

【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

(4)欠測メカニズムに対する感度分析

駒寄弘<sup>1)2)</sup> 高橋文博<sup>1)3)</sup> 横溝孝明<sup>1)4)</sup>

- 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) マルホ株式会社 3) 田辺三菱製薬株式会社 4) 大正製薬株式会社

Sensitivity analysis for the missing  
mechanism.

Hiroshi Komazaki<sup>1)2)</sup>, Fumihiro Takahashi<sup>1)3)</sup>, Takaaki Yokomizo<sup>1)4)</sup>

- 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) Maruho Co., Ltd. 3) Mitsubishi Tanabe Pharma Corporation 4) Taisho Pharmaceutical Co., Ltd.

## 要旨：

Selection Model 及びPattern Mixture modelを用いて、  
欠測メカニズムに対する感度分析の方法を説明する。また、  
SASによるシミュレーション結果を提示する。

キーワード：欠測メカニズム，感度分析，MAR，MNAR，Type(i)・  
Type(ii)の仮定，感度パラメータ，Selection Model，Pattern  
Mixture model，

# Contents

- NRC(2010)で提案された感度分析の種類
- 欠測メカニズムに対する感度分析
  - ✓ 感度分析を行う理由
    - 主要な解析はMAR? or MNAR?
    - Type(i), Type(ii)の仮定と感度パラメータ
  - ✓ SMを用いた感度分析
    - モデルの説明, 感度パラメータによるMAR, MNARの仮定
    - シミュレーションデータを用いた解析例
  - ✓ PMMを用いた感度分析
    - モデルの説明, 感度パラメータによるMAR, MNARの仮定
    - シミュレーションデータを用いた解析例
  - ✓ SMとPMMの感度分析の違い

# Contents

- 本感度分析の結果の考察方法～NRC(2010)の内容より～

## 記号の整理 (Selection Model)

対象となるデータ: 経時データ(連続値)

$$\mathbf{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in_i} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_i^o \\ \mathbf{Y}_i^m \end{pmatrix}$$

( $i = 1, \dots, N$ )

$n_i$  : 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

$$R_{ij} = \begin{cases} 1 & \text{被験者 } i \text{ の } j \text{ 時点でのデータが観測された} \\ 0 & \text{被験者 } i \text{ の } j \text{ 時点でのデータが欠測} \end{cases}$$

$$\mathbf{R}_i = \begin{pmatrix} R_{i1} \\ \vdots \\ R_{in_i} \end{pmatrix}$$

## パラメータ

$\theta$  : 応答変数モデルのパラメータ

$\psi$  : 欠測モデルのパラメータ

## 記号の整理 (Pattern Mixture Model)

対象となるデータ: ある測定時点のデータ(連続値)

$Y_i^m$  : 被験者*i*の未観測データ

$Y_i^o$  : 被験者*i*の観測データ

$\mu_0$  : 未観測データの平均

$\mu_1$  : 観測データの平均

### 欠測識別変数

$R = \begin{cases} 1 & \text{データが観測} \\ 0 & \text{データが欠測} \end{cases} \quad \pi : \text{観測割合}$

### パラメータ

$\beta = (\beta_0, \beta_1)^T$  : 応答変数モデルのパラメータ



## NRC(2010)で提案された感度分析の種類

□ 完全データの分布の仮定

□ 外れ値の影響

□ 欠測メカニズムの仮定  
(MAR or MNAR)



本セッションで取り上げる感度分析  
NRC (2010)で『最も重要』と指摘

## 主要な解析はMAR or MNAR？

### □ MARを仮定した解析を主解析とすることの正当性と問題点

- ✓ 途中脱落が多いと想定される臨床試験では、周到に計画された上で、欠測がMARとなるよう十分な情報を収集できる試験デザイン考えるべき
- ✓ しかしながらMARの仮定はデータからは確認できない
- ✓ 実際のデータは欠測メカニズムがMARやMNARの混合である可能性もある

