

博士論文

欠測を伴う経時的測定データに対する  
MMRM (Mixed-effects Model Repeated Measures)  
における小標本下での推測手法

2019 年 9 月 20 日

総合研究大学院大学  
複合科学研究科統計科学専攻

右京 芳文

# 目次

1. 序文.....	5
1.1 経時測定データと欠測.....	5
1.2 小標本の経時測定データ解析における問題点 .....	6
2. MMRM.....	8
2.1 経時測定データと MMRM.....	8
2.2 検定問題の設定.....	8
2.3 最尤法と制限付き最尤法 .....	9
2.3.1 最尤法.....	9
2.3.2 制限付き最尤法 .....	9
2.4 Kenward and Roger 法 .....	10
3. Bartlett 補正法とブートストラップ検定に基づく近似推測手法.....	13
3.1 背景と目的.....	13
3.2 尤度比検定 .....	14
3.3 ブートストラップ法を用いた Bartlett 補正 .....	14
3.4 ブートストラップ検定 .....	15
3.5 シミュレーション実験.....	16
3.5.1 シナリオ .....	16
3.5.2 相関構造 .....	18
3.5.3 欠測メカニズム .....	19
3.5.4 評価方法.....	21
3.5.5 従来法に対する提案法の基礎的挙動 .....	21
3.5.6 $N = 10$ (1 群 5 例) .....	23
3.5.7 $N = 20$ (1 群 10 例) .....	26
3.5.8 $N = 40$ (1 群 20 例) .....	35
3.6 補足的検討 .....	43
3.6.1 リサンプリング回数に関する検討.....	43
3.6.2 ブートストラップ法による Bartlett 補正項の推定値 .....	45
3.6.3 Kenward and Roger 法の性能に関する検討.....	51
3.6.4 収束に関する検討.....	71
3.7 実データへの適用 .....	72
3.7.1 産後うつ臨床試験.....	72
3.7.2 認知行動療法の臨床試験.....	80
3.7.3 クロスオーバー試験.....	87
3.8 考察 .....	90
3.8.1 欠測の単調性/非単調性 .....	90

3.8.2 KR 法が過度に保守的となる原因 .....	90
3.8.3 $t$ 検定の MAR 下における性能.....	92
3.8.4 リサンプリング回数.....	92
3.8.5 ブートストラップ法による Bartlett 補正項の推定値 .....	93
3.8.6 収束条件 .....	93
3.8.7 提案法の限界.....	93
3.8.8 結論 .....	94
4. 並べ替え検定による推測手法.....	95
4.1 背景と目的 .....	95
4.2 並び替え検定による提案法.....	96
4.2.1 並べ替え検定による推測手法における記法 .....	96
4.2.2 既知のコンセプト.....	97
4.2.3 既知のコンセプトに対する問題点 .....	97
4.2.4 方法 1：投与群指示変数の並べ替え .....	97
4.2.5 方法 2：重み付き残差の並べ替え.....	99
4.3 シミュレーション実験.....	100
4.3.1 シナリオ .....	100
4.3.2 相関構造 .....	100
4.3.3 欠測メカニズム .....	100
4.3.4 評価方法 .....	100
4.3.5 従来法に対する提案法の基礎的挙動 .....	101
4.3.6 $N = 10$ （1 群 5 例） .....	103
4.3.7 $N = 20$ （1 群 10 例） .....	105
4.3.8 $N = 40$ （1 群 20 例） .....	112
4.4 実データへの適用 .....	112
4.5 考察 .....	116
4.5.1 並べ替え検定とパラメトリックブートストラップの比較.....	116
4.5.2 計算の効率性.....	116
4.5.3 調整する共変量が多い場合の性能.....	116
4.5.4 結論 .....	116
5. 結論.....	118
6. 謝辞.....	120
7. 参考文献.....	121
8. 付録.....	127
8.1 シミュレーションデータの要約 .....	127
8.2 ブートストラップ法による Bartlett 補正項の推定値.....	128
8.2.1 Bartlett 補正項推定値のヒストグラム： $N = 20$ （1 群 10 例） .....	129
8.2.2 Bartlett 補正項推定値のヒストグラム： $N = 40$ （1 群 20 例） .....	132

8.3 Kenward and Roger 法の性能に関する検討 .....	135
8.3.1 Type I error rate 及び検出力.....	135
8.3.2 収束割合 .....	143
8.3.3 （固定効果の）標準誤差.....	153
8.3.4 （ $t$ 検定の）自由度 .....	161
8.4 産後うつ臨床試験個別値一覧 .....	175
8.5 認知行動療法の臨床試験個別値一覧 .....	176



# 1. 序文

## 1.1 経時測定データと欠測

新医薬品及び新治療に関する臨床試験や一般的な臨床研究・疫学研究では、経時的に評価項目を測定するいわゆる経時測定データとなっていることが多い。同一個人内で繰り返し測定されるデータは、一般には相関しているため独立ではない。そのため、経時測定データをモデル化して解析を行う場合では、相関を考慮した解析方法を適用する必要がある。また、経時測定データでは複数回の測定が実施されるため、必ずと言ってよいほど欠測が発生する。したがって、興味ある治療効果を推定する時点として多くの臨床試験・臨床研究で主要評価時点として採用されている最終測定時点では、欠測の影響により研究開始時点で想定していたよりも少ないデータ数しか得られていないことが多い。欠測により情報量が減少すると想定していた検出力が得られないだけでなく、治療効果の推定精度も低下するため、問題である (Hogan et al. 2004)。

こうした状況の中で、新医薬品の審査を担当する欧米の規制当局から 2010 年に相次いで欠測データの取り扱いに関するガイドラインや報告書が公布された (European Medicines Agency 2010; National Research Council 2010; Ratitch et al. 2013)。NRC (National Research Council; 全米学術研究会議) の報告書は、正確には米国食品医薬品局 (Food and Drug Administration; FDA) が直接公布しているガイドラインではないが、FDA が全米学術研究会議に検討を依頼して交付された文書であり、参考とされることが多い。いずれのガイドラインも発生してしまった欠測データへの解析手法のみならず、臨床試験の計画段階から欠測データの発生頻度を最小にするための手段を講じるべきであると提言している。ガイドラインや報告書の公布後、生物統計の研究者の中で欠測に対して様々な議論が展開された (Little et al. 2012a; Little et al. 2012b; O'Neill and Temple 2012)。日本においても、製薬企業の業界団体である日本製薬工業協会が学術調査報告書を公開しており欠測データに関する関心が高まった (日本製薬工業協会, 2014)。

計画時に欠測データが発生しないように十分考慮すべきであることは言うまでもないが、実際には経時測定データにおいて欠測データの発生を回避することはできない。Rubin (1976) は経時測定データにおける欠測データの発生メカニズムを 3 種類に分類することを提案した。欠測が完全にランダムに発生しており観測データ及び非観測データに依存していない場合を、完全にランダムな欠測 (Missing completely at random; MCAR)、欠測が観測データのみ依存して発生する場合をランダムな欠測 (Missing at random; MAR)、欠測が非観測データにも依存する場合をランダムでない欠測 (Missing not at random; MNAR) と定義した。

欠測データへの対処法の 1 つに、最後に観測された値で補完する方法 (Last observation carried forward; LOCF 法) がある (Gillings and Koch 1991)。ある対象者のデータが欠測した場合、その対象者の最後の観測値を欠測データに代入し、完全なデータセットとみなして解析する方法である。多くの臨床試験で LOCF 法を使って欠測データを補完した擬似的完全データに対して、共変量を調整した線形モデルによる解析等を適用した解析をこれまで主解析として標準的に使用してきた。しかし、近年、主解析に LOCF 法を使用することに対する批判が急速に高まっている。実質的に FDA の欠測データに関するガイドラインとも評される NRC レポート (National Research Council 2010) は、LOCF 法の科学的正当化が難しいため、主解析として LOCF 法を使用すべきではないと述べている。また、欧州医薬品庁が公布したガイドライン

(European Medicines Agency 2010) でも、LOCF 法が群間差を不偏に推定できる状況はかなり限定的であり、必ずしも保守的な結果を与えるものではないとしている。また、同ガイドラインは、LOCF 法は治療効果の指標における標準誤差を過小推定する可能性があり、それに伴い治療効果の信頼区間が狭くなる点を問題視している。ただし、事例を挙げて LOCF 法が保守的な推定結果を与えることがある点についても触れている。

LOCF 法は、欠測メカニズムが MCAR で、かつ対象者の反応が最終観測時点から試験終了時点まで一定であるという仮定の下で妥当だが、これは極めて強い仮定である。例えば、症状が改善、あるいは治癒したことを理由に脱落した場合のように、LOCF 解析の仮定が合理的と思える状況も存在するかもしれないが、一般的にはそのような仮定を保証するだけの生物学的根拠は存在しないし、そもそもデータから仮定を検証することもできない。実際、これまでに多くの研究者が LOCF 法を使用することの問題点を指摘してきた (Mallinckrodt et al. 2001a, 2001b, 2004a, 2004b; Barnes et al. 2008; DeSouza et al. 2009; Lane 2008; Liu and Gould 2002; Liu et al. 2006; Siddiqui et al. 2009)。ほとんどの臨床試験では、LOCF 法を使用する上での前提を正当化することは困難なため、主解析に LOCF 法を使用するのは避けた方が無難とされている。

## 1.2 小標本の経時測定データ解析における問題点

一方で、LOCF 法よりも Mixed-effects models for repeated measures (MMRM) を用いた解析法が連続量を対象とした主解析の方法として妥当であるとされ、近年その利用が急速に広がっている (Mallinckrodt et al. 2001a, 2004a, 2003, 2004c, 2008; Barnes et al. 2008; Lane 2008)。MMRM は尤度に基づく方法であり、欠測値に補完値を代入したりせずに観測された値をそのまま解析に利用する。結果変数の欠測メカニズムが MAR である場合に、直接尤度法により治療効果等の興味あるパラメータの一致推定量が得られることから、臨床試験の主要な解析として採用されるようになってきている (Mallinckrodt et al. 2004c; Mallinckrodt et al. 2008)。

MMRM は線形混合モデル (Laird and Ware 1982; Diggle 2002; Fitzmaurice, Laird and Ware 2012) の一種であり、経時測定データにおける時点間の相関を考慮した結果変数の分散共分散構造を直接的にモデル化している (Mallinckrodt et al. 2008)。この場合、変量効果は結果変数の周辺分布の分散共分散行列に含まれた形でパラメタライズされている。MMRM の利点の 1 つとして、時点間相関構造を柔軟にモデリング可能である点が挙げられる。

MMRM のスタンダードな推定方法は、先に述べた最尤法もしくは制限付き最尤法 (Restricted maximum likelihood: REML, Patterson and Thompson 1971) である。いずれの方法も尤度原理に基づくため、MAR の欠測メカニズムのもとでは、推定した固定効果パラメータは一貫性及び漸近正規性を有する。しかしながら、サンプルサイズが小さい下で最尤法ではバイアスが大きくなるため、より妥当な推測結果を一般的に与える制限付き最尤法が用いられることが多い。

また、MMRM の推測手法は大標本近似に基づいているため、小標本下ではその妥当性が破綻してしまう可能性がある。サンプルサイズが十分に大きくない下では、大標本理論に基づいて構成される検定・信頼区間の近似が十分に成り立たず、第 1 種の過誤確率を名目水準以下に保持できないなどの問題が知られている (Schluchter and Elashoff 1990; Kenward and Roger 1997)。欠測により観測データの情報量が小さくなると、大標本近似はさらに悪化するため、推測の妥当性の問題はますます深刻なものとなる。Gosho et al. (2017) は、欠測を伴う小標本下で

MMRM により固定効果パラメータの推測をした場合に妥当性が失われてしまうことを示した。特に、試験終了時点における治療効果の検定では、第 1 種の過誤確率が名義水準以下に保持できなくなった。

希少疾患を対象とした臨床試験及び疫学研究では、欠測を伴う小標本の経時測定データを解析対象とせざるを得ない状況が容易に想定される。一般的に普及している経時測定データの標準的な解析方法は大標本近似に基づく手法であり、小標本ではその統計的推測の妥当性が失われる。また、経時測定データにほぼ確実に発生する欠測により、統計的推測の妥当性はさらに失われることとなる。近年、新医薬品の開発対象や観察研究の研究対象も common disease から rare disease に移行しており、実務で遭遇する問題に対して実践的な解決方法が求められている。

本論文の構成は、第 2 章で MMRM を説明し、第 3 章で Bartlett 補正法とブートストラップ検定に基づく近似推測手法に関する研究を紹介し、第 4 章で並べ替え検定による推測手法に関する研究を紹介する。最後に第 5 章で結論を述べる。

## 2. MMRM

### 2.1 経時測定データと MMRM

2 群比較のランダム化臨床試験を想定する。連続量の結果変数が  $n$  回繰り返し測定される状況を考える ( $j = 1, 2, \dots, n$ )。2 群の合計被験者数を  $N$  とし、被験者  $i$  の測定回数を  $n_i$  とおく ( $j = 1, 2, \dots, n_i, n_i \leq n$ )。主要評価項目は、最終測定時点である  $n$  時点目の 2 群の治療効果の群間差とする。本研究では、単調な欠測を仮定する。単調な欠測とは、欠測が 1 回発生したら以降の測定値が全て欠測となる状況を指す。

MMRM は線形混合モデル (Laird and Ware 1982; Verbeke and Molenberghs 2000) の特別な形であり、以下の式で表現される。

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i \quad (1)$$

$Y_i$  は、被験者  $i$  における  $n_i \times 1$  の繰り返し測定データベクトルを表す。 $\beta$  は、 $p \times 1$  の固定効果パラメータベクトルである。 $b_i$  は、変数効果 (個人間変動) を表す  $q \times 1$  のベクトルで多変量正規分布  $MVN(0, D)$  に従って分布する。ここで、 $D$  は  $(i, j)$  成分  $d_{ij}$  に対して  $d_{ij} = d_{ji}$  となる  $q \times q$  の分散共分散行列である。 $X_i$  と  $Z_i$  は、それぞれ固定効果及び変数効果に対応する  $n_i \times p$  及び  $n_i \times q$  のデザイン行列である。 $\varepsilon_i$  は、 $n_i \times 1$  の誤差 (個人内変動) ベクトルであり、それぞれ独立に多変量正規分布  $MVN(0, \Sigma_i)$  の分布に従う。 $\Sigma_i$  は、全時点のデータが観測された分散共分散行列の  $(1, 1)$  成分から  $(n_i, n_i)$  成分までを取り出した  $n_i \times n_i$  の対角行列であり分散共分散行列である。変数効果  $b_i$  と誤差  $\varepsilon_i$  は独立であり、異なる個人間のデータも全て独立と仮定する。

MMRM では、変数効果  $b_i$  に対して周辺化した結果変数  $Y_i$  の周辺分布を考える。 $Y_i$  の周辺分布は、以下のようになる。

$$Y_i \sim MVN(X_i\beta, V_i), \quad V_i = Z_i D Z_i^T + \Sigma_i \quad (2)$$

MMRM では、式(1)の中で変数効果を明示的に与えて解析するというよりは、式(2)で示した  $Y_i$  の周辺分布の分散共分散行列  $V_i$  に変数効果を含めてパラメタライズし、結果変数の共分散構造の一部としてモデル化する (Mallinckrodt et al. 2008)。つまり、一般的な MMRM では周辺化した分布に基づきパラメータ推定するため、変数効果  $b_i$  を推定することはない。この点が、他の線形混合モデルに基づく解析と異なる特徴の 1 つである。さらに結果変数  $Y_i$  の時点間相関構造が  $V_i$  で規定できるため、 $V_i$  を複合対称 (Compound Symmetry: CS)、自己回帰 (Auto-regression: AR)、無構造 (Unstructured: UN) などで直接モデル化できる利点がある。

また、MMRM は欠測メカニズムが MAR の下で妥当となる直接尤度法によりパラメータ推定を行う (Verbeke and Molenberghs 2000)。欠測メカニズムはデータから検証することはできないが、欠測を含んだデータ解析において可能な限り妥当な手法を適用すべきである。欧州医薬品庁からのガイドライン (European Medicines Agency 2010) や NRC レポート (National Research Council 2010) 公布後、MMRM の利用が増えている背景にはこうした理由もあると推測される。

### 2.2 検定問題の設定

ここで、臨床試験の主要評価項目を MMRM で検定することを考える。主要評価項目は臨床試験で示すべき仮説と通常対応している。本研究が想定する経時測定データ型の臨床試験の場合、繰り返し測定の最終測定時点での群間差の評価 (検定) が主要評価項目の解析となることが多い。そこで、以下に示す検定問題を設定する。

固定効果パラメータ  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$  のうち、一般性を失うことなく、興味ある固定効果パラメータ  $\beta_1$  の検定を以下のように定義する。

