$\gamma$

モデル1

Mixed Model

モデル2

```
RANDOM_SLOPE =%STR(LINEAR)
```

変量効果は，切片，及び時点に対する効果も含む

# 最後に

- 欠測がMNARの場合における，感度分析の手法として，PMM，及びSPMを取り上げた

PMM $$f(Y_i, R_i) = f({Y_i}^o, {Y_i}^m | R_i) f(R_i)$$

SPM $$f(Y_i, R_i, b_i) = f({Y_i}^o | b_i) f({Y_i}^m | b_i) f(R_i | b_i) f(b_i)$$

SM $$f(Y_i, R_i) = f({Y_i}^o, {Y_i}^m) f(R_i | {Y_i}^o, {Y_i}^m)$$

# 参考文献


- Dmitrienko, A., Molenberghs, G., Chuang-Stein, C., Offen, W., (2005) Analysis of Clinical Trials Using SAS: *A Practical Guide* .Carry, NC: SAS Institute Inc.
- Kenward, M.G., Molenberghs, G. & Thijs, H. (2003) Pattern mixture models with proper time dependence. *Biometrika*. **90**(1), 53-71
- Mallinckrodt, C. H. (2013). *Preventing and Treating Missing Data in Longitudinal Clinical Trials.* Cambridge University press.
- Molenberghs, G., Kenward, M.G. (2007) Missing Data in Clinical Studies. *Chichester. UK*, *John Wiley*.
- National Research Council (2010). *The Prevention and Treatment of Missing Data in Clinical Trials*. Washington, DC: The National Academies Press.
- Ratitch, B., O'Kelly, M., Tosiello, R. (2013) Missing data in clinical trials from clinical assumptions to statistical analysis using pattern mixture models. *Phramaceutical Statistics*. 12, 337-347

# 参考文献


- Rubin DB (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
- Rubin DB (1987). Multiple Imputation for Nonresponse in Surveys. *John Wiley & Sons*.
- Yang Yuan (2011). Multiple Imputation Using SAS Software. *Journal of Statistical Software*. **45**(6), 1-25 .
- Thijs, H., Molenberghs, G. (2002) Strategies to fit pattern mixture models. Biostatistics. **3**(2), 245-265
- Follmann, D. and Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. Biometrics 51, 151–168.
- Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. Journal of the American Statistical Association **88**, 125–134.

# 参考文献


- Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. Biometrika **81**, 471–483.
- Little, R. J. A. (1995). Modelling the drop-out mechanism in repeated-measures studies. Journal of the American Statistical Association **90**, 1112–1121.
- Wu M., & Follmann, D.A. “Use of Summary Measures to Adjust for Informative Missingness in Repeated Measures with Random Effects,” *Biometrics*, (1999); 55 75-84