### □ MNARの仮定の下で解析する際の問題点

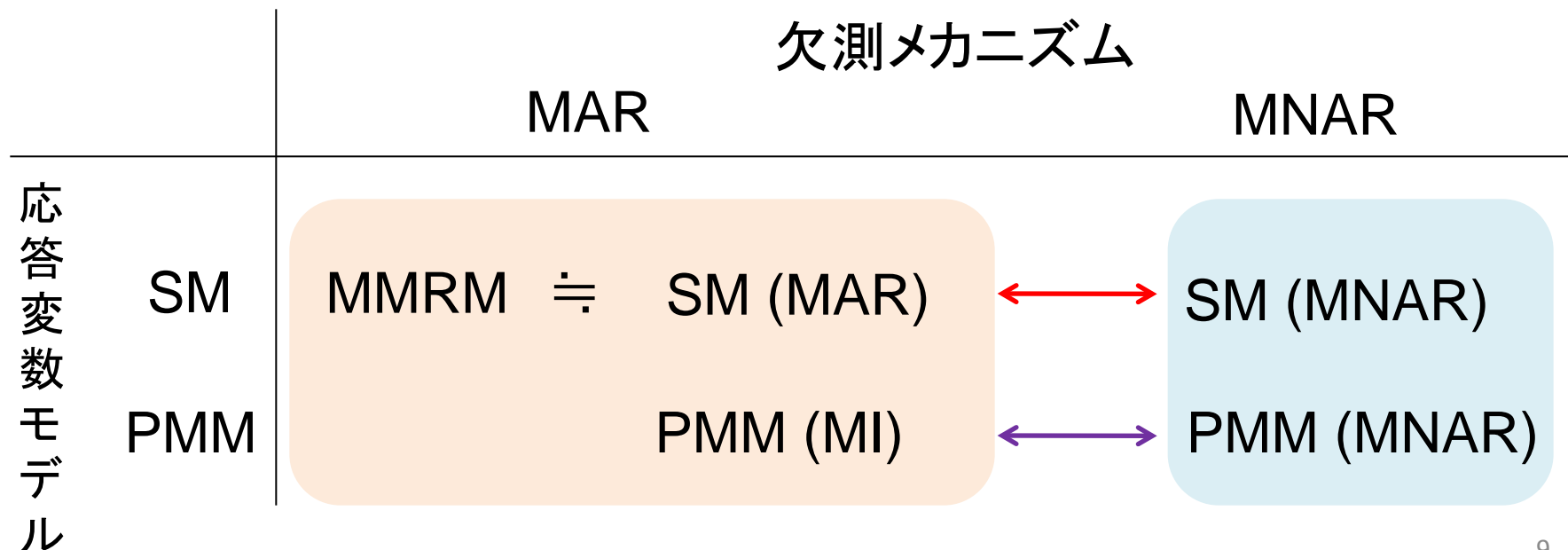
- ✓ 応答変数モデルまたは欠測モデルの仮定を恣意的に置く必要がある
- ✓ 妥当性の検証できない仮定をしている
- ✓ 仮定が誤っていた場合の結果の頑健性は脆い



**□欠測メカニズムに対する感度分析を含む解析手順の提案**

1. 主解析はMARを仮定した解析
2. 感度分析として、MNARの仮定の下で解析
3. MARを仮定した解析の結果の頑健性を確認し、試験結果の妥当性を述べる

□ これら感度分析はSM及びPMMより実行可能.



## Type(i), Type(ii)の仮定と感度パラメータ

欠測を含むデータの解析を行う際、モデルは部分的にtype(i), (ii)の二種類の仮定に分けられ、それぞれ感度分析の方法は異なる。

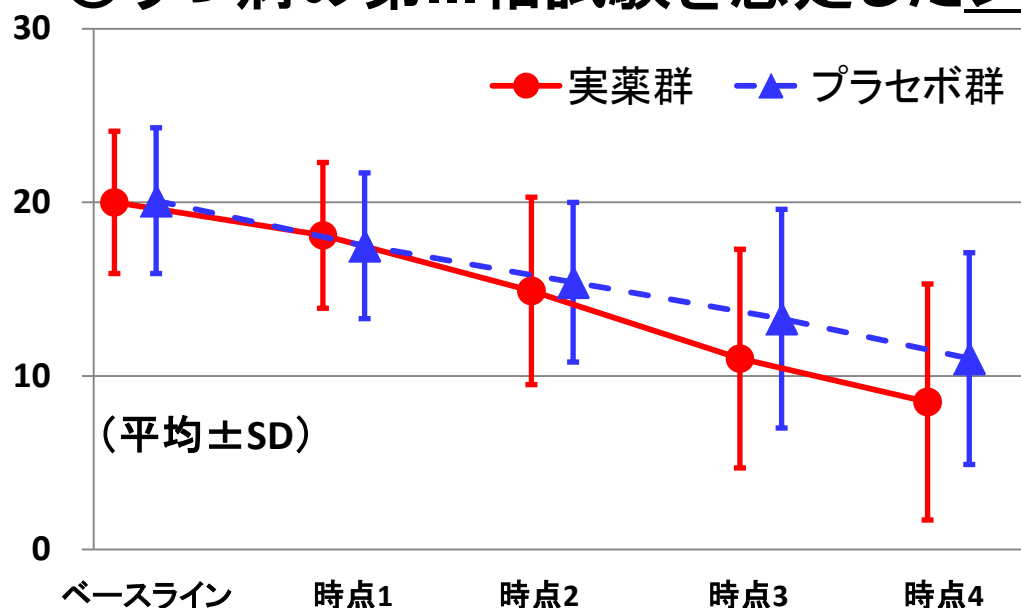
Type (ii) : **観測データの分布**に対する**検証可能**な仮定  
→ 仮定の正しさ(モデルの適合度など)をみる。