帰無仮説  $H_0: \beta_1 = \beta_1^{null}$

対立仮説  $H_1: \beta_1 \neq \beta_1^{null}$

ここで、 $\boldsymbol{\beta}_c = (\beta_2, \beta_3, \dots, \beta_p)^T$  とし、固定効果パラメータ  $\boldsymbol{\beta}$  の分散共分散成分を  $\mathbf{v}$  とおく。また  $\mathbf{v}$  は、周辺化した結果変数  $\mathbf{Y}_i$  の分散共分散行列  $\mathbf{V}_i$  を構成成分とするパラメータベクトルである。結果変数  $\mathbf{Y}_i$  の確率密度関数は、

$$f(\mathbf{y}_i) = \frac{1}{(2\pi)^{\frac{n_i}{2}} \cdot |\mathbf{V}_i|^{\frac{1}{2}}} \exp \left\{ -\frac{(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})}{2} \right\}$$

であり、尤度関数  $L(\beta_1, \boldsymbol{\beta}_c, \mathbf{v})$  は、

$$L(\beta_1, \boldsymbol{\beta}_c, \mathbf{v}) = \prod_{i=1}^N f(\mathbf{y}_i)$$

となる。

多変量正規分布の対数尤度関数  $l_{ML}$  を用いて、最尤法により  $l_{ML}(\beta_1, \boldsymbol{\beta}_c, \mathbf{v})$  を最大化する推定値  $(\hat{\beta}_1, \hat{\boldsymbol{\beta}}_c, \hat{\mathbf{v}})$  を得ることができる。実際には、分散成分  $\mathbf{v}$  について固定効果の自由度を考慮した推定値を得るために制限付き最尤法 (Restricted maximum likelihood: REML) が利用されることが一般的である (Patterson and Thompson 1971)。制限付き最尤法の長所については、Harville や McCulloch らが言及しており、実務上も制限付き最尤法の使用が推奨されている (Harville 1977; Mallinckrodt et al. 2008; McCulloch et al. 2008)。

## 2.3 最尤法と制限付き最尤法

### 2.3.1 最尤法

多変量正規分布の尤度関数を用いて、興味あるパラメータ (本研究では薬剤の固定効果) を推定する。最尤法により  $l_{ML}(\beta_1, \boldsymbol{\beta}_c, \mathbf{v})$  を最大化する推定値  $(\hat{\beta}_1, \hat{\boldsymbol{\beta}}_c, \hat{\mathbf{v}})$  を得ることができる。

$$l_{ML}(\beta_1, \boldsymbol{\beta}_c, \mathbf{v}) = -\frac{1}{2} \sum_{i=1}^N \{ \log |\mathbf{V}_i| + (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) + n_i \log 2\pi \} \quad (3)$$

### 2.3.2 制限付き最尤法

多変量正規分布の尤度関数を用いて、最尤法と同じ要領で興味あるパラメータ (本研究では薬剤の固定効果) を推定する。制限付き最尤法では、ある一定の条件を満たした場合に分散成分  $\mathbf{v}$  の不偏推定量を得ることができる (Patterson and Thompson 1971)。制限付き最尤法における尤度関数を以下に示す。

$$l_{REML}(\beta_1, \boldsymbol{\beta}_c, \mathbf{v}) = l_{ML}(\beta_1, \boldsymbol{\beta}_c, \mathbf{v}) - \frac{1}{2} \log \left| \sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right|$$

実際には  $l_{ML}(\beta_1, \boldsymbol{\beta}_c, \mathbf{v})$  から  $\boldsymbol{\beta}$  の推定値  $\hat{\boldsymbol{\beta}}$  を得て  $l_{REML}(\beta_1, \boldsymbol{\beta}_c, \mathbf{v})$  の  $\boldsymbol{\beta}$  にプラグインした上で、 $l_{REML}(\beta_1, \boldsymbol{\beta}_c, \mathbf{v})$  を最大化する  $\mathbf{v}$  を求める。

REML 法では、セクション 2.2 で示した仮説検定を Wald-type の検定統計量を構成することにより検定する。Wald-type の検定統計量は  $\hat{\beta}_1^2 / \widehat{Var}(\hat{\beta})|_{\beta=\hat{\beta}_1}$  で算出可能であり、漸近的に自由度 1

のカイ 2 乗分布に従う。カイ 2 乗分布を  $t$  分布または  $F$  分布の近似で置換する。 $t$  分布または  $F$  分布を用いる時に自由度が必要になるが、本研究では、その推定に Containment 法 (Schaalje et al. 2002) を用いた。Containment 法はバランスデータである分割実験の自由度に基づいた方法である。固定効果に対する自由度は固定効果を含む変量効果に依存している。仮に変量効果が特定の固定効果を含んでいない場合、自由度は  $K - \text{rank}(X, Z)$  となる。また、最尤法及び制限付き最尤法を用いて推定した結果、相関構造の誤特定により  $\alpha$  エラーのインフレーションが起こることがある (Mallinckrodt et al. 2004b)。そうでなくても相関構造をデータから検証することはほぼ不可能であるので、恣意性を除外するために特定の相関構造を指定しないことが多い。したがって、MMRM では無構造 (Unstructured) を前提としたモデリングが基本となっている。まずは、相関構造を Unstructured と仮定して解析を実行するが、推定するパラメータ数が多いために収束が得られない場合がある。こうした場合に、収束値を得るために仮定する相関構造をより簡略化した (推定するパラメータ数を減少させた) 相関構造を仮定して解析を実施する手順とすることが多い。

## 2.4 Kenward and Roger 法

平均構造及び相関構造が正しく特定されている場合の固定効果  $\beta$  の推定量の分散を、モデル分散と呼ぶ。モデル分散の推定量は、以下の式で表される。

$$\widehat{\text{Var}}(\hat{\beta}) = \left( \sum_{i=1}^N (X_i^T \hat{V}_i^{-1} X_i) \right)^{-1}$$

相関構造 (共分散行列)  $V_i$  を誤特定した場合、標準誤差が過小評価され、多くの場合 Type I error rate のインフレーションにつながることを報告されている (Mallinckrodt et al. 2004b; Lu and Mehrotra 2010)。

MMRM は尤度に基づくアプローチであり、欠測メカニズムに応じて欠測値を補完したり調整したりする方法ではない。モデルのパラメータは REML で推定され、モデル分散の共分散パラメータは観測データから推定した Fisher 情報行列により推定値を得る。

モデル分散を  $\widehat{\text{Var}}_{\text{MO}}(\hat{\beta})$  とし、以下  $\hat{\Phi}$  と表記することにする、

$$\widehat{\text{Var}}_{\text{MO}}(\hat{\beta}) = \hat{\Phi} = \left( \sum_{i=1}^N (X_i^T \hat{V}_i^{-1} X_i) \right)^{-1}$$

ここで、 $\hat{V}_i$  は  $V_i$  の観測データからの推定値である。

同様にロバスト分散を  $\widehat{\text{Var}}_{\text{RO}}(\hat{\beta})$  と表記すると、

$$\widehat{\text{Var}}_{\text{RO}}(\hat{\beta}) = \left( \sum_{i=1}^N (X_i^T \hat{V}_i^{-1} X_i) \right)^{-1} \sum_{i=1}^N X_i^T \hat{V}_i^{-1} \text{Var}(Y_i) \hat{V}_i^{-1} X_i \left( \sum_{i=1}^N (X_i^T \hat{V}_i^{-1} X_i) \right)^{-1}$$

$\text{Var}(Y_i)$  は通常残差  $\hat{\epsilon}_i$  により推定される。

$$\hat{\epsilon}_i = Y_i - X_i \hat{\beta}$$

したがって、ロバスト分散  $\widehat{\text{Var}}_{\text{RO}}(\hat{\beta})$  は、

$$\widehat{\text{Var}}_{\text{RO}}(\hat{\beta}) = \hat{\Phi}_{\text{RO}} = \left( \sum_{i=1}^N (X_i^T \hat{V}_i^{-1} X_i) \right)^{-1} \sum_{i=1}^N X_i^T \hat{V}_i^{-1} \hat{\epsilon}_i \hat{\epsilon}_i^T \hat{V}_i^{-1} X_i \left( \sum_{i=1}^N (X_i^T \hat{V}_i^{-1} X_i) \right)^{-1}$$

残差の推定値  $\hat{\epsilon}_i$  は過小推定される傾向にあることがよく知られている。すなわち、真値よりも観測値に近い値になってしまう。したがって、分散の式にある  $\text{Var}(Y_i)$  は過小推定されて

しまう傾向にある (Mancl and DeRouen 2001)。

小標本下において  $Var(\mathbf{Y}_i)$  が過小推定されることにより、結果的に  $\widehat{Var}_{RO}(\hat{\boldsymbol{\beta}})$  に発生するバイアスは大きくなる (Mancl and DeRouen 2001)。

小標本下におけるバイアスの問題に対処するため、KR (Kenward and Roger's) 法が提案されている (Kenward and Roger 1997)。KR 法における固定効果パラメータの分散は、 $Var_{KR}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\Phi}} + \boldsymbol{\Lambda}$  のかたちに表示される (Kackar and Harville 1984)。Kenward と Roger は  $\hat{\boldsymbol{\Phi}}$  を  $\boldsymbol{\Phi}(\hat{\boldsymbol{\sigma}})$  の 2 次のテイラー展開で近似し、 $\boldsymbol{\Lambda}$  をバイアス補正したテイラー近似で和をとった式で  $\widehat{Var}_{KR}(\hat{\boldsymbol{\beta}})$  を提案した。KR 法では固定効果  $\boldsymbol{\beta}$  の共分散構造パラメータ推定量と  $t$  分布及び  $F$  分布の分母の自由度を調整する。 $\mathbf{V}_i$  の要素  $\sigma_i (m_i \times 1)$  を  $m_i$  個のパラメータによる関数と仮定する。KR 法による  $\hat{\boldsymbol{\beta}}$  の分散を  $Var_{KR}(\hat{\boldsymbol{\beta}})$  と表記すると、

$$\widehat{Var}_{KR}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\Phi}} + 2\hat{\boldsymbol{\Phi}} \left\{ \sum_{i=1}^N \sum_{j=1}^{m_i} \sum_{l=1}^{m_i} w_{ijl} \left( \mathbf{Q}_{ijl} - \mathbf{P}_{ij} \hat{\boldsymbol{\Phi}} \mathbf{P}_{il} - \frac{1}{4} \mathbf{R}_{ijl} \right) \right\} \hat{\boldsymbol{\Phi}}$$

ここで、

$$\mathbf{P}_{ij} = \mathbf{X}_i^T \frac{\partial \mathbf{V}_i^{-1}}{\partial \hat{\sigma}_j} \mathbf{X}_i, \mathbf{Q}_{ijl} = \mathbf{X}_i^T \frac{\partial \mathbf{V}_i^{-1}}{\partial \hat{\sigma}_j} \mathbf{V}_i \frac{\partial \mathbf{V}_i^{-1}}{\partial \hat{\sigma}_l} \mathbf{X}_i, \mathbf{R}_{ijl} = \mathbf{X}_i^T \mathbf{V}_i^{-1} \frac{\partial^2 \mathbf{V}_i}{\partial \hat{\sigma}_j \partial \hat{\sigma}_l} \mathbf{V}_i^{-1} \mathbf{X}_i$$

$w_{ijl}$  は  $\hat{\sigma}_i$  の  $(j, l)$  成分であり、 $\sigma_i$  の REML 推定量である。

第 2 項が小標本に対するバイアス補正項と解釈できる (Gosho et al. 2017)。

松山・山口 (2001) を参考に本研究に沿って固定効果の検定を考える。固定効果に対する推測に対して、 $t$  統計量と  $F$  統計量が使われる。 $\boldsymbol{\beta}$  は  $(p \times 1)$  のベクトルで、 $\boldsymbol{\beta}$  の線形結合  $\mathbf{h}^T \boldsymbol{\beta}$  の推定に関心があるとする。ただし、 $\mathbf{h}^T \boldsymbol{\beta}$  は推定可能であるとする。ここで  $\mathbf{h}$  は関心のある推定を決定する  $(p \times 1)$  の対比ベクトルである。

仮説は以下の通りである。

$$H_0: \mathbf{h}^T \boldsymbol{\beta} = 0, H_1: \mathbf{h}^T \boldsymbol{\beta} \neq 0$$

仮説に対応する  $t$  検定統計量は、

$$t = \frac{\mathbf{h}^T \boldsymbol{\beta}}{\sqrt{\mathbf{h}^T (\sum_{i=1}^N (\mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i)) \mathbf{h}}} \quad (4)$$

となり、近似的に自由度  $v$  の  $t$  分布に従うことを利用して検定を行う。ここで自由度  $v$  を決定する必要があるが、これは分散  $\mathbf{h}^T (\sum_{i=1}^N (\mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i)) \mathbf{h}$  を推定する際に利用される自由度に等しい。この仮説を拡張した固定効果の一般的な線形仮説は次のようになる。

$$H_0: \mathbf{H}^T \boldsymbol{\beta} = 0, H_1: \mathbf{H}^T \boldsymbol{\beta} \neq 0$$

ここで、 $\mathbf{H}$  は固定効果  $\boldsymbol{\beta}$  に関する線形仮説を決定する  $(p \times q)$  のフルランク ( $\text{rank}(\mathbf{H}) = q$ ) の対比行列である。 $\mathbf{h}_j = (h_{j1}, h_{j2}, \dots, h_{jp})^T$  を行列  $\mathbf{H}$  の  $j$  番目の列ベクトルとし、 $j = 1, 2, \dots, q$  に対して  $\mathbf{h}^T \boldsymbol{\beta}$  が推定可能であると仮定する。

$\mathbf{H}^T \boldsymbol{\beta}$  の推定量として  $\mathbf{H}^T \hat{\boldsymbol{\beta}}$  を用いるとし、 $\mathbf{H}^T \hat{\boldsymbol{\beta}}$  の分散共分散行列を求めると  $\mathbf{H}^T (\sum_{i=1}^N (\mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i))^{-1} \mathbf{H}$  となる。

仮説検定に対応する検定統計量は、

$$F = \frac{(\mathbf{H}^T \hat{\boldsymbol{\beta}})^T (\mathbf{H}^T \sum_{i=1}^N (\mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i) \mathbf{H})^{-1} (\mathbf{H}^T \hat{\boldsymbol{\beta}})}{\text{rank}(\mathbf{H})} \quad (5)$$

となり，近似的に分子の自由度が $\text{rank}(\mathbf{H})$ の $F$ 分布に従う。ただし，分母の自由度は一般に容易に決定することができず，何等かの方法で求める必要がある。 $F$ 検定統計量は，一般的なワルド検定統計量である。 $q = 1$ の場合， $\mathbf{H} \equiv \mathbf{h}$ となり，仮説検定( $H_0: \mathbf{h}^T \boldsymbol{\beta} = 0, H_1: \mathbf{h}^T \boldsymbol{\beta} \neq 0$ )は $H_0: \mathbf{H}^T \boldsymbol{\beta} = 0, H_1: \mathbf{H}^T \boldsymbol{\beta} \neq 0$ の仮説に含まれ， $\text{rank}(\mathbf{H}) = 1$ となることにより $F$ 検定統計量(2)は $t$ 検定統計量(1)の2乗に帰着する。

KR法の基本的な考え方は， $F$ 分布の平均（期待値）と分散を推定し， $F$ 分布のモーメントを一致させるような $F$ 分布の自由度を得ることにある。KR法（Kenward and Roger 1997）の他にも、Kenward and Roger (2)法（Kenward and Roger 2009），Mancl and DeRouenによる方法（Mancl and DeRouen 2001）及びKauermann and Carrollによる方法（Kauermann and Carroll 2001）も提案されている。