Type (i) : **欠測データの分布**に対する**検証不可能**な仮定  
→ 観測データの分布情報 + **感度パラメータ**を用いることで、解析者が欠測データの分布を**任意**に仮定する。(MNARの仮定の下での解析が可能となる。)  
→ 欠測メカニズムがMARの仮定から離れることに対する(MARを仮定した)主解析の結果の頑健性を確認する。

本セッションでは、NRC(2010)で紹介されている方法に準拠し、SMとPMMのモデル及び**感度パラメータを用いた**欠測メカニズム(MAR or MNAR)に対する感度分析の方法を紹介する。

## 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)

## SMを用いた感度分析

$$\begin{aligned} f(\mathbf{Y}_i, \mathbf{R}_i | \boldsymbol{\theta}, \boldsymbol{\psi}) &= f(\mathbf{Y}_i | \boldsymbol{\theta}) \cdot f(\mathbf{R}_i | \mathbf{Y}_i, \boldsymbol{\psi}) \\ &= f(\mathbf{Y}_i^o, \mathbf{Y}_i^m | \boldsymbol{\theta}) \cdot \underbrace{f(\mathbf{R}_i | \mathbf{Y}_i^o, \mathbf{Y}_i^m, \boldsymbol{\psi})}_{\text{Type (i) の仮定}} \end{aligned}$$

$$\begin{aligned} \text{logit}\{Pr(R_{ij} = 0 | R_{i1} = 1, \dots, R_{i,j-1} = 1, \mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\psi})\} \\ = \psi_0 + \psi_1 Y_{i,j-1} + \psi_2 Y_{ij} \end{aligned}$$

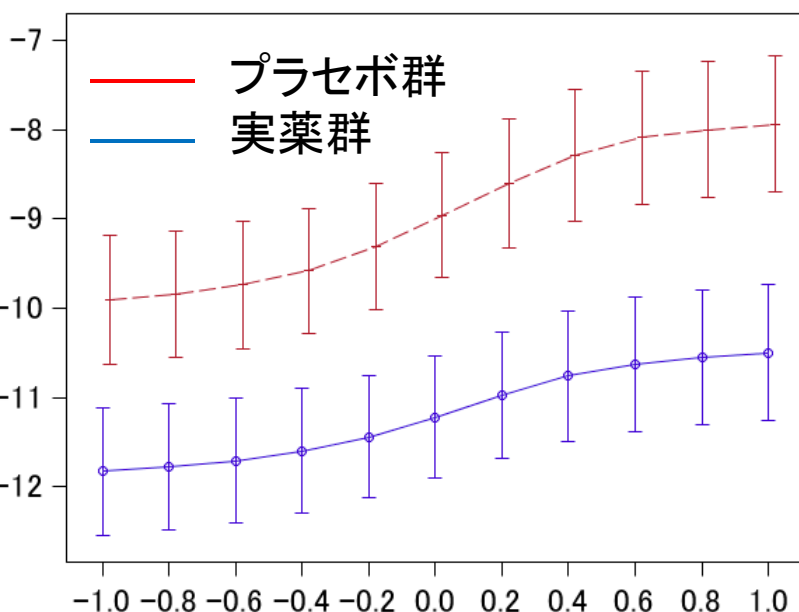
**感度パラメータ:** 値を解析者が設定  
色々な値を代入して結果の頑健性をみる

## SMを用いた感度分析

- SMでは、欠測モデル、応答変数モデル共にType(i) の仮定である.
- 応答変数について、群間差が評価できる範囲で、可能な限りモデルを仮定しないセミパラメトリックな方法もある. (IPW法: Inverse Probability Weighting)
- SMの感度パラメータは、現在測定された応答変数の大きさが欠測の有無に与える影響度合いを表している.
- $\psi_2 = 0$  のとき, MAR. それ以外の場合, MNARを意味している.

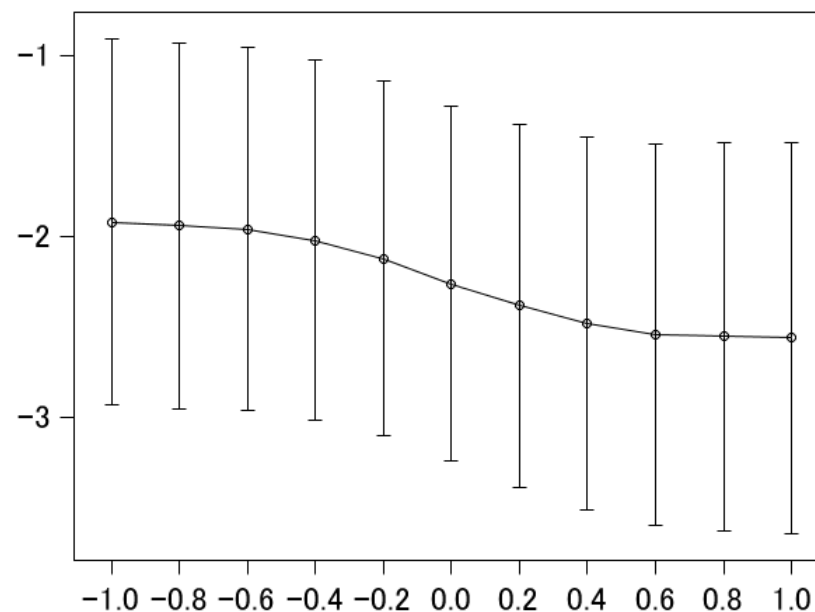
# シミュレーションデータを用いた解析例 ～SMによる感度分析の結果～

## □ 各群の点推定値±SE



感度パラメータ  $\psi_2$

## □ 群間の差の点推定値±SE



感度パラメータ  $\psi_2$



## PMMを用いた感度分析

$$f(Y_i^o, Y_i^m, R_i \setminus X_i, \beta) \\ = f(R_i \setminus X_i) f(Y_i^o \setminus R_i, X_i, \beta) \boxed{f(Y_i^m \setminus Y_i^o, R_i, X_i, \beta)}$$

未観測のデータのモデルに対する感度  
をみる

### 【目的】

制約条件を置くことに加えて、未観測のデータの応答変数の分布と観測データの応答変数の分布の乖離を評価することで欠測データのメカニズムMARの乖離を評価する。

PMMの経時測定モデル 今回の解析では、全ての未観測データに対して-Δを付加

$$\begin{aligned}
 f_t(y_1, \dots, y_T) &= f_t(y_1, \dots, y_t) f_t(y_{t+1} | y_1, \dots, y_t) f_t(y_{t+2}, \dots, y_T | y_1, \dots, y_{t+1}) \\
 &= f_t(y_1, \dots, y_t) \boxed{f_t(y_{t+1} | y_1, \dots, y_t)} \prod_{s=t+2}^T \boxed{f_t(y_s | y_1, \dots, y_{s-1})}
 \end{aligned}$$

$s \geq t+2$

NFMVの制約条件  
 $s \geq t+2$ に対して、

経時測定データにおける  
未観測のデータの感度対象

$$\boxed{f(y_s | y_1, \dots, y_{s-1}, r = t)} = f(y_s | y_1, \dots, y_{s-1}, \boxed{r \geq s-1})$$

ただし、上記の $r=s-1$ のときの条件付き分布と下記の分布

$$\boxed{f(y_{t+1} | y_1, \dots, y_t, r = t)}$$

は特定されていない。よって、追加の**NCMV**(+Δ)の制約条件を付加

$$\begin{aligned}
 f(y_s \setminus y_1, \dots, y_{s-1}, r = s-1) &= f(y_s \setminus y_1, \dots, y_{s-1}, r = s) \\
 f(y_{t+1} \setminus y_1, \dots, y_t, r = t) &= f(y_{t+1} \setminus y_1, \dots, y_t, r = t+1)
 \end{aligned}$$

OR

$$\begin{aligned}
 f(y_s - \Delta \setminus y_1, \dots, y_{s-1}, r = s) \\
 f(y_{t+1} - \Delta \setminus y_1, \dots, y_t, r = t+1)
 \end{aligned}$$

## 2時点(シンプル)の場合

興味あるパラメータ: Yの平均 $\mu$ 未観測の人(R=0)の平均と観測されている人(R=1)の平均の乖離を感度パラメータ $\Delta$ で単純に表現

$$\mu_0 = \mu_1 + \Delta \Leftrightarrow E(Y / R = 0) = E(Y / R = 1) + \Delta$$

$$\text{一般的な場合: } g(\mu_0) = g(\mu_1) + \Delta \Leftrightarrow E(Y / R = 0) = g^{-1}(g(E(Y / R = 1)) + \Delta)$$

解析担当者が関数gを指定, 単純ケースでは  $g(\mu) = \mu$ 

検証不可能な仮定

Yの平均 $\mu$ は下記のように表現

$$\mu = \Pr(R = 1) \times \mu_1 + \Pr(R = 0) \times \mu_0 = \pi\mu_1 + (1 - \pi)g^{-1}(g(\mu_1) + \Delta)$$

$$\forall a \in R \text{ に対して } g(a) = a \text{ の場合, } \mu = \pi\mu_1 + (1 - \pi)(\mu_1 + \Delta)$$