### 3. Bartlett 補正法とブートストラップ検定に基づく近似推測手法

#### 3.1 背景と目的

小標本下で大標本近似の妥当性が破綻してしまう問題に対して、従来の線形混合モデルに関して複数の先行研究がなされてきた。1 つは高次漸近理論を適用する方法である。Zucker et al. (2000)は、尤度比検定 (Likelihood ratio test: LRT) と Bartlett 補正 (Bartlett 1937) や Cox-Reid (Cox and Reid 1987) の調整尤度を組み合わせた方法を提案した。Lyons と Peters (Lyons and Peters 2000) や Guolo et al. (2006)は、Barndorff-Nielsen (Barndorff-Nielsen 1986) により導入された Skovgaard の修正符号付対数尤度比 (Skovgaard 1996) を適用した高次漸近論的アプローチを提案した。Stein et al. (2014)は、Severini の近似方法 (Severini 1998) に基づいた Barndorff-Nielsen (Barndorff-Nielsen 1983) のアプローチを用いたプロファイル尤度による方法を検討した。Stein et al. (2014)による改善法は、シミュレーション研究で第 1 種の過誤確率を制御する点に関して比較的良好な性能を示したが、従来の尤度比検定との比較のみを示すに留まった。さらに、Stein et al. (2014)の方法は、対象とする解析データごとに対数尤度関数の高次微分が必要であり複雑な解析的計算を要するものであった。特に経時測定データに MMRM を適用する場合、モデルの誤特定を回避するため結果変数の周辺分布に関する分散共分散構造に複雑な構造を仮定することが多い。医薬品開発の臨床試験では、特に解析者の恣意性を排除することを目的に無構造 (unstructured) を仮定することが一般的である。このような条件下で、Stein et al. (2014)の提案法はケースバイケースでの複雑な計算が必要となり、臨床試験の実務に応用する上では困難を伴う。Stein et al. (2014)は、ブートストラップ法を用いた近似による推測方法についても言及している。しかしながら、Stein et al. (2014)の提案法では適用することができる変量効果モデルの構造も限定されており、結果変数間に複雑な相関構造を仮定する状況は検討されていない。さらに、シミュレーションによる評価では、欠測を伴う経時測定データに関して既存法と比較した検討はされておらず、近年の臨床試験のスタンダードな方法のひとつとして急速に普及した MMRM の解析における実践的に有用な方法は十分に確立していない。

本研究では、欠測を伴う小標本の経時測定データに対して MMRM におけるブートストラップ法を用いた高次漸近理論に基づく Bartlett 補正項の近似推測手法を 2 種類提案する。先行研究で提案されている解析的な計算法では実践的には困難が伴う。これを回避するために、ブートストラップ法による近似推測手法を適用する (Tibshirani and Efron 1993; Cox and Hinkley 1974; Davison and Hinkley 1997; Cordeiro and Cribari-Neto 2014)。1 つ目の方法は、Bartlett 補正項 (Bartlett 1937) をブートストラップ法により推定する方法 (Cordeiro and Cribari-Neto 2014) である。本提案法は、統一した計算手順により複雑な線形モデルや欠測パターンに関係なく適用可能であり、対象データに応じたケースバイケースの複雑な解析的計算を回避することができる有用な方法である。2 つ目の方法は、ブートストラップ法により推定された帰無仮説下の経験分布を用いたモンテカルロ検定である。本法は簡便なアプローチであり、欠測を伴う小標本である状況下で有用であることが広く知られている。両方法に共通するリサンプリングの手順は、結果変数に欠測が発生していても適用可能である。小標本かつ欠測を伴うより実践的な状況下の臨床試験における経時測定データを模したシミュレーションにより統計的性能を評価した。現時点の小標本の経時測定データの標準的な解析方法として、結果変数の分散共分散構造に無構造を仮定した Kenward and Roger 法 (Kenward and Roger 1997) が広く普及している。

Kenward and Roger 法は、固定効果パラメータを REML 法により推定し、分散の推定値と  $t$  検定における自由度を補正する方法である。シミュレーションでは Kenward and Roger 法との比較も行った。

適用事例として、Lancet で発表された産後うつ臨床試験データ (Gregoire et al. 1996) とうつ症状に対して認知行動療法の効果を検討した臨床試験データ (Proudfoot et al. 2003) の実データに本提案法を適用し有用性を検討した。さらに、本提案法は新医薬品の開発時に異なる製剤の同等性を評価するために実施される生物学的同等性試験で頻繁に適用されるクロスオーバー試験データにも適用可能であり、その有用性を併せて検討した。

### 3.2 尤度比検定

はじめに MMRM による尤度比検定を導入する。固定効果パラメータ  $\beta_1$  に関する尤度比検定統計量は、以下のようになる。

$$T(\beta_1^{null}) = -2\{l(\beta_1^{null}, \tilde{\beta}_c, \tilde{v}) - l(\hat{\beta}_1, \hat{\beta}_c, \hat{v})\}$$

ここで、 $(\tilde{\beta}_c, \tilde{v})$  は帰無仮説下で推定された制約付きの最尤推定値である。 $(\hat{\beta}_1, \hat{\beta}_c, \hat{v})$  は対立仮説下で推定された最尤推定値となる。帰無仮説  $H_0: \beta_1 = \beta_1^{null}$  の下で、尤度比検定統計量  $T(\beta_1^{null})$  は自由度 1 のカイ 2 乗分布に従う。仮説に示したように帰無仮説は対立仮説にネストされた構造となっているため、本研究の尤度比検定統計量は最尤推定値によってのみ算出する (Cnaan et al. 1997; Verbeke 1997; Pinheiro and Bates 2000)。

### 3.3 ブートストラップ法を用いた Bartlett 補正

対象とする解析データが小標本である場合に、尤度比検定統計量  $T(\beta_1^{null})$  のカイ 2 乗分布に対する大標本近似は理想的な近似にはなっていない。これまでにいくつかの高次漸近理論に基づくアプローチが提案されてきており、その一つに Bartlett 補正がある (Bartlett 1937)。小標本の状況下でも  $T(\beta_1^{null})$  の従う分布がカイ 2 乗分布に近似するように、Bartlett は尤度比検定統計量の補正方法を提案した。Bartlett の補正方法は、補正項で従来の尤度比検定統計量を割ることにより、参照するカイ 2 乗分布への近似を改善しようとするものである。特に補正項は Bartlett 補正項  $\xi$  と呼ばれ、尤度比検定統計量の 1 次モーメントの推定値になっている。すなわち、次のような式で表される。

$$\xi = \hat{E}[T(\beta_1^{null})]$$

Lawley は対数尤度関数を微分してモーメントを導出し Bartlett 補正項  $\xi$  の一般式を示した。Barndorff-Nielsen and Hall は、Bartlett 補正により尤度比検定統計量のカイ 2 乗分布に対する誤差が  $O(N^{-1})$  から  $O(N^{-2})$  に減少することを証明した (Barndorff-Nielsen 1988)。

本研究では、欠測を伴う小標本下の経時測定データに対して Bartlett 補正を導入した尤度比検定の実践的な手順を提案する。これまでに行われた研究は、Bartlett 補正項  $\xi$  を解析的な方法で推定しようとするものがほとんどであった (Lawley 1956; Zucker et al. 2000; Melo et al. 2009)。しかしながら、仮定されるモデルが複雑であったり、対象データに欠測が発生していたりすると、補正項の推定が常に容易ではない状況がある。Rocke はパラメトリックブートストラップ法によりリサンプリングのアプローチを使って Bartlett 補正項  $\xi$  を推定する方法を提案した (Rocke 1989)。本研究で提案する方法でも、帰無仮説下で MMRM を適用して推定したパラメータによりパラメトリックブートストラップ法を適用して、ブートストラップサンプルを生成

する。生成したブートストラップサンプルから、Bartlett 補正項 $\hat{\xi}$ を計算する。技術的にはノンパラメトリックブートストラップ法を適用することも可能であるが、小標本下で Bartlett 補正項 $\hat{\xi}$ を推定するにはパラメトリックブートストラップ法の使用が推奨されている (Wehrens et al. 2000; Noma et al. 2018)。

複雑なモデルが仮定されている状況であっても、統一化された手順が適用可能なので、パラメトリックブートストラップ法により Bartlett 補正項 $\hat{\xi}$ を推定することに利点がある。ブートストラップ法を基にした手順は以下のように定式化される。

#### アルゴリズム 1 (ブートストラップリサンプリング法を用いた Bartlett 補正)

- (1) 経時測定データ解析モデルの対数尤度関数 (式(3)) により、帰無仮説 $H_0: \beta_1 = \beta_1^{null}$ の下で最尤推定値 $\{\tilde{\beta}_c, \tilde{\mathbf{v}}\}$ を計算
- (2) パラメトリックブートストラップ法により、パラメータ $\{\tilde{\beta}_c, \tilde{\mathbf{v}}\}$ により経時測定データ解析モデル (式(1)) の帰無分布から $(\mathbf{Y}_1^{(b)}, \mathbf{Y}_2^{(b)}, \dots, \mathbf{Y}_N^{(b)})$ をリサンプリング (これを $B$ 回繰り返す( $b = 1, 2, \dots, B$ ))。なお、各 $\mathbf{Y}_i^{(b)}$ に解析対象データの欠測パターンを反映させる (すなわち、 $\mathbf{Y}_i^{(b)}$ に $n_i$ までのデータを持たせる)。
- (3) 生成した $(\mathbf{Y}_1^{(b)}, \mathbf{Y}_2^{(b)}, \dots, \mathbf{Y}_N^{(b)})$ に MMRM を適用して、最尤推定値 $\{\hat{\beta}_1^{(b)}, \hat{\beta}_c^{(b)}, \hat{\mathbf{v}}^{(b)}\}$ 及び帰無仮説下の制約付き最尤推定値 $\{\tilde{\beta}_c^{(b)}, \tilde{\mathbf{v}}^{(b)}\}$ を推定 (これを $B$ 回繰り返す( $b = 1, 2, \dots, B$ ))
- (4) 以下の式により $B$ 回分の尤度比検定統計量を計算

$$T^{(b)}(\beta_1^{null}) = -2 \left\{ l(\beta_1^{null}, \tilde{\beta}_c^{(b)}, \tilde{\mathbf{v}}^{(b)}) - l(\hat{\beta}_1^{(b)}, \hat{\beta}_c^{(b)}, \hat{\mathbf{v}}^{(b)}) \right\}$$

Bartlett 補正項 $\hat{\xi}$ を次の式により推定する。

$$\hat{\xi} = \frac{1}{B} \sum_{b=1}^B T^{(b)}(\beta_1^{null}).$$

- (5) Bartlett 補正された尤度比検定統計量を以下の式により得る

$$T_{BS}^*(\beta_1^{null}) = T(\beta_1^{null}) / \hat{\xi}.$$

自由度1のカイ2乗分布を参照しアルゴリズム中のステップ(5)で得た $T_{BS}^*(\beta_1^{null})$ を利用して、カイ2乗検定が可能である。さらに、以下に示す不等式を満たす $\beta_1^{null}$ の集合を求めることにより、 $100(1 - \alpha)\%$ 信頼区間を構成することもできる。

$$T_{BS}^*(\beta_1^{null}) \leq \chi_{1,\alpha}^2$$

ここで、 $\chi_{1,\alpha}^2$ は自由度1のカイ2乗分布における上側 $\alpha$ 分位点である。

### 3.4 ブートストラップ検定

Bartlett 補正は、検定統計量を補正することによりカイ2乗分布への近似を改善する方法であった。パラメトリックブートストラップ法では、モンテカルロ法によりブートストラップ標本を得て直接的に検定統計量が従う帰無分布を推定することができる (Efron 1992)。パラメトリックブートストラップ法により、帰無仮説下での尤度比検定統計量を複数リサンプリングし、

$T(\beta_1^{null})$ の帰無分布をモンテカルロ近似により推定している。提案法 2 は、アルゴリズム 1 で示したようなカイ 2 乗分布への近似を用いる代わりに、 $T(\beta_1^{null})$ の帰無分布を推定しようとする方法である。

アルゴリズム 1 のプロセス(1)～(4)を実行し、 $T^{(b)}(\beta_1^{null})$  ( $b = 1, 2, \dots, B$ )を得る。これは、 $B$  個の検定統計量で構成される $T(\beta_1^{null})$ の経験分布である。ブートストラップ法に基づく名義上の $\alpha$ 水準に対応する棄却限界値は、 $T(\beta_1^{null})$ の経験分布の上側 $\alpha$ 分位点に相当する。Rocke (1989)によれば、帰無分布の上側 $\alpha$ 分位点のような分布の裾を推定する場合には、推定精度を確保するために 1000 回以上のリサンプリングが推奨されている。

検定統計量の推定された帰無分布を利用した提案法 2 の手順は以下のように定式化される。

#### アルゴリズム 2 (ブートストラップリサンプリング法による経験分布を利用した補正)

- (1) アルゴリズム 1 のステップ(1)～(4)を実行
- (2) 以下の式により $p$ 値を計算 (Davison and Hinkley 1997)

$$p = \frac{1}{B+1} \left\{ 1 + \sum_{b=1}^B I[T^{(b)}(\beta_1^{null}) > T(\beta_1^{null})] \right\}$$

ここで、 $I(x)$ は指示関数であり、 $x$ が真のとき 1、 $x$ が偽のとき 0 を返す関数である。

以下に示す不等式を満たす $\beta_1^{null}$ の集合を求めることにより、 $100(1-\alpha)\%$ 信頼区間を構成することもできる (Zeng and Davidian 1997)。

$$T(\beta_1^{null}) \leq \hat{q}_{bs,(1-\alpha)}$$

ここで、 $\hat{q}_{bs,(1-\alpha)}$ はリサンプリングされた帰無仮説下の検定統計量で構成される経験分布の上側 $\alpha$ 分位点に相当する。

### 3.5 シミュレーション実験

#### 3.5.1 シナリオ

現実的な状況下における提案法の性能を評価するために、シミュレーション実験を実施した。シミュレーション実験の設定は、Gosho et al. (2017)の方法に従った。実験では、プラセボ群及び実薬群の 2 群の比較試験を想定する。対象被験者数は、2 群合計 $N$ とし各被験者 $i$  ( $i = 1, \dots, N$ ) について、経時的に 7 回 ( $t = 1, 2, \dots, 7$ ) の測定を仮定する。最終測定時点 (7 回目の測定時点) の群間差に興味があるものとする。被験者 $i$ 、時点 $t$ の測定値 $Y_{it}$ は以下の式により多変量正規分布を利用して生成した。

$$Y_{it} = \text{mean}_{it} + \text{subject}_i + \text{error}_{it}$$

ここで、 $\text{mean}$ は固定効果、 $\text{subject}$ を変量効果、 $\text{error}$ をランダム誤差とする。また、 $\text{subject}$ 及び $\text{error}$ は互いに独立と仮定する。

小標本下での性能評価が目的であるため、合計例数は $N = 10, 20, 40$  (1 群あたりでは、それぞれ 5, 10, 20 例) と設定した。測定値 $Y_{it}$ の平均構造について図 3.5-1 に示す 4 種類のシナリオを仮定した。シナリオ 1 及び 2 は、帰無仮説に対応しており最終測定時点の各群の測定値の平均値は同じである。一方、シナリオ 3 及び 4 は、対立仮説に対応しており最終測定時点の各

群の測定値の平均値は異なっている。シナリオ 1 及び 2 を用いて Type I error rate の検討，シナリオ 3 及び 4 を用いて検出力の検討が可能である。

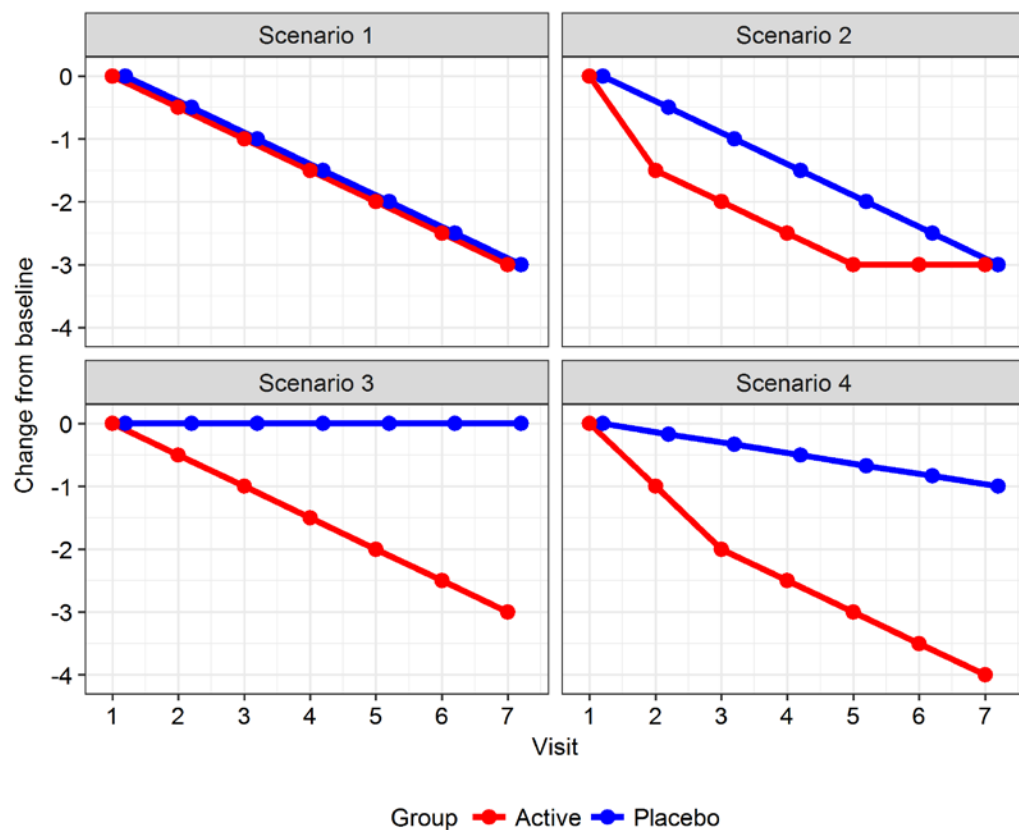


図 3.5-1 シミュレーション実験で用いた 4 種類のシナリオ

なお、一般の臨床試験では検出力を 80% もしくは 90% として設定する場合が多い。こうした現実的な想定を設定するために、さらに 2 種類のシナリオ（シナリオ 5 及び 6）を設定し検討に加えた。図 3.5-2 に追加したシナリオを示す。

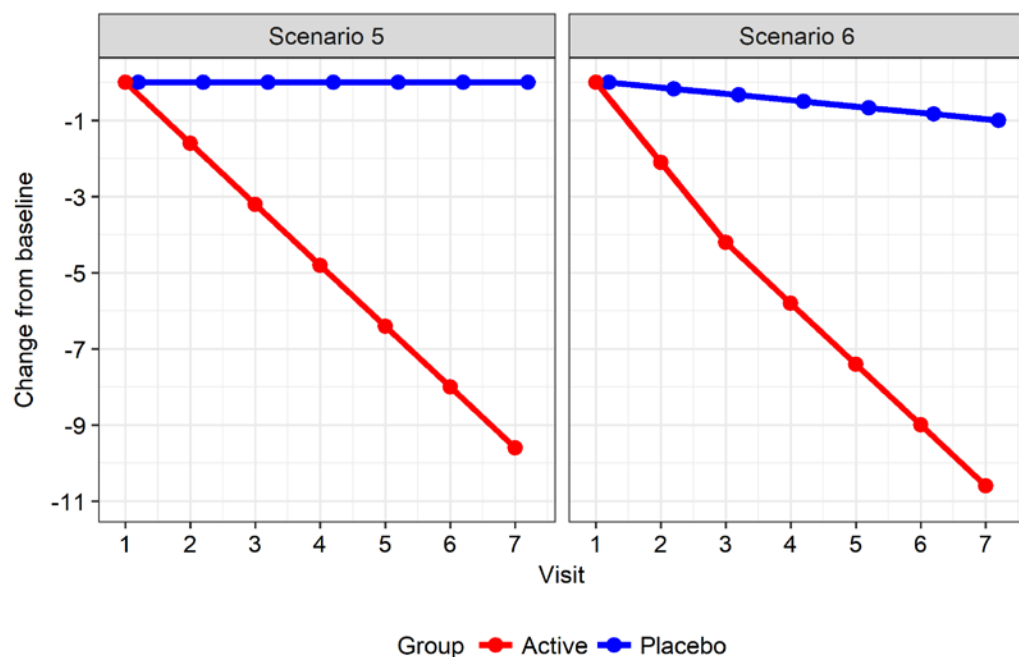


図 3.5-2 シミュレーション実験で用いた 2 種類の追加シナリオ

表 3.5-1 シナリオ 1, 2, 3 及び 4 における平均構造の設定値

Scenario	Group	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7
1	Placebo	0.0	-0.5	-1.0	-1.5	-2.0	-2.5	-3.0
	Active	0.0	-0.5	-1.0	-1.5	-2.0	-2.5	-3.0
2	Placebo	0.0	-0.5	-1.0	-1.5	-2.0	-2.5	-3.0
	Active	0.0	-1.5	-2.0	-2.5	-3.0	-3.0	-3.0
3	Placebo	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Active	0.0	-0.5	-1.0	-1.5	-2.0	-2.5	-3.0
4	Placebo	0.0	-0.17	-0.33	-0.5	-0.67	-0.83	-1.0
	Active	0.0	-1.0	-2.0	-2.5	-3.0	-3.5	-4.0

表 3.5-2 シナリオ 5 及び 6 における平均構造の設定値

Scenario	Group	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7
5	Placebo	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Active	0.0	-1.6	-3.2	-4.8	-6.4	-8.0	-9.6
6	Placebo	0.0	-0.17	-0.33	-0.5	-0.67	-0.83	-1.0
	Active	0.0	-2.1	-4.2	-5.8	-7.4	-9.0	-10.6

各シナリオの状況設定について触れる。シナリオ 1 では平均構造がすべての時点で同じであるのに対し、シナリオ 2 では最終測定時点の直前まで群間に差があり、最終測定時点で群間差が消失するシナリオである。シナリオ 3 及び 5 はプラセボ群と薬効のある実薬群を仮定した経時測定データであり、シナリオ 4 及び 6 は実薬群と標準治療群の比較試験のように両群ともに薬効があり、各測定時点で群間差が存在する経時測定データを想定している。

### 3.5.2 相関構造

本シミュレーションでは誤差の相関構造に ARH(1)を仮定した。ARH(1)は成分 $(t, t')$ が $\sigma_t \sigma_{t'} \rho^{|t'-t|}$  ( $\sigma_t$ は時点 $t$ における誤差項の標準偏差、 $\rho$ は隣接する時点間の相関係数を表す)で定義された分散共分散行列である。本定義に従えば、ARH(1)は、時点ごとに分散が異なり、相関は時点が遠くなるほど減少する相関構造を模した分散共分散行列と言える。対角成分 $\sigma_t^2$ は $\sigma_t^2 = 9\{1 + 3(t - 1)/6\}$ 、相関係数 $\rho$ は $\rho = 0.7$ と定義した。

7 時点を対象とする場合の ARH(1)は以下のように表すことができる。7 時点の場合には 8 個のパラメータ ( $\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7, \rho$ ) を推定することになる。

$$\begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho & \sigma_1 \sigma_3 \rho^2 & \sigma_1 \sigma_4 \rho^3 & \sigma_1 \sigma_5 \rho^4 & \sigma_1 \sigma_6 \rho^5 & \sigma_1 \sigma_7 \rho^6 \\ \sigma_2 \sigma_1 \rho & \sigma_2^2 & \sigma_2 \sigma_3 \rho & \sigma_2 \sigma_4 \rho^2 & \sigma_2 \sigma_5 \rho^3 & \sigma_2 \sigma_6 \rho^4 & \sigma_2 \sigma_7 \rho^5 \\ \sigma_3 \sigma_1 \rho^2 & \sigma_3 \sigma_2 \rho & \sigma_3^2 & \sigma_3 \sigma_4 \rho & \sigma_3 \sigma_5 \rho^2 & \sigma_3 \sigma_6 \rho^3 & \sigma_3 \sigma_7 \rho^4 \\ \sigma_4 \sigma_1 \rho^3 & \sigma_4 \sigma_2 \rho^2 & \sigma_4 \sigma_3 \rho & \sigma_4^2 & \sigma_4 \sigma_5 \rho & \sigma_4 \sigma_6 \rho^2 & \sigma_4 \sigma_7 \rho^3 \\ \sigma_5 \sigma_1 \rho^4 & \sigma_5 \sigma_2 \rho^3 & \sigma_5 \sigma_3 \rho^2 & \sigma_5 \sigma_4 \rho & \sigma_5^2 & \sigma_5 \sigma_6 \rho & \sigma_5 \sigma_7 \rho^2 \\ \sigma_6 \sigma_1 \rho^5 & \sigma_6 \sigma_2 \rho^4 & \sigma_6 \sigma_3 \rho^3 & \sigma_6 \sigma_4 \rho^2 & \sigma_6 \sigma_5 \rho & \sigma_6^2 & \sigma_6 \sigma_7 \rho \\ \sigma_7 \sigma_1 \rho^6 & \sigma_7 \sigma_2 \rho^5 & \sigma_7 \sigma_3 \rho^4 & \sigma_7 \sigma_4 \rho^3 & \sigma_7 \sigma_5 \rho^2 & \sigma_7 \sigma_6 \rho & \sigma_7^2 \end{pmatrix}$$

前述した設定に従い誤差の相関構造 $\Sigma_i$ について具体的に計算すると以下の分散共分散行列となる。

$$\begin{pmatrix} 9.00 & 7.72 & 6.24 & 4.88 & 3.74 & 2.83 & 2.12 \\ 7.72 & 13.50 & 10.91 & 8.54 & 6.55 & 4.95 & 3.71 \\ 6.24 & 10.91 & 18.00 & 14.09 & 10.80 & 8.17 & 6.11 \\ 4.88 & 8.54 & 14.09 & 22.50 & 17.25 & 13.04 & 9.76 \\ 3.74 & 6.55 & 10.80 & 17.25 & 27.00 & 20.41 & 15.28 \\ 2.83 & 4.95 & 8.17 & 13.04 & 20.41 & 31.50 & 23.57 \\ 2.12 & 3.71 & 6.11 & 9.76 & 15.28 & 23.57 & 36.00 \end{pmatrix}$$

変量効果 $subject_i$ は $N(0, D)$ に従うと仮定した。本シミュレーションでは、 $subject_i \sim N(0, 3^2)$ である。変量効果の分散共分散行列は $Z_i D Z_i'$ であり $Z_i = 1_7$ とすると、 $Z_i D Z_i' = 1_7 \cdot Var(subject_i) \cdot 1_7'$ とできる。すなわち、変量効果に関する相関構造は次のようになる。

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \cdot Var(subject_i) \cdot (1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1) = Var(subject_i) \cdot \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 \\ 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 \\ 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 \\ 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 \\ 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 \\ 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 \\ 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 \end{pmatrix}$$

結果変数 $Y_{it}$ の分散共分散行列は $V_i (V_i = Z_i D Z_i' + \Sigma_i)$ なので、ここまでの結果を使ってまとめると以下のようになる。

$$V_i = Z_i D Z_i' + \Sigma_i = \begin{pmatrix} 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 \\ 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 \\ 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 \\ 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 \\ 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 \\ 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 \\ 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 & 9.0 \end{pmatrix} + \begin{pmatrix} 9.00 & 7.72 & 6.24 & 4.88 & 3.74 & 2.83 & 2.12 \\ 7.72 & 13.50 & 10.91 & 8.54 & 6.55 & 4.95 & 3.71 \\ 6.24 & 10.91 & 18.00 & 14.09 & 10.80 & 8.17 & 6.11 \\ 4.88 & 8.54 & 14.09 & 22.50 & 17.25 & 13.04 & 9.76 \\ 3.74 & 6.55 & 10.80 & 17.25 & 27.00 & 20.41 & 15.28 \\ 2.83 & 4.95 & 8.17 & 13.04 & 20.41 & 31.50 & 23.57 \\ 2.12 & 3.71 & 6.11 & 9.76 & 15.28 & 23.57 & 36.00 \end{pmatrix}$$

$$= \begin{pmatrix} 18.00 & 16.72 & 15.24 & 13.88 & 12.74 & 11.83 & 11.12 \\ 16.72 & 22.50 & 19.91 & 17.54 & 15.55 & 13.95 & 12.71 \\ 15.24 & 19.91 & 27.00 & 23.09 & 19.80 & 17.17 & 15.11 \\ 13.88 & 17.54 & 23.09 & 31.50 & 26.25 & 22.04 & 18.76 \\ 12.74 & 15.55 & 19.80 & 26.25 & 36.00 & 29.41 & 24.28 \\ 11.83 & 13.95 & 17.17 & 22.04 & 29.41 & 40.50 & 32.57 \\ 11.12 & 12.71 & 15.11 & 18.76 & 24.28 & 32.57 & 45.00 \end{pmatrix}$$

### 3.5.3 欠測メカニズム

本シミュレーションでは、MCAR (missing completely at random) 及び MAR (missing at random) の2種類の欠測メカニズムを仮定する。MARは結果変数 $Y_i$ が欠測するかどうかは、観測値にのみ依存するとする仮定である。MCARであれば、測定データは単純に完全データからのランダムサンプリングと考えることができるので観測値のみから完全データのパラメータを

推定可能である。しかしながら、臨床試験などではMCARは通常成り立つとは考えづらい。

被験者 $i$ の測定時点 $t$ の測定値 $Y_{it}$ とし、 $Y_{it}$ が欠測する確率を $\lambda_{it}$ とする。なお、 $\lambda_{it}$ は時点 $(t-1)$ の測定値が観測されているもとの条件付き確率である。欠測確率 $\lambda_{it}$ が観測値及び欠測値を説明変数としたロジスティック回帰モデルで説明できると仮定すると、以下の式で欠測確率を表現できる。

$$\text{logit}(\lambda_{it}) = \gamma_0 + \gamma_1 Y_{it-1}, t = 2, 3, \dots, 7$$

シミュレーションでは単調な欠測を仮定する。すなわち、一度欠測が発生したら、欠測発生時点以降の測定値はすべて欠測値として取り扱う。

以上の定義を踏まえて、2種類の欠測メカニズムについて欠測確率に関するロジスティック回帰モデルの回帰係数を次のように設定する。すなわち、MCARでは、 $\gamma_1 = 0$ であり、MARでは、 $\gamma_1 = -1$ となる。

欠測メカニズムと各投与群の最終的な脱落割合及びその脱落割合を達成するロジスティック回帰モデルの係数を表 3.5-3 にまとめた。2群を合わせた全体での最終的な脱落割合は、20%及び40%の2種類を設定する。臨床試験の実際の期間や対象とする疾患領域次第で脱落割合も変動するが、脱落が多いときでも約半数の被験者が試験を完了している試験の設定は現実的であると考えられる。例えば、精神疾患領域での臨床試験では脱落率40%は珍しくないとの報告もある (Rabinowitz et al. 2008)。これらの設定に加えて、欠測のない完全データを含めて検討を行った。

表 3.5-3 シナリオ別の最終時点における累積脱落割合とパラメータ設定値

Scenario	Missing mechanism	Overall dropout (%)	Dropout (%)		$\gamma_0$	
			Placebo	Active	Placebo	Active
1	MCAR	20	20	20	3.2	3.2
		40	40	40	2.4	2.4
	MAR	20	20	20	7.1	7.1
		40	40	40	4.2	4.2
2	MCAR	20	22	18	3.1	3.4
		40	44	36	2.3	2.6
	MAR	20	22	18	6.6	6.6
		40	44	36	3.7	3.7
3	MCAR	20	24	16	3.0	3.5
		40	46	34	2.2	2.6
	MAR	20	24	16	7.8	7.8
		40	46	34	4.8	4.8
4	MCAR	20	24	16	3.0	3.5
		40	46	34	2.2	2.6
	MAR	20	24	16	7.2	7.2
		40	46	34	4.2	4.2
5	MCAR	20	26	17	3.0	3.5
		40	47	36	2.2	2.6
	MAR	20	29	11	7.4	7.4
		40	51	27	4.4	4.4
6	MCAR	20	26	17	3.0	3.5
		40	47	36	2.2	2.6
	MAR	20	28	11	7.0	7.0
		40	51	28	4.0	4.0

MCAR: missing completely at random; MAR: missing at random



### 3.5.4 評価方法

本研究で提案する 2 種類の方法を含めて合計 5 種類の方法を評価した。比較対象とした方法の略語を説明する。 $LR_{Bart}$  及び  $LR_{Boot}$  は、それぞれ尤度比検定に Bartlett 補正を施した方法（アルゴリズム 1）及び尤度比検定をブートストラップ法により補正した方法（アルゴリズム 2）である。LR は従来の尤度比検定であり提案法の直接的な比較対象である。 $t$ -test は最終測定時点のデータのみを対象とした  $t$  検定であり、検定統計量を計算する際に併合分散を用いた。

本研究では、小標本下での漸近近似を改善した方法として提案されている Kenward and Roger 法（Kenward and Roger 1997）を適用した MMRM（ $MMRM_{KR}$ ）を MMRM の比較対象として設定した。

以下の表 3.5-4 に略語と方法の対応表を示す。

表 3.5-4 略語と方法の対応表

略語	方法
$LR_{Bart}$	ブートストラップ法により推定した Bartlett 補正項で尤度比検定を補正する方法（アルゴリズム 1 に対応）
$LR_{Boot}$	ブートストラップ法により推定した帰無仮説下の尤度比検定統計量が従う経験分布を利用する方法（アルゴリズム 2 に対応）
LR	従来の尤度比検定
$MMRM_{KR}$	固定効果の標準誤差及び自由度調整法に Kenward and Roger 法を適用した MMRM
$t$ -test	最終測定時点のデータのみを対象とした 2 標本 $t$ 検定（検定統計量を計算する際の標準誤差は、併合分散により算出）