- ある範囲の感度パラメータ $\Delta$ に対して, MNARのもとでの感度解析の一つとなる.

$$\begin{cases} \Delta = 0 \Leftrightarrow \mu = \mu_1 \Leftrightarrow MAR \\ \Delta \neq 0 \Leftrightarrow \mu \neq \mu_1 \Leftrightarrow MNAR \end{cases}$$

欠測メカニズムMARの仮定から乖離した  $\mu$ にどの程度影響が出るか調査

### 【PMMの枠組み】

- 観測されている同じ測定時点のデータ $Y_0$ をもち, 未観測の人と観測されている人の $Y_1$ の分布を特定

$$g(E(Y_1 \setminus Y_0, R = 0)) = g(E(Y_1 \setminus Y_0, R = 1)) + \Delta$$

関数 $g$ 及び $\Delta$ を含めたモデルを規定が必要. ここでは単純な回帰モデルを示す.

## ベースライン調整

- ここでは単純な回帰モデルの場合を示す.

$E(Y_1 \mid Y_0, R = 1) = \beta_0 + \beta_1 Y_0$  のモデルを仮定すると,

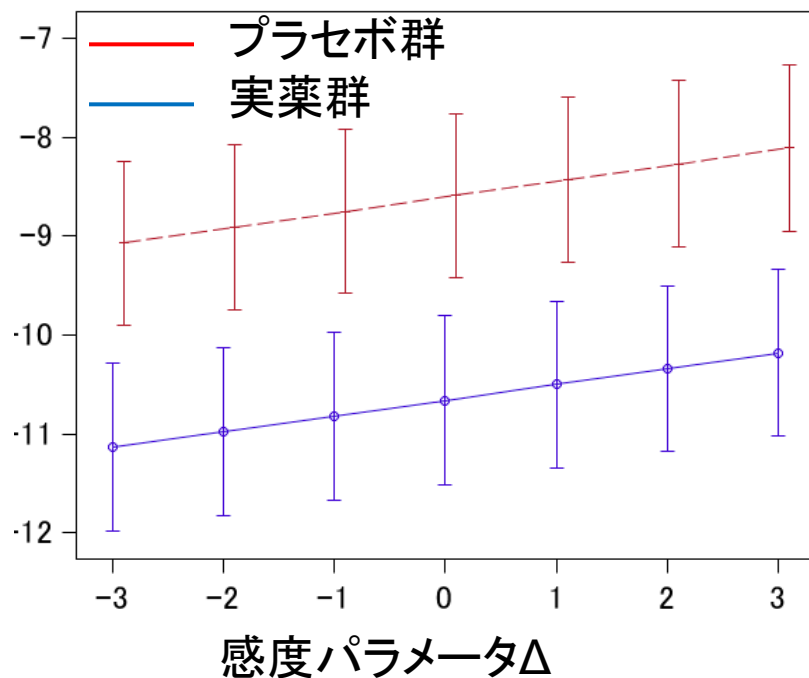
$$E(Y_1 \mid Y_0, R = 0) = E(Y_1 \mid Y_0, R = 1) + \Delta = \beta_0 + \beta_1 Y_0 + \Delta$$

- 欠測データの分布の平均は回帰予測の標本平均により推定
- 得られた $\hat{\beta}_0, \hat{\beta}_1$  は $R=1$ において,  $Y_0$ を与えたうえでの $Y_1$ への回帰により得られた推定値