MMRM では、投与群及び時点を因子（クラス変数）として扱い、投与群と時点の交互作用項を含めたモデルにより解析を実施した。分散共分散構造は、無構造（Unstructured）を指定した。5 種類の比較方法ごとに最終測定時点における群間差に関する検定を実施し  $p$  値を計算した。 $MMRM_{KR}$  については、固定効果の標準誤差が 50 を超えた場合に推定が不適切であると判断し評価から除外する処理を行った。反復回数を分母としたときに  $p < 0.05$  を満たす  $p$  値の割合を、Type I error rate として計算した。さらに  $t$  検定を除いた 4 種類の比較方法について収束割合を算出した。本研究におけるシミュレーションプログラムは SAS Ver 9.4（SAS Institute, Cary, USA）で作成しており、収束判定の基準は SAS GLIMMIX プロシージャのデフォルト設定を適用した。

$LR_{Bart}$  及び  $LR_{Boot}$  でブートストラップ法により帰無仮説下の尤度比検定統計量をリサンプリングするときに、シミュレーションデータに依存して最尤法によるパラメータの収束が得られず結果として対数尤度が求められないため、ブートストラップ標本が得られない場合がある。 $LR_{Bart}$  及び  $LR_{Boot}$  でブートストラップ法では、収束が得られなかった場合を除外して合計 3000 回のリサンプリングがされるように調整した。先行研究より、3000 回のリサンプリング回数は必要十分な回数であると考えた。リサンプリング回数については検討した結果を別途示す（3.6.1 節を参照）。また、シミュレーションの反復回数は 1000 回とした。

### 3.5.5 従来法に対する提案法の基礎的挙動

従来法に対する提案法の基礎的挙動を確認するために、非欠測下の帰無仮説を仮定し従来の尤度比検定 LR と提案法（ $LR_{Bart}$  及び  $LR_{Boot}$ ）について QQ プロットを作成した。

従来の尤度比検定統計量  $LR$  は、小標本下でカイ 2 乗分布に対する近似が悪くなり提案法  $LR_{Bart}$  及び  $LR_{Boot}$  では近似が改善されることが予想される。実際に近似が改善されていること及び改善の程度を視覚的及び定量的に評価した。図 3.5-3 及び図 3.5-4 に帰無仮説の非欠測下における QQ プロットを示した。

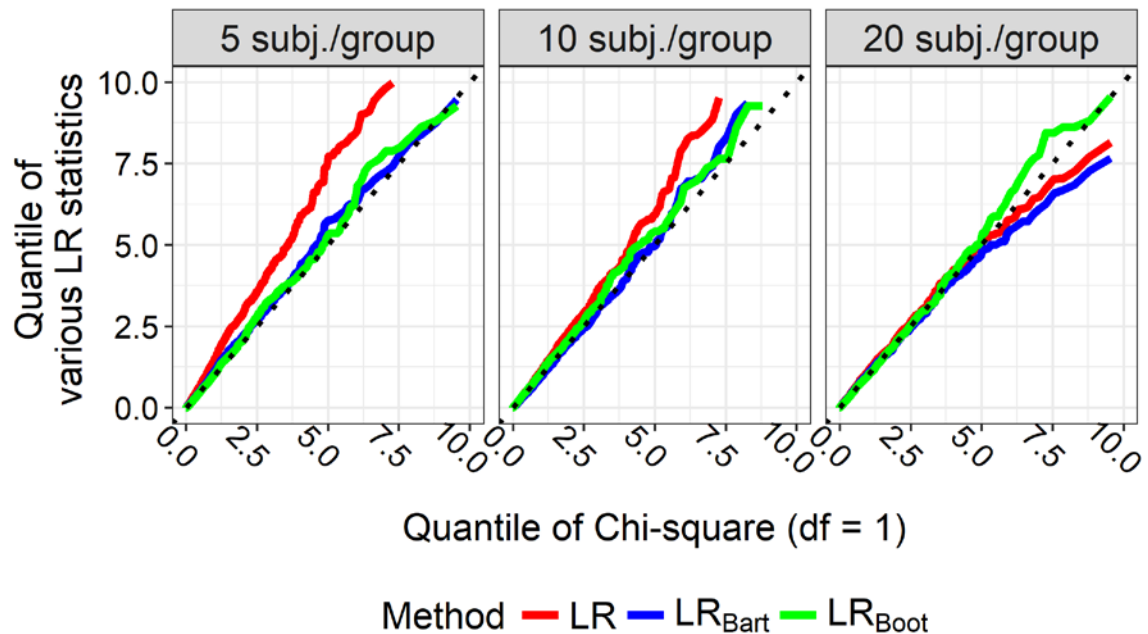


図 3.5-3 シナリオ 1 の非欠測下における QQ プロット

横軸，自由度 1 のカイ 2 乗分布の分位点；縦軸，従来の尤度比検定及び提案法 1 及び 2 の対応する分位点； $LR$ ，従来の尤度比検定； $LR_{Bart}$ ，Bartlett 型補正による尤度比検定； $LR_{Boot}$ ，ブートストラップ検定に基づく尤度比検定

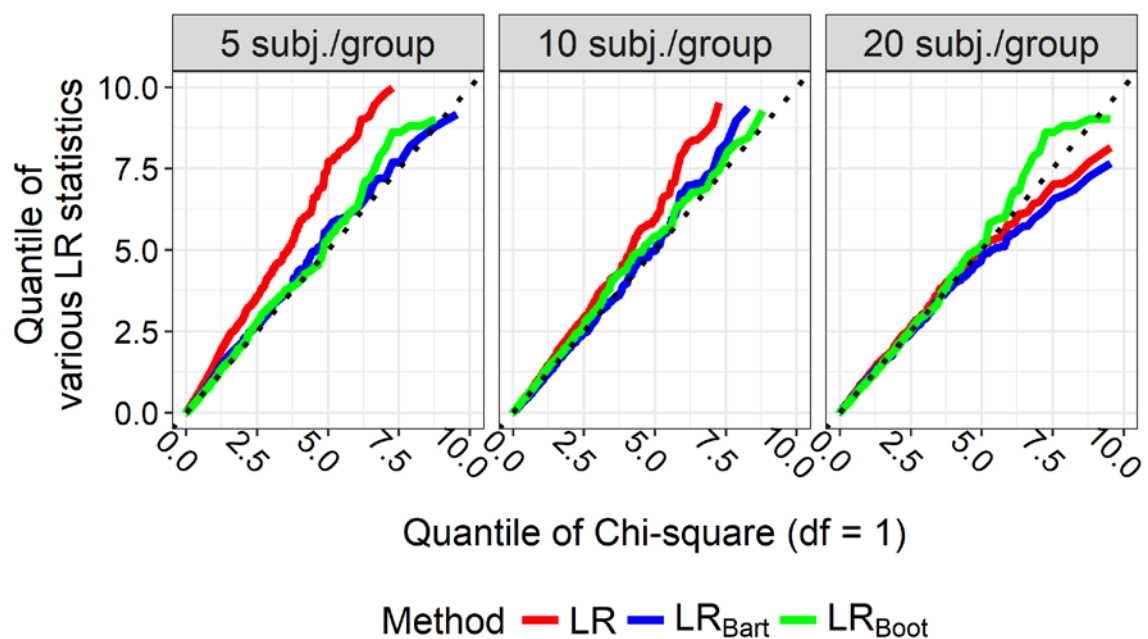


図 3.5-4 シナリオ 2 の非欠測下における QQ プロット

横軸，自由度 1 のカイ 2 乗分布の分位点；縦軸，従来の尤度比検定及び提案法 1 及び 2 の対応する

分位点；LR，従来の尤度比検定；LR<sub>Bart</sub>，Bartlett 型補正による尤度比検定；LR<sub>Boot</sub>，ブートストラップ検定に基づく尤度比検定

図 3.5-3 及び図 3.5-4 より上側 5%点を表 3.5-5 にまとめた。自由度 1 のカイ 2 乗分布における理論上の上側 5%点は 3.841 である。理論値に対する近似の程度を定量的に確認できる。

また，表 3.5-6 に帰無仮説の非欠測下における Type I error rate に関するシミュレーション実験結果をまとめた。カイ 2 乗分布への近似の程度と実際の Type I error rate の挙動を対応させて評価できる。

表 3.5-5 帰無仮説の非欠測下におけるカイ 2 乗分布の上側 5%点

シナリオ	1 群例数	LR	LR <sub>Bart</sub>	LR <sub>Boot</sub>
1	5	5.584	4.144	3.982
1	10	4.533	3.824	4.306
1	20	4.040	3.824	4.057
2	5	5.584	4.178	3.934
2	10	4.544	3.826	4.321
2	20	4.040	3.822	4.044

自由度 1 のカイ 2 乗分布における理論上の上側 5%点=3.841

表 3.5-6 帰無仮説の非欠測下における Type I error rate (%)

シナリオ	1 群例数	LR	LR <sub>Bart</sub>	LR <sub>Boot</sub>
1	5	9.90	5.71	5.91
1	10	7.30	5.00	4.90
1	20	5.70	4.80	4.80
2	5	9.90	5.61	5.81
2	10	7.30	5.00	4.90
2	20	5.70	4.80	4.80

反復回数 1000 回のシミュレーション実験結果

自由度 1 のカイ 2 乗分布における理論上の上側 5%点は，3.841 であり，表 3.5-5 より LR の上側 5%点は 3.841 よりも大きいため，Type I error rate は 5%よりも大きく inflation することが予想される。シミュレーション実験から得られた Type I error rate（表 3.5-6）より，LR では，予想通り Type I error rate が inflation していることが確認できた。

一方，提案法である LR<sub>Bart</sub> 及び LR<sub>Boot</sub> では上側 5%点が 3.841 に近似した値となっていることが分かる（表 3.5-5）。実際，シミュレーション実験から得られた Type I error rate も 5%に近い値となっていることが確認できた（表 3.5-6）。

### 3.5.6 $N = 10$ （1 群 5 例）

$N = 10$  の検討で，定義した欠測メカニズムにより欠測を発生させたデータを生成し，比較方法を適用した。 $t$ 検定以外の比較方法は最尤法（もしくは制限付最尤法）を用いた方法であるため，小標本かつ欠測が発生する  $N = 10$  の状況ではパラメータの収束がほとんど得られなかった。そこで， $N = 10$  のセクションでは，欠測がない完全データのみの検討結果を示す。

#### 3.5.6.1 Type I error rate

$N = 10$ （1 群 5 例）における Type I error rate の結果を，表 3.5-7 に示す。シナリオ 1 において，LR は 9.90% の Type I error rate を示した。シナリオ 1 の真の平均構造は群間に差がない。し

たがって、Type I error rate は 5% が理想である。大標本近似の下で成立する最尤法を用いている LR では既に報告されているように、固定効果の標準誤差が過小評価され結果として 5% を大きく上回る Type I error rate が観測された。

一方、提案法である LR<sub>Bart</sub> 及び LR<sub>Boot</sub> ではシナリオ 1 でそれぞれ 5.71% 及び 5.91% の Type I error rate を示した。理想である 5% を上回っているものの従来の尤度比検定である LR と比較して 5% に接近した Type I error rate であった。

小標本下での近似を改善する目的で提案された MMRM<sub>KR</sub> は 4.61% であり、真の Type I error rate である 5% を下回った。本検討では、保守的な Type I error rate となった。

シナリオ 2 もシナリオ 1 と同様な結果であった。

表 3.5-7 Type I error rate (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 1 及び 2, 欠測なし, 1 群 5 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	LR <sub>Bart</sub>	5.71	---	---	5.71	---	---
	LR <sub>Boot</sub>	5.91	---	---	5.91	---	---
	LR	9.90	---	---	9.90	---	---
	MMRM <sub>KR</sub>	4.61	---	---	4.61	---	---
	t-test	4.60	---	---	4.60	---	---
2	LR <sub>Bart</sub>	5.61	---	---	5.61	---	---
	LR <sub>Boot</sub>	5.81	---	---	5.81	---	---
	LR	9.90	---	---	9.90	---	---
	MMRM <sub>KR</sub>	4.72	---	---	4.72	---	---
	t-test	4.60	---	---	4.60	---	---

<sup>†</sup>LR<sub>Bart</sub>, Bartlett 型補正による尤度比検定: LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定: LR, 従来の尤度比検定: MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定: t-test, t 検定

### 3.5.6.2 検出力

N = 10 (1 群 5 例) における検出力の結果を、表 3.5-8 に示す。シナリオ 3 において、最も高い検出力を示した方法は LR であり 17.50% であった。LR<sub>Bart</sub> 及び LR<sub>BOOT</sub> はそれぞれ 11.51% 及び 11.71% であり、MMRM<sub>KR</sub> の 10.63% よりも高い検出力であった。

シナリオ 4 もシナリオ 3 と同様な結果であった。

シナリオ 5 及び 6 でも LR の検出力が最も高く、MCAR 及び MAR 下でそれぞれ 65.35% 及び 65.49% であった。

表 3.5-8 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 3, 4, 5 及び 6, 欠測なし, 1 群 5 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
3	LR <sub>Bart</sub>	11.51	---	---	11.51	---	---
	LR <sub>Boot</sub>	11.71	---	---	11.71	---	---
	LR	17.50	---	---	17.50	---	---
	MMRM <sub>KR</sub>	10.63	---	---	10.63	---	---
	t-test	10.60	---	---	10.60	---	---
4	LR <sub>Bart</sub>	11.21	---	---	11.21	---	---

	LR <sub>Boot</sub>	11.51	---	---	11.51	---	---
	LR	17.50	---	---	17.50	---	---
	MMRM <sub>KR</sub>	10.75	---	---	10.75	---	---
	<i>t</i> -test	10.60	---	---	10.60	---	---
5	LR <sub>Bart</sub>	52.44	---	---	52.44	---	---
	LR <sub>Boot</sub>	53.56	---	---	53.56	---	---
	LR	65.35	---	---	65.35	---	---
	MMRM <sub>KR</sub>	50.45	---	---	50.45	---	---
	<i>t</i> -test	50.70	---	---	50.70	---	---
6	LR <sub>Bart</sub>	52.83	---	---	52.83	---	---
	LR <sub>Boot</sub>	52.94	---	---	52.94	---	---
	LR	65.49	---	---	65.49	---	---
	MMRM <sub>KR</sub>	50.50	---	---	50.50	---	---
	<i>t</i> -test	50.70	---	---	50.70	---	---

<sup>†</sup>LR<sub>Bart</sub>, Bartlett 型補正による尤度比検定： LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定： LR, 従来の尤度比検定： MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定： *t*-test, *t*検定

### 3.5.6.3 収束割合（シナリオ 1 及び 2）

$N = 10$ （1 群 5 例）におけるシナリオ 1 及び 2 の収束割合の結果を、表 3.5-9 に示す。欠測がない完全データであれば、いずれの方法であってもほぼ 100%の収束割合が得られた。なお、*t*-test は対象データが存在すれば、*t*検定統計量を計算することができる。つまり、反復計算によりパラメータを求めるアプローチを取っていない。したがって、収束割合の検討から除外しているため、表中に結果を表示していない。

表 3.5-9 収束割合（%）のシミュレーション実験結果<sup>†</sup>（シナリオ 1 及び 2, 欠測なし, 1 群 5 例）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	LR <sub>Bart</sub>	99.9	---	---	99.9	---	---
	LR <sub>Boot</sub>	99.9	---	---	99.9	---	---
	LR	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
2	LR <sub>Bart</sub>	99.8	---	---	99.8	---	---
	LR <sub>Boot</sub>	99.8	---	---	99.8	---	---
	LR	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---

<sup>†</sup>LR<sub>Bart</sub>, Bartlett 型補正による尤度比検定： LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定： LR, 従来の尤度比検定： MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定

### 3.5.6.4 収束割合（シナリオ 3, 4, 5 及び 6）

$N = 10$ （1 群 5 例）におけるシナリオ 3, 4, 5 及び 6 の収束割合の結果を、表 3.5-10 に示す。欠測がない完全データであれば、いずれの方法であってもほぼ 100%の収束割合が得られた。シナリオ 1 及び 2 で得られた結果と、同等である。

表 3.5-10 収束割合（%）のシミュレーション実験結果<sup>†</sup>（シナリオ 3, 4, 5 及び 6, 欠測なし, 1 群 5 例）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%

3	LR <sub>Bart</sub>	99.9	---	---	99.9	---	---
	LR <sub>Boot</sub>	99.9	---	---	99.9	---	---
	LR	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
4	LR <sub>Bart</sub>	99.9	---	---	99.9	---	---
	LR <sub>Boot</sub>	99.9	---	---	99.9	---	---
	LR	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
5	LR <sub>Bart</sub>	98.4	---	---	98.4	---	---
	LR <sub>Boot</sub>	98.4	---	---	98.4	---	---
	LR	98.4	---	---	98.4	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
6	LR <sub>Bart</sub>	98.8	---	---	98.8	---	---
	LR <sub>Boot</sub>	98.8	---	---	98.8	---	---
	LR	98.8	---	---	98.8	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---

<sup>†</sup>LR<sub>Bart</sub>, Bartlett 型補正による尤度比検定： LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定： LR, 従来の尤度比検定： MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定

### 3.5.7 $N = 20$ (1 群 10 例)

#### 3.5.7.1 Type I error rate

$N = 20$  (1 群 10 例) における Type I error rate の結果を、表 3.5-11 及び図 3.5-5 に示す。LR の Type I error rate は、累積欠測割合が増加するにつれて 5% から大きく上昇した。シナリオ 1 の欠測割合が 40% の状況では、MCAR の下で LR の Type I error rate はそれぞれ 11.33% であった。また、同様に MAR の下では LR は 9.76% の Type I error rate であった。

図 3.5-5 の水平な青点線は真の Type I error rate が 5% であると仮定した場合に 1000 回の反復を実施した場合のモンテカルロエラーを示している。そこで、上側と下側の青点線内にシミュレーションで得られた Type I error rate が存在する場合に本研究では、Type I error rate が 5% に維持されていると判断する目安とした。水平な赤点線は Type I error rate が 5% となる位置を示している。LR<sub>Bart</sub>, LR<sub>Boot</sub> 及び  $t$ -test の 3 つの方法は欠測メカニズムや累積欠測割合に関わらず Type I error rate が 5% に維持されていると考えることができる。例えば、シナリオ 1 の欠測割合 40% の MAR 下で LR<sub>Bart</sub>, LR<sub>Boot</sub> 及び  $t$ -test はそれぞれ、5.47%, 5.58% 及び 4.70% であった。

一方、MMRM<sub>KR</sub> は MAR 下で 5% を維持しているとは言えず、保守的な値を示した。累積欠測割合 40% の MAR 下で MMRM<sub>KR</sub> の Type I error rate は、シナリオ 1 及び 2 でそれぞれ 3.62% 及び 3.47% であった。MCAR 下においては MMRM<sub>KR</sub> も Type I error rate が 5% を維持できていた。 $t$ -test の Type I error rate はシミュレーションによる検討では 5% の名目水準を維持できているように見えた。しかしながら、群間差の標準偏差が増大することにより、 $t$ -test による最終時点の群間差の評価にはバイアスが発生していると考えられた。

1 群 10 例の場合に累積欠測割合 40% とは、完了例が平均的に 6 例程度であることを意味している。前述したように、1 群 5 例の完全データに対して同様なシミュレーションを実施したところ、シナリオ 1 で LR<sub>Bart</sub>, LR<sub>Boot</sub> 及び  $t$ -test はそれぞれ、5.71%, 5.91% 及び 4.60% であった。欠測メカニズムが MCAR 及び MAR のいずれの場合であっても、完全データと同等な検定サイズに関する性能を持つことが示唆された。

表 3.5-11 Type I error rate (%)のシミュレーション実験結果<sup>†</sup> (シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 10 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	LR <sub>Bart</sub>	5.00	5.07	5.79	5.00	5.13	5.47
	LR <sub>Boot</sub>	4.90	5.07	6.13	4.90	5.33	5.58
	LR	7.30	8.21	11.33	7.30	7.34	9.76
	MMRM <sub>KR</sub>	5.10	5.00	5.70	5.10	3.40	3.62
	<i>t</i> -test	5.10	5.50	4.70	5.10	5.30	4.70
2	LR <sub>Bart</sub>	5.00	5.27	4.76	5.00	5.23	4.95
	LR <sub>Boot</sub>	4.90	5.27	5.19	4.90	5.23	5.19
	LR	7.30	8.32	11.04	7.30	7.64	9.41
	MMRM <sub>KR</sub>	5.10	4.70	5.68	5.10	3.30	3.47
	<i>t</i> -test	5.10	5.20	4.80	5.10	5.50	5.01

<sup>†</sup>LR<sub>Bart</sub>, Bartlett 型補正による尤度比検定： LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定： LR, 従来の尤度比検定： MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定： *t*-test, *t*検定

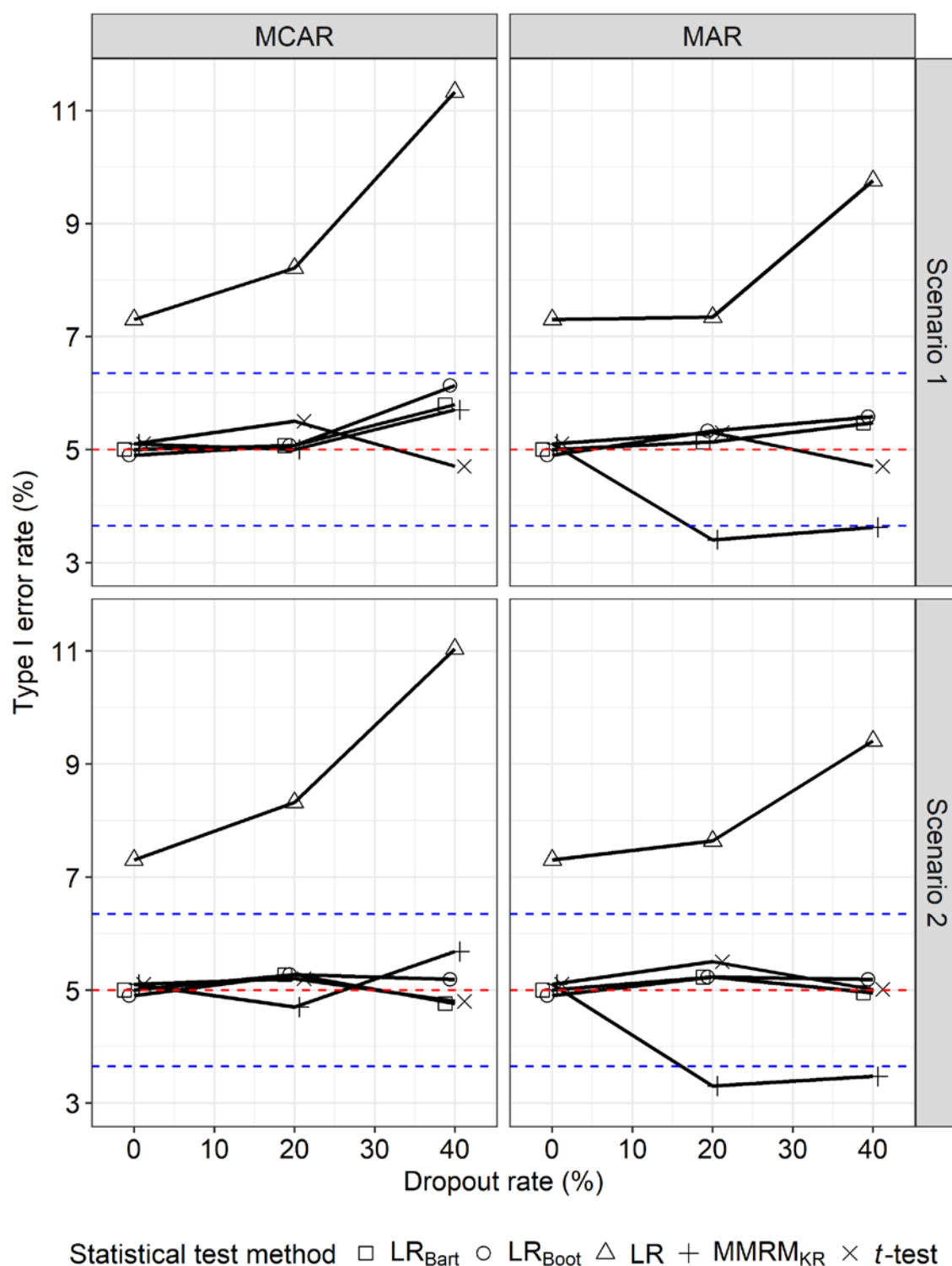


図 3.5-5 Type I error rates (%)シミュレーション実験結果<sup>†</sup> (シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 10 例)

赤点線, 5%; 青点線, 95%信頼限界 (真の Type I error rate を 5% と仮定した場合の 1000 回反復によるモンテカルロ誤差); <sup>†</sup>LR<sub>Bart</sub>, Bartlett 補正による尤度比検定; LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定; LR, 従来の尤度比検定; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定; *t*-test, *t*検定



### 3.5.7.2 検出力

$N = 20$  (1 群 10 例) における検出力の結果を、表 3.5-12 及び図 3.5-6 に示す。シナリオ 3 及び 4 の最終測定時点における群間差はいずれも-3.0 であることから、有意水準を 5% としたときに検定の回数に対して有意となる割合は検出力と解釈することができる。

図 3.5-6 より、いずれのパターンであっても LR の検出力は他の方法よりも高く累積欠測割合に応じて約 14% から 20% の値を示した。しかしながら、LR の Type I error rate は 5% に維持されていなかった (3.5.7.1 を参照)。Type I error rate の維持を最優先に考えるとすれば、検出力だけが他の方法よりも優れていることをもって良い性能を持つと評価することはできない。

LR 以外の 4 つの方法は、いずれも完全データ (累積欠測割合 0%) から累積欠測割合が増加するに従い検出力が減少した。累積欠測割合が増加した結果、対象となるデータ数が減少し検出力が低下したことが原因として考えられた。1 群 10 例のシナリオ 3 で累積欠測割合 40% の MAR 下における LR<sub>Bart</sub>, LR<sub>Boot</sub>, MMRM<sub>KR</sub> 及び  $t$ -test の検出力は、それぞれ 7.06%, 7.41%, 6.53% 及び 7.51% であった。

表 3.5-12 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 10 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
3	LR <sub>Bart</sub>	15.90	12.98	11.46	15.90	11.24	7.06
	LR <sub>Boot</sub>	15.80	13.18	12.24	15.80	11.75	7.41
	LR	19.60	18.41	17.78	19.60	17.17	13.94
	MMRM <sub>KR</sub>	16.20	13.20	11.71	16.20	12.00	6.53
	$t$ -test	16.20	12.70	10.50	16.20	11.00	7.51
4	LR <sub>Bart</sub>	15.90	12.78	11.65	15.90	11.65	7.56
	LR <sub>Boot</sub>	15.70	13.28	12.10	15.70	12.25	7.91
	LR	19.60	18.41	17.52	19.60	16.97	14.29
	MMRM <sub>KR</sub>	16.20	13.20	11.67	16.20	11.71	7.66
	$t$ -test	16.20	12.70	10.50	16.20	11.40	7.01
5	LR <sub>Bart</sub>	86.60	72.95	56.47	86.60	72.75	50.76
	LR <sub>Boot</sub>	86.40	73.25	56.25	86.40	72.75	50.86
	LR	90.60	80.96	69.93	90.60	80.46	63.82
	MMRM <sub>KR</sub>	87.10	74.10	58.82	87.10	74.77	51.01
	$t$ -test	87.10	72.40	57.30	87.10	68.90	48.70
6	LR <sub>Bart</sub>	86.60	72.95	56.70	86.60	73.35	48.38
	LR <sub>Boot</sub>	86.40	72.95	57.14	86.40	73.85	48.70
	LR	90.60	80.96	69.82	90.60	80.46	62.28
	MMRM <sub>KR</sub>	87.10	74.10	58.84	87.10	74.50	50.43
	$t$ -test	87.10	72.40	57.30	87.10	70.10	48.60

<sup>†</sup>LR<sub>Bart</sub>, Bartlett 型補正による尤度比検定 : LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定 : LR, 従来の尤度比検定 : MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 :  $t$ -test,  $t$ 検定

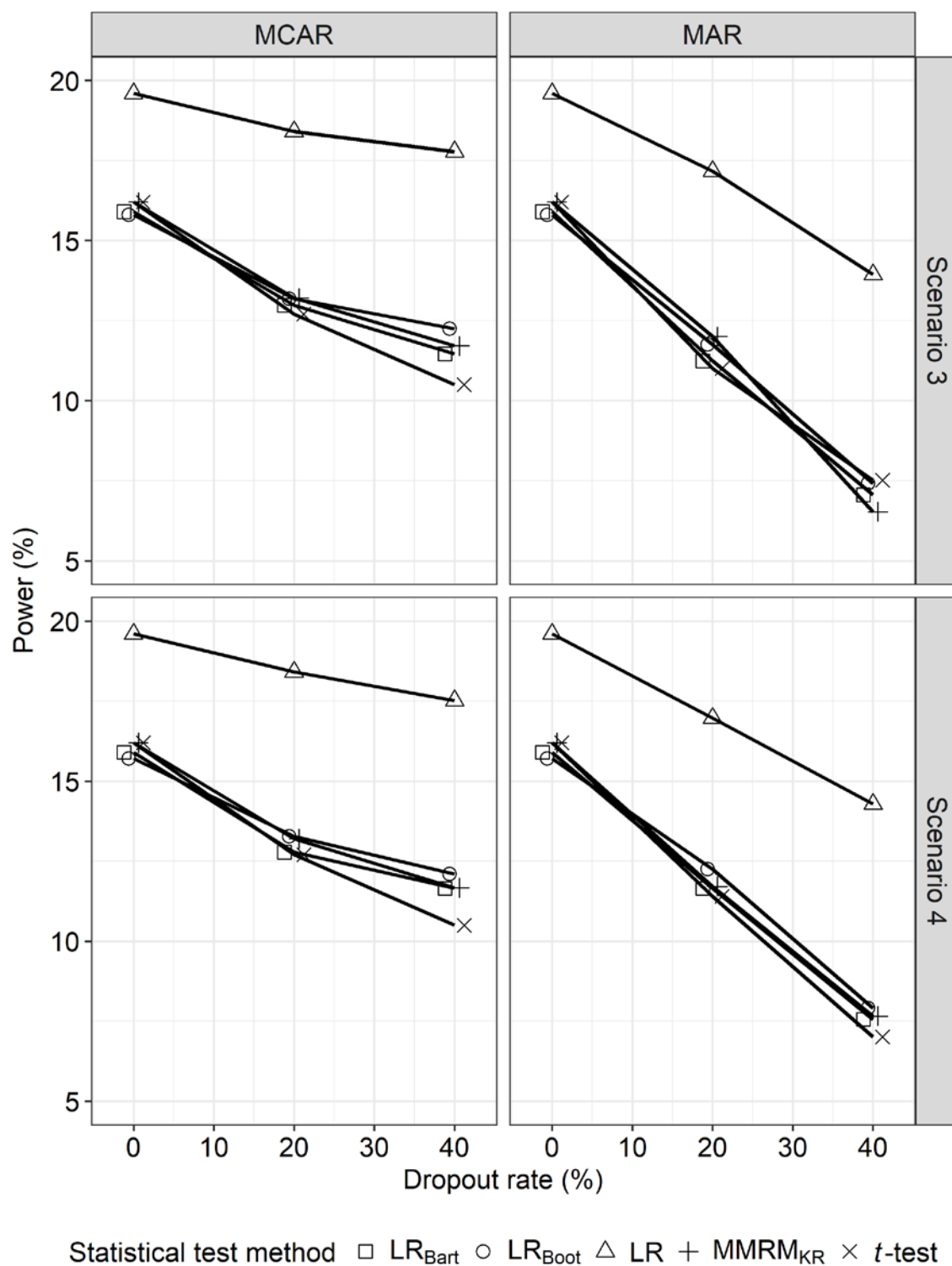


図 3.5-6 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 10 例)

<sup>†</sup>  $LR_{Bart}$ , Bartlett 補正による尤度比検定 ;  $LR_{Boot}$ , ブートストラップ検定に基づく尤度比検定 ;  $LR$ , 従来の尤度比検定 ;  $MMRM_{KR}$ , Kenward and Roger 法による REML 推定を用いた Wald 型検定 ;  $t$ -test,  $t$  検定