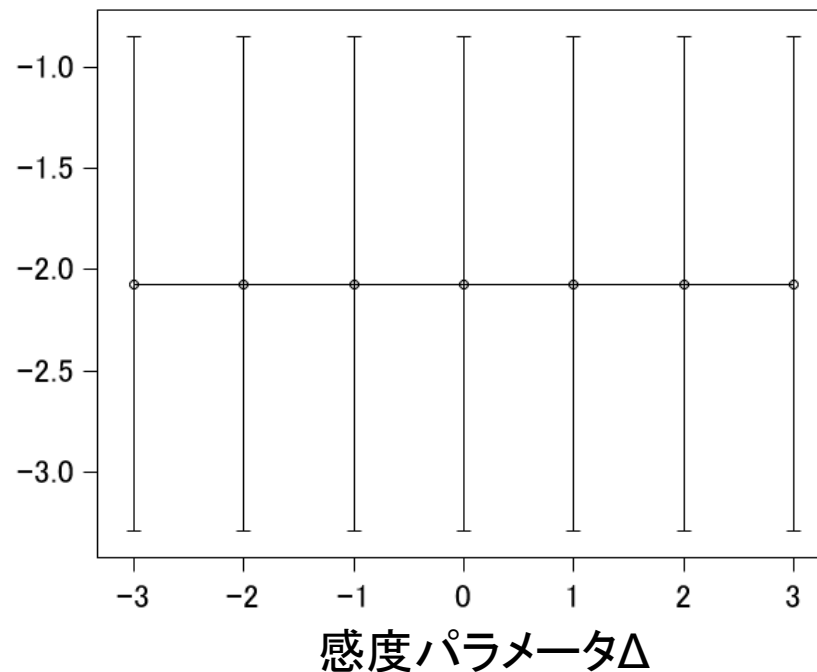


# シミュレーションデータを用いた解析例 ～PMM(NFMV-NCMV+ $\Delta$ )による感度分析の結果～

## □ 各群の点推定値±SE



## □ 群間の差の点推定値±SE





## SMとPMMの感度分析の違い

SM

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

脱落確率のモデルに対する  
感度を見る

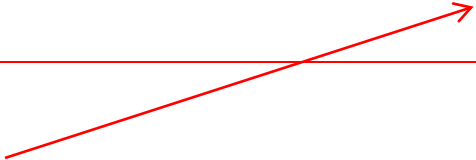
PMM

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