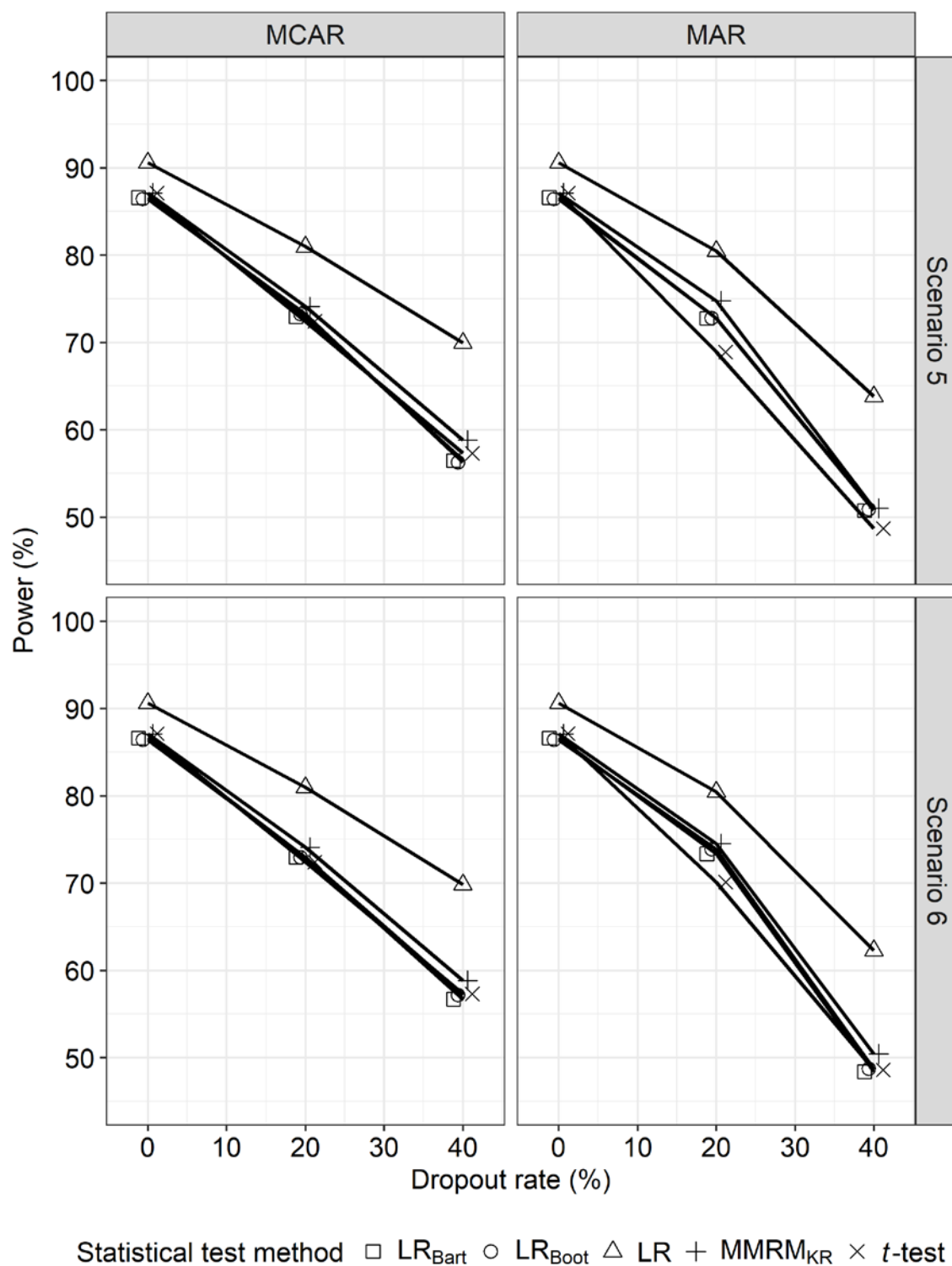


図 3.5-7 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 10 例)

<sup>†</sup>  $LR_{Bart}$ , Bartlett 補正による尤度比検定 ;  $LR_{Boot}$ , ブートストラップ検定に基づく尤度比検定 ;  $LR$ , 従来の尤度比検定 ;  $MMRM_{KR}$ , Kenward and Roger 法による REML 推定を用いた Wald 型検定 ;  $t$ -test,  $t$  検定

### 3.5.7.3 収束割合（シナリオ 1 及び 2）

$N = 20$ （1 群 10 例）におけるシナリオ 1 及び 2 の収束割合の結果を、表 3.5-13 及び図 3.5-8 に示す。欠測のない完全データを適用した場合には、1 群 5 例の小標本下であっても本検討で対象としたいずれの方法もほぼ 100% の収束割合を示した。1 群 10 例の検討では、欠測のない完全データを適用した場合には、1 群 5 例の場合と同様に 100% の収束割合を示した。累積欠測割合が高くなるにつれて、収束割合は低下した。例えば、シナリオ 1 の累積欠測割合 40% の MAR 下では、 $LR_{Bart}$ 、 $LR_{Boot}$ 、LR 及び  $MMRM_{KR}$  の収束割合は、それぞれ 86.0%、86.0%、86.7% 及び 87.4% であった。図表中に  $t$  検定の結果を表示していない理由は、3.5.6.3 で既に述べた通りである。

表 3.5-13 収束割合（%）のシミュレーション実験結果<sup>†</sup>（シナリオ 1 及び 2，MCAR 及び MAR 下，1 群 10 例）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	$LR_{Bart}$	100.0	98.7	88.1	100.0	99.4	86.0
	$LR_{Boot}$	100.0	98.7	88.1	100.0	99.4	86.0
	LR	100.0	99.9	90.8	100.0	100.0	86.7
	$MMRM_{KR}$	100.0	100.0	91.8	100.0	100.0	87.4
2	$LR_{Bart}$	100.0	98.6	92.4	100.0	99.5	82.8
	$LR_{Boot}$	100.0	98.6	92.4	100.0	99.5	82.8
	LR	100.0	99.9	94.8	100.0	99.9	83.2
	$MMRM_{KR}$	100.0	100.0	95.7	100.0	99.9	84.6

<sup>†</sup> $LR_{Bart}$ , Bartlett 型補正による尤度比検定：  $LR_{Boot}$ , ブートストラップ検定に基づく尤度比検定： LR, 従来の尤度比検定：  $MMRM_{KR}$ , Kenward and Roger 法による REML 推定を用いた Wald 型検定

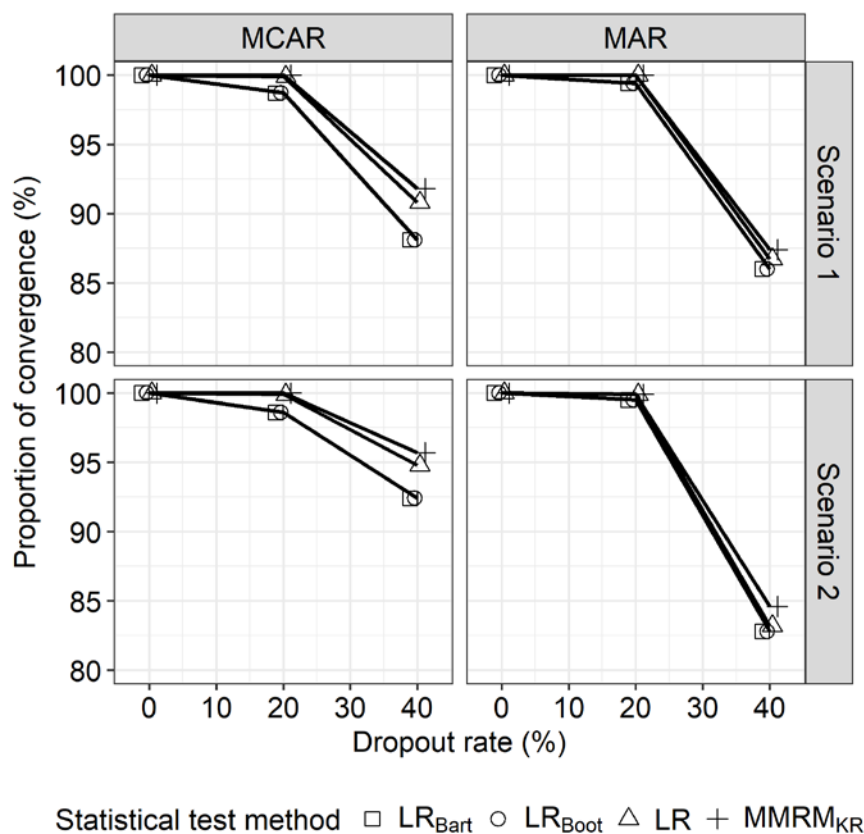


図 3.5-8 収束割合（％）のシミュレーション実験結果<sup>†</sup>（シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 10 例）

<sup>†</sup> LR<sub>Bart</sub>, Bartlett 補正による尤度比検定；LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定；LR, 従来の尤度比検定；MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定

### 3.5.7.4 収束割合（シナリオ 3, 4, 5 及び 6）

$N = 20$ （1 群 10 例）におけるシナリオ 3, 4, 5 及び 6 の収束割合の結果を、表 3.5-14, 図 3.5-9 及び図 3.5-10 に示す。収束割合の結果は、これまでに検討したシナリオ 1 及び 2（群間差がない帰無仮説下のシナリオ）における結果と同等であった。すなわち、欠測のない完全データに対しては、100%の収束割合が得られ、累積欠測割合が増加するにつれて収束割合が低下した。ただし、シナリオ 5 及び 6 では累積欠測割合が 40%の場合でも約 90%前後の収束割合が得られた。

表 3.5-14 収束割合（％）のシミュレーション実験結果<sup>†</sup>（シナリオ 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 10 例）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
3	LR <sub>Bart</sub>	100.0	99.4	89.9	100.0	99.6	86.4
	LR <sub>Boot</sub>	100.0	99.4	89.9	100.0	99.6	86.4
	LR	100.0	100.0	91.1	100.0	99.8	87.0
	MMRM <sub>KR</sub>	100.0	100.0	92.1	100.0	100.0	88.5
4	LR <sub>Bart</sub>	100.0	99.4	90.1	100.0	99.6	84.7
	LR <sub>Boot</sub>	100.0	99.4	90.1	100.0	99.6	84.7
	LR	100.0	100.0	91.4	100.0	99.9	85.0
	MMRM <sub>KR</sub>	100.0	100.0	92.2	100.0	99.9	86.5
5	LR <sub>Bart</sub>	100.0	99.8	89.6	100.0	99.8	92.6
	LR <sub>Boot</sub>	100.0	99.8	89.6	100.0	99.8	92.6
	LR	100.0	100.0	90.2	100.0	99.8	92.7
	MMRM <sub>KR</sub>	100.0	100.0	92.3	100.0	99.9	94.0
6	LR <sub>Bart</sub>	100.0	99.8	89.6	100.0	99.8	92.6
	LR <sub>Boot</sub>	100.0	99.8	89.6	100.0	99.8	92.6
	LR	100.0	100.0	90.3	100.0	99.8	93.0
	MMRM <sub>KR</sub>	100.0	100.0	92.3	100.0	100.0	93.9

<sup>†</sup> LR<sub>Bart</sub>, Bartlett 型補正による尤度比検定；LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定；LR, 従来の尤度比検定；MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定

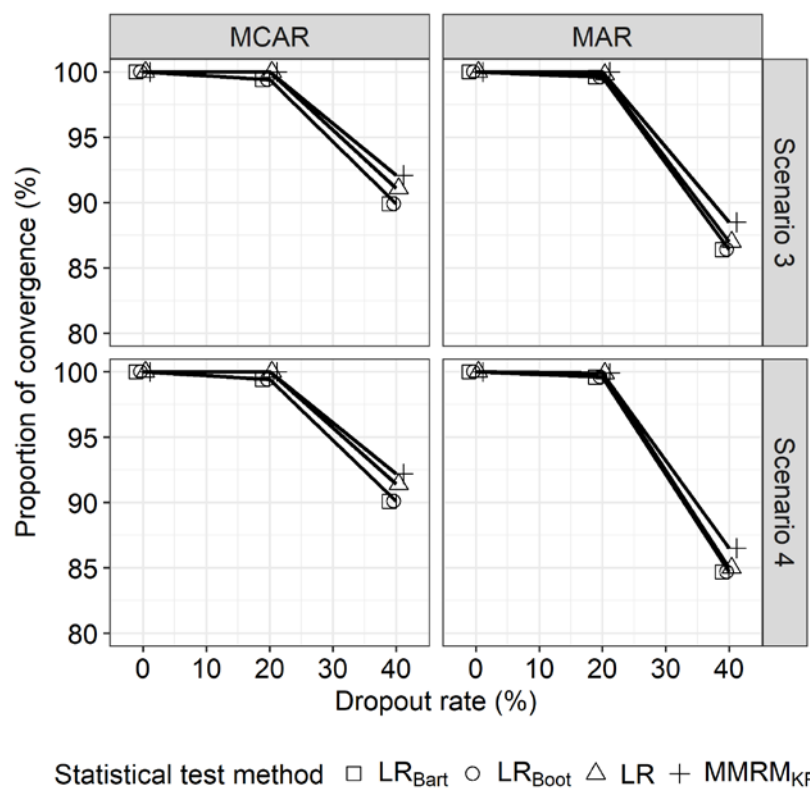


図 3.5-9 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 10 例)

<sup>†</sup>  $LR_{Bart}$ , Bartlett 補正による尤度比検定;  $LR_{Boot}$ , ブートストラップ検定に基づく尤度比検定;  $LR$ , 従来の尤度比検定;  $MMRM_{KR}$ , Kenward and Roger 法による REML 推定を用いた Wald 型検定

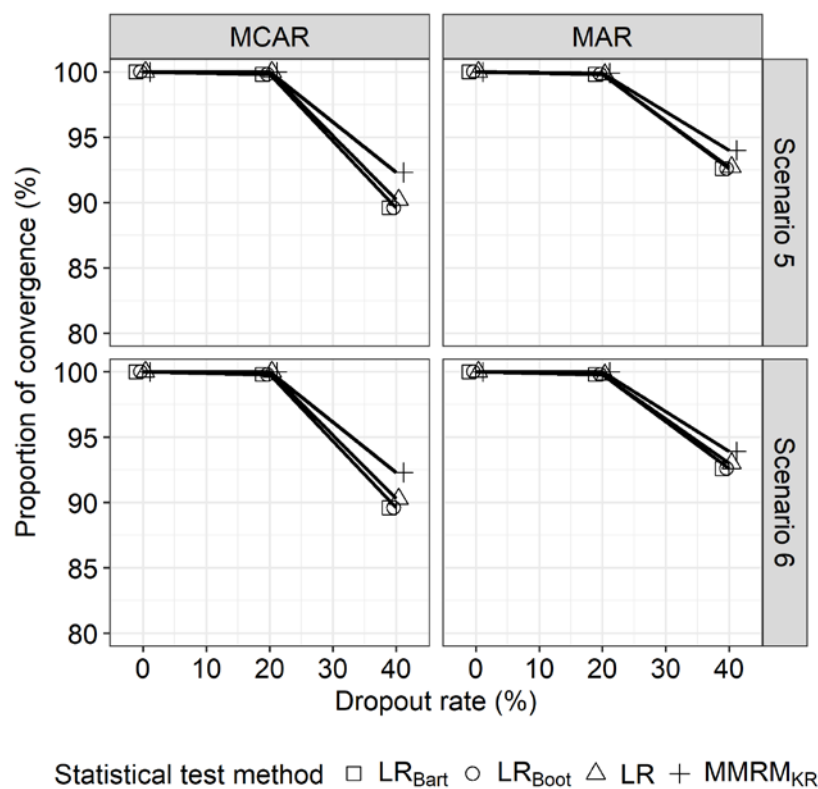


図 3.5-10 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 5 及び 6, MCAR 及び MAR

下, 1 群 10 例)

† LR<sub>Bart</sub>, Bartlett 補正による尤度比検定 ; LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定 ; LR, 従来の尤度比検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定

### 3.5.8 $N = 40$ (1 群 20 例)

#### 3.5.8.1 Type I error rate

$N = 40$  (1 群 20 例) における Type I error rate の結果を, 表 3.5-15 及び図 3.5-11 に示す。 $N = 20$  (1 群 10 例) の検討結果と同様に, LR は累積欠測割合が高くなるにつれて Type I error rate が上昇し, MCAR 及び MAR の累積欠測割合 40%の下では Type I error rate を 5%に維持できなかった。しかしながら, 1 群 20 例の場合と比較して標本数が多くなったため, Type I error rate は累積欠測割合が高い状況でも名目有意水準である 5%に近づいた。

シナリオ 1 及び 2 ともに同様な傾向が認められた。

表 3.5-15 Type I error rate (%)のシミュレーション実験結果† (シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 20 例)

シナリオ	方法†	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	LR <sub>Bart</sub>	4.80	5.82	5.27	4.80	5.02	4.31
	LR <sub>Boot</sub>	4.80	6.12	5.37	4.80	4.91	4.31
	LR	5.70	6.83	7.40	5.70	6.42	6.91
	MMRM <sub>KR</sub>	4.50	4.70	5.00	4.50	5.10	4.40
	<i>t</i> -test	4.50	4.50	3.80	4.50	5.20	4.80
2	LR <sub>Bart</sub>	4.80	5.74	5.16	4.80	4.91	4.00
	LR <sub>Boot</sub>	4.80	5.74	5.46	4.80	4.91	4.10
	LR	5.70	6.65	7.28	5.70	6.42	6.81
	MMRM <sub>KR</sub>	4.50	5.30	6.10	4.50	4.80	5.30
	<i>t</i> -test	4.50	5.00	4.50	4.50	5.00	4.60