欠測データのモデルに対する  
感度を見る

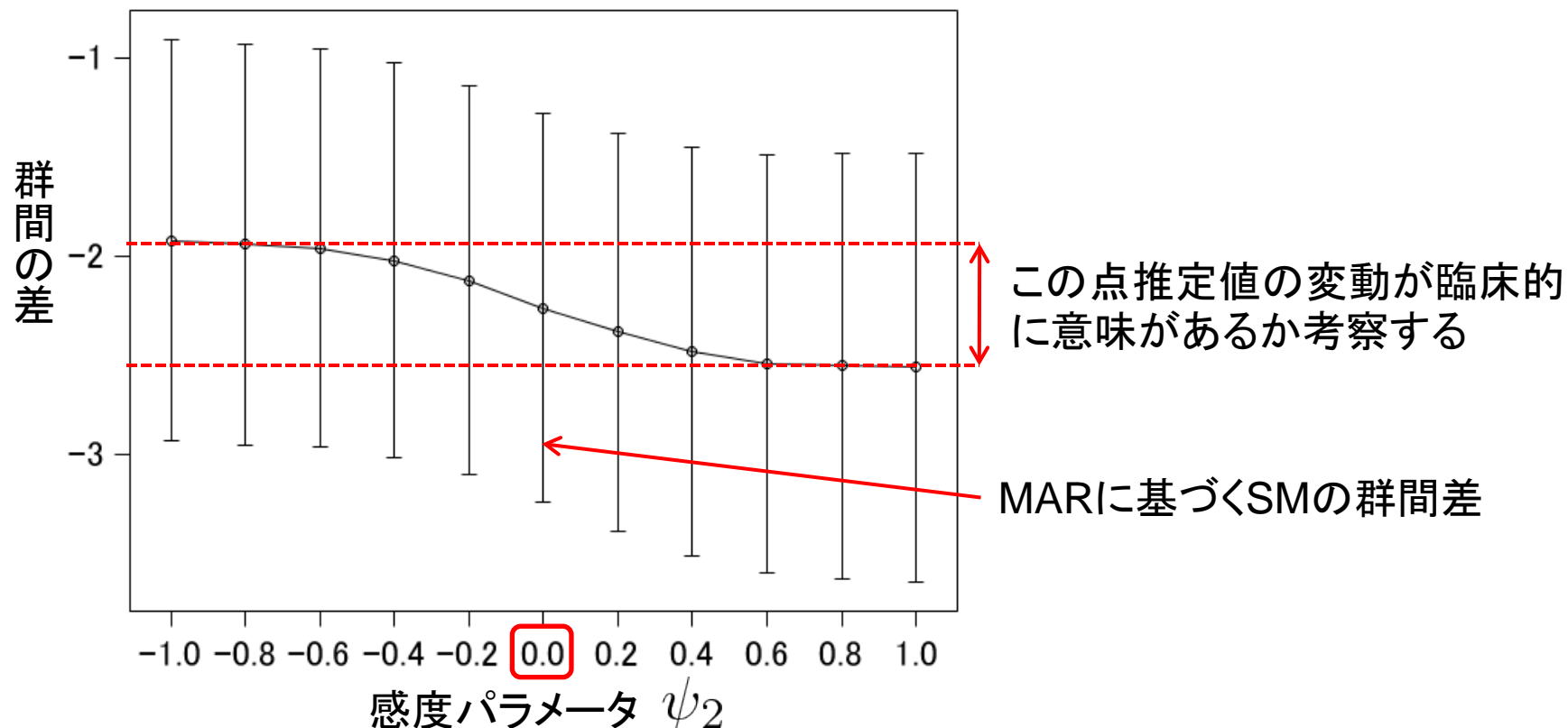
## 本感度分析の結果の考察方法

### □ 欠測メカニズムに対する感度分析を含む解析手順の提案(前スライドより)

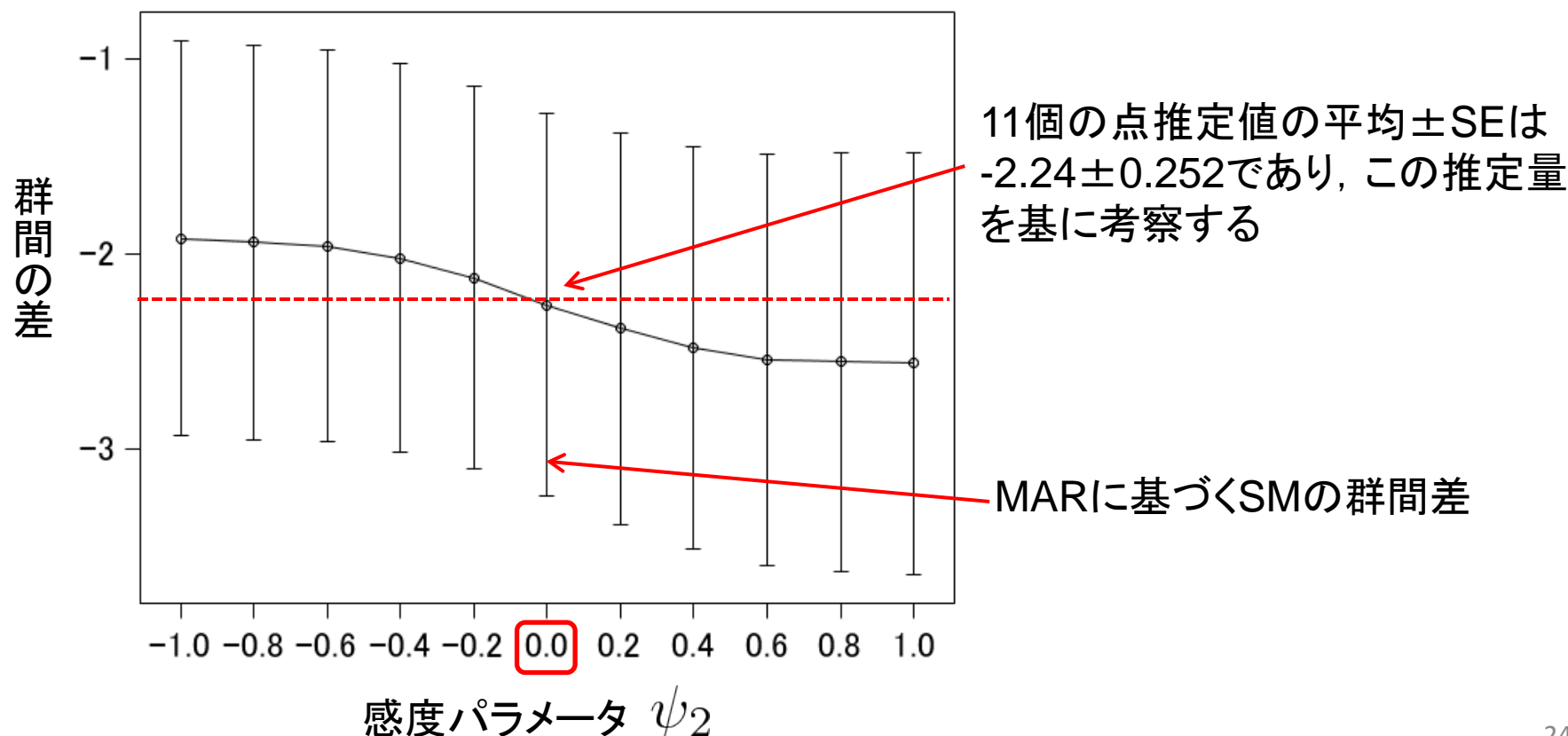
1. 主解析はMARを仮定した解析
  2. 感度分析として, MNARの仮定の下で解析
  3. MARを仮定した解析の結果の頑健性を確認し, 試験結果の妥当性を述べる
- 

□ **NRC (2010)**では3種類の頑健性の確認方法が提案されている。(あくまで一案であり, 今後も研究が必要)

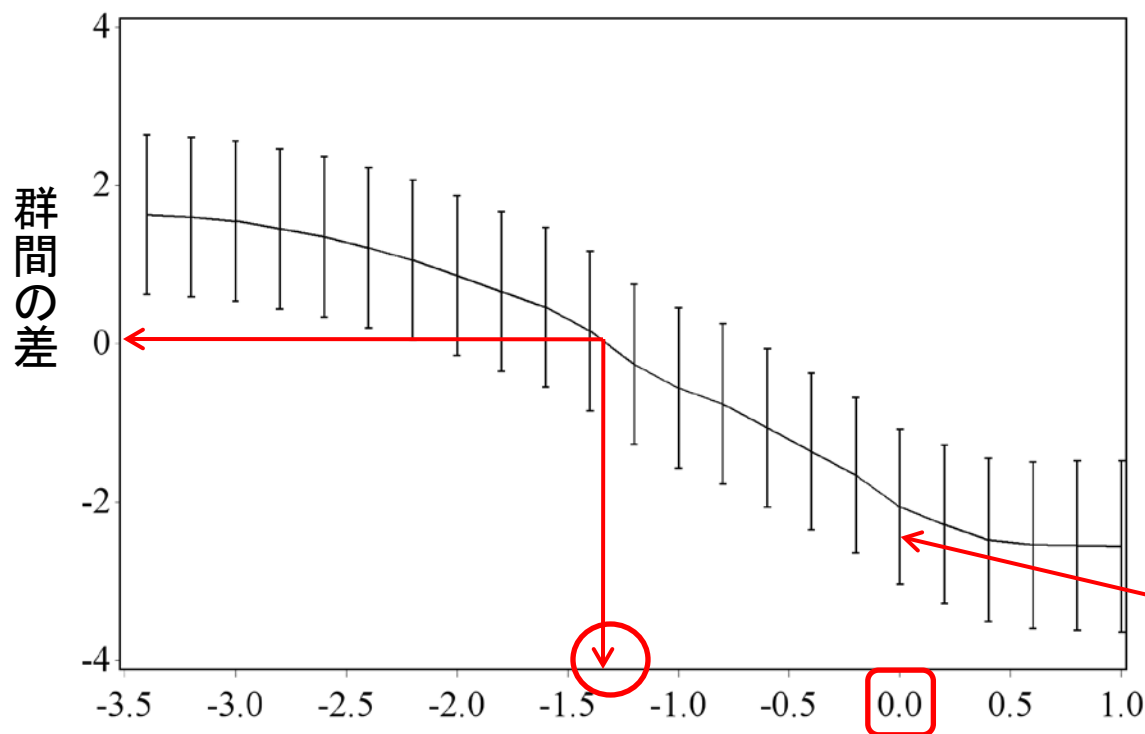
- ①感度パラメータに関して妥当な範囲(下限値～上限値)を指定し, この範囲内で点推定や95%信頼区間を算出  
(ここでの95%信頼区間は, 標本のばらつき+モデルの不確実性におけるばらつきを意味しているため, 95% Confidence Region と呼ばれることもある.)



- ②感度パラメータに関して妥当な範囲(下限値～上限値)を指定し、複数の点推定値に対して平均やSEを求め、一つの結果にまとめる。



- ③MARに基づく推測を行い, MARの結果が変わるような感度パラメータの値を特定すること. もし, この値が妥当な範囲に含まれているならば, MARの結果が疑わしい可能性がある.



群間の差が0になる感度パラメータ $\psi_2$ の値が科学的に妥当な範囲であるか考察する

MARに基づくSMの群間差

感度パラメータ  $\psi_2$

☆NRC(2010)では、以下のコメントも残している

## 主解析及び感度分析の解析結果のどちらを重視すべきか？

- 極端な仮定をおいた場合の感度分析の結果はあまり重視されないが、主解析での仮定と同等の範囲内で行われた感度分析の結果は重視される。
- 主解析の結果が感度分析の結果と反対になることがあったとしても、そのときの感度分析の仮定が極端なものであるならば、主解析の結果を支持するのは合理的

=>極端な仮定ではなく、妥当な仮定の範囲で主解析の結果が支持されていることが重要



## 参考文献

- National Research Council (NRC). The Prevention and Treatment of Missing Data in Clinical Trials. Washington, DC: The National Academies Press, 2010
- Craig H. Mallinckrodt (2013), Preventing and Treating Missing Data in Longitudinal Clinical Trials: A Practical Guide, Cambridge University Press
- Bohdana Ratitch, Michael O'Kelly, and Robert Tosiello, Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models, Pharmaceutical. Statistics. 2013, 12, 337-347.
- SAS macro. missingdata.org.uk .  
[http://missingdata.lshtm.ac.uk/index.php?view=category&id=61%3Amnar-methods&option=com\\_content&Itemid=137](http://missingdata.lshtm.ac.uk/index.php?view=category&id=61%3Amnar-methods&option=com_content&Itemid=137)