†LR<sub>Bart</sub>, Bartlett 型補正による尤度比検定 ; LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定 ; LR, 従来の尤度比検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; *t*-test, *t*検定

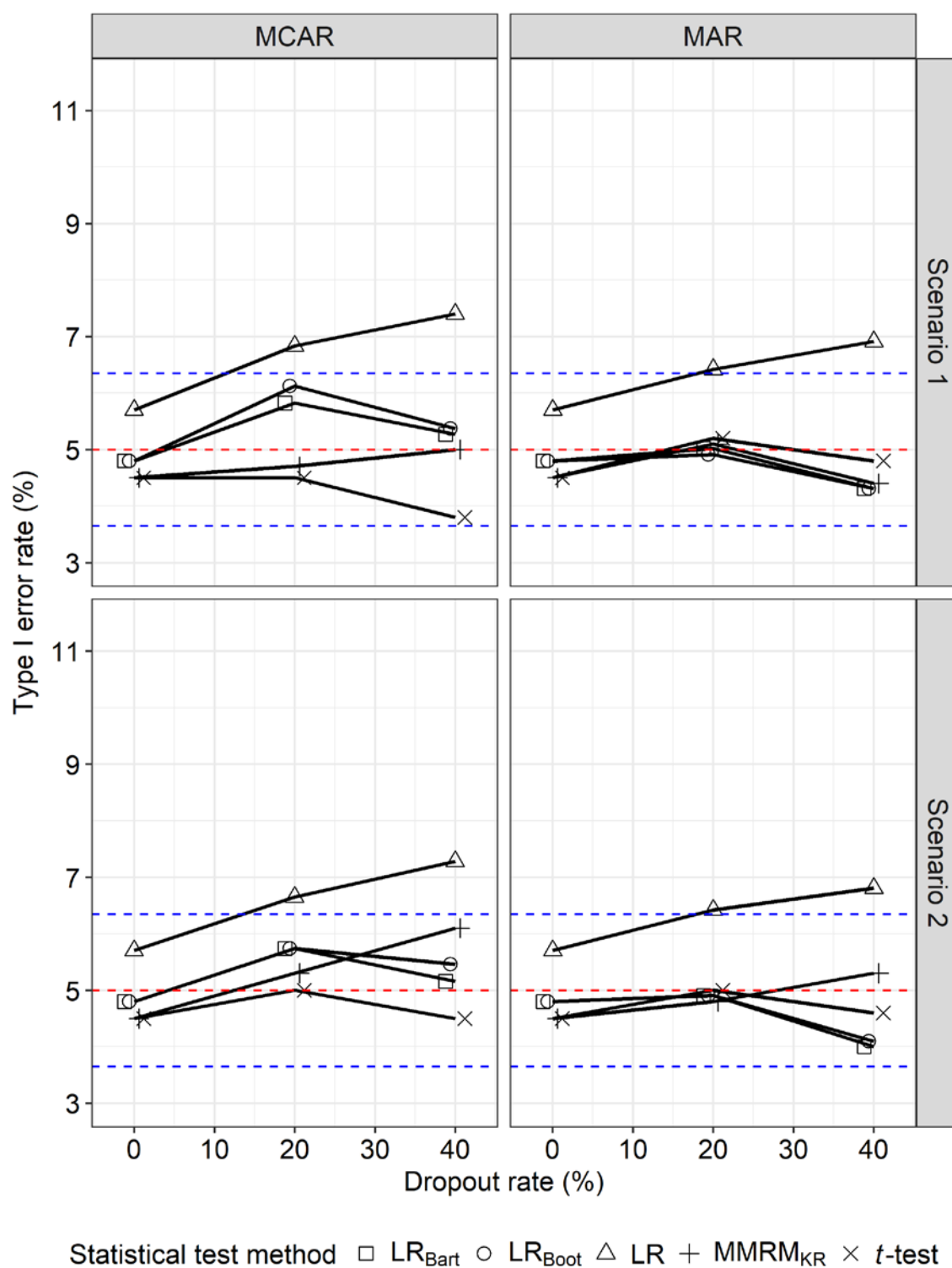


図 3.5-11 Type I error rates (%)のシミュレーション実験結果<sup>†</sup> (シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 20 例)

赤点線, 5% ; 青点線, 95%信頼限界 (真の Type I error rate を 5% と仮定した場合の 1000 回反復によるモンテカルロ誤差) ; <sup>†</sup>LR<sub>Bart</sub>, Bartlett 補正による尤度比検定 ; LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定 ; LR, 従来の尤度比検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; *t*-test, *t*検定



### 3.5.8.2 検出力

$N = 40$  (1 群 20 例) における検出力の結果を、表 3.5-16、図 3.5-12 及び図 3.5-13 に示す。 $N = 20$  (1 群 10 例) と比較して、検出力は高い値を示した。シナリオ 3 及び 4 は、各投与群の効果の平均構造に差があるモデルが真であり、標本数が増加したため差を検出しやすくなっている。

また、1 群 10 例の検討結果では、LR が他の 4 種類の手法と比較して明確に高い検出力を示していた。1 群 20 例の本検討でもその傾向は認められるが、その差は明確ではなくなっている。これも標本数が増加したため、LR 以外の手法でも真の差を検出しやすくなっているためと考えられた。

また、シナリオ 5 及び 6 でも  $t$ -test を除外すれば、すべての方法でほぼ同等の検出力が得られており、方法間で大きな差は認められなかった。

表 3.5-16 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 20 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
3	LR <sub>Bart</sub>	30.80	26.15	21.53	30.80	26.25	18.32
	LR <sub>Boot</sub>	30.80	26.65	22.13	30.80	26.75	18.92
	LR	33.60	29.96	26.46	33.60	29.26	22.52
	MMRM <sub>KR</sub>	31.70	26.40	21.60	31.70	25.30	18.10
	$t$ -test	31.70	25.00	19.70	31.70	20.70	14.40
4	LR <sub>Bart</sub>	30.80	26.03	21.66	30.80	25.53	17.70
	LR <sub>Boot</sub>	31.10	26.63	22.47	31.10	25.83	17.70
	LR	33.70	29.83	26.38	33.70	28.53	23.20
	MMRM <sub>KR</sub>	31.70	26.40	21.60	31.70	25.40	17.80
	$t$ -test	31.70	25.00	19.70	31.70	20.00	13.10
5	LR <sub>Bart</sub>	99.40	98.40	91.80	99.40	97.90	90.60
	LR <sub>Boot</sub>	99.40	98.50	91.80	99.40	97.90	90.80
	LR	99.50	98.60	93.60	99.50	98.20	94.40
	MMRM <sub>KR</sub>	99.50	97.60	93.50	99.50	97.30	91.70
	$t$ -test	99.50	97.10	89.40	99.50	94.80	84.20
6	LR <sub>Bart</sub>	99.40	98.30	91.90	99.40	98.10	90.70
	LR <sub>Boot</sub>	99.40	98.40	92.00	99.40	98.10	91.00
	LR	99.50	98.60	93.60	99.50	98.30	93.80
	MMRM <sub>KR</sub>	99.50	97.60	91.10	99.50	97.30	91.50
	$t$ -test	99.50	97.10	89.40	99.50	95.20	83.80

<sup>†</sup>LR<sub>Bart</sub>, Bartlett 型補正による尤度比検定: LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定: LR, 従来の尤度比検定: MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定:  $t$ -test,  $t$ 検定

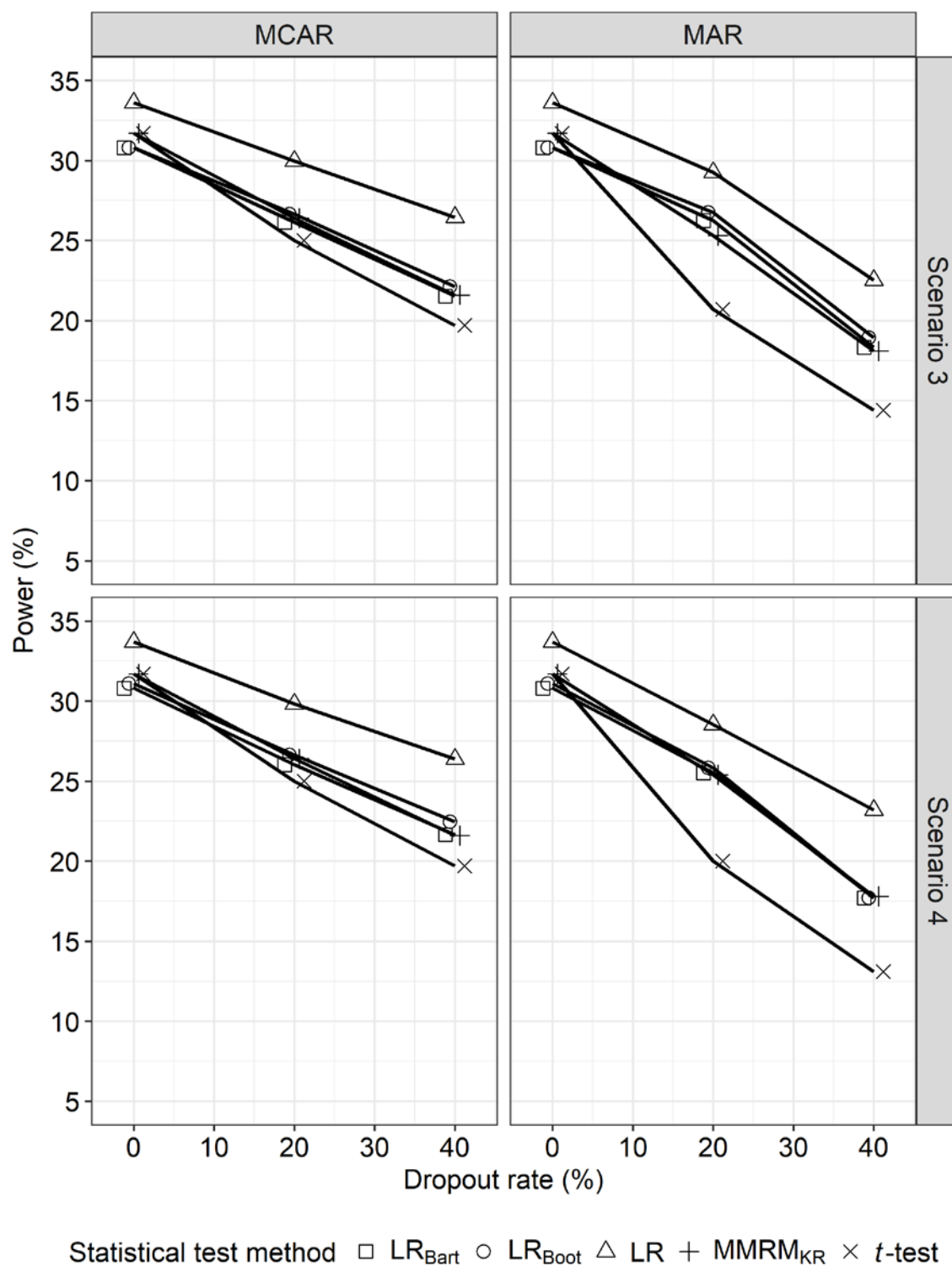


図 3.5-12 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ3及び4, MCAR及びMAR下, 1群20例)

<sup>†</sup> LR<sub>Bart</sub>, Bartlett補正による尤度比検定; LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定; LR, 従来の尤度比検定; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定;  $t$ -test,  $t$  検定

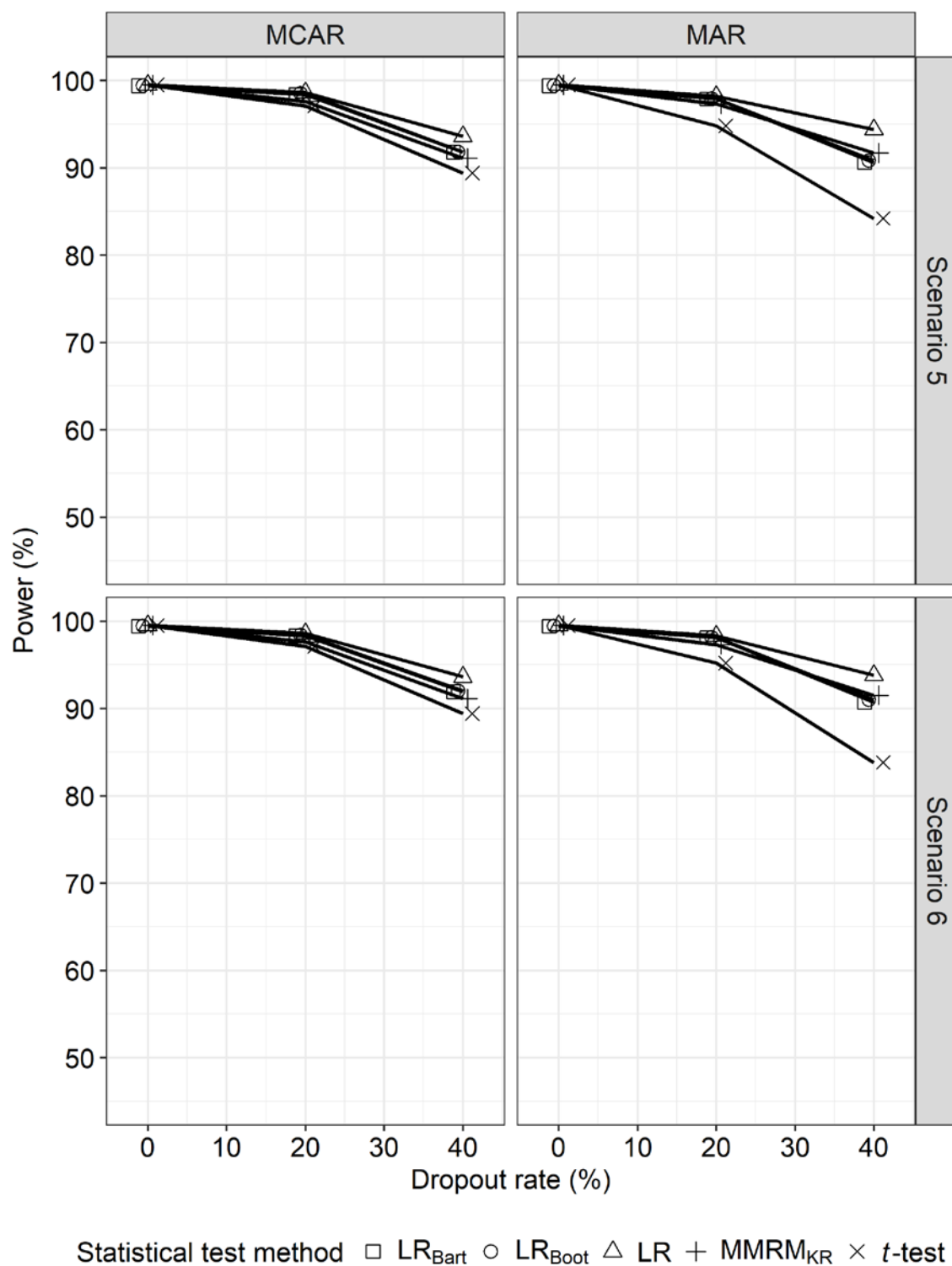


図 3.5-13 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 20 例)

<sup>†</sup>  $LR_{Bart}$ , Bartlett 補正による尤度比検定 ;  $LR_{Boot}$ , ブートストラップ検定に基づく尤度比検定 ;  $LR$ , 従来の尤度比検定 ;  $MMRM_{KR}$ , Kenward and Roger 法による REML 推定を用いた Wald 型検定 ;  $t$ -test,  $t$  検定

### 3.5.8.3 収束割合（シナリオ 1 及び 2）

$N = 40$ （1 群 20 例）におけるシナリオ 1 及び 2 の収束割合の結果を、表 3.5-17 及び図 3.5-14 に示す。標本数  $N = 40$ （1 群 20 例）では、欠測のない完全データであっても、累積欠測割合が高い場合であっても、ほぼ 100% の収束が得られた。

表 3.5-17 収束割合（%）のシミュレーション実験結果<sup>†</sup>（シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 20 例）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	LR <sub>Bart</sub>	100.0	99.6	98.7	100.0	99.7	99.8
	LR <sub>Boot</sub>	100.0	99.6	98.7	100.0	99.7	99.8
	LR	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0
2	LR <sub>Bart</sub>	100.0	99.3	98.9	100.0	99.7	99.9
	LR <sub>Boot</sub>	100.0	99.3	98.9	100.0	99.7	99.9
	LR	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0

<sup>†</sup>LR<sub>Bart</sub>, Bartlett 型補正による尤度比検定；LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定；LR, 従来の尤度比検定；MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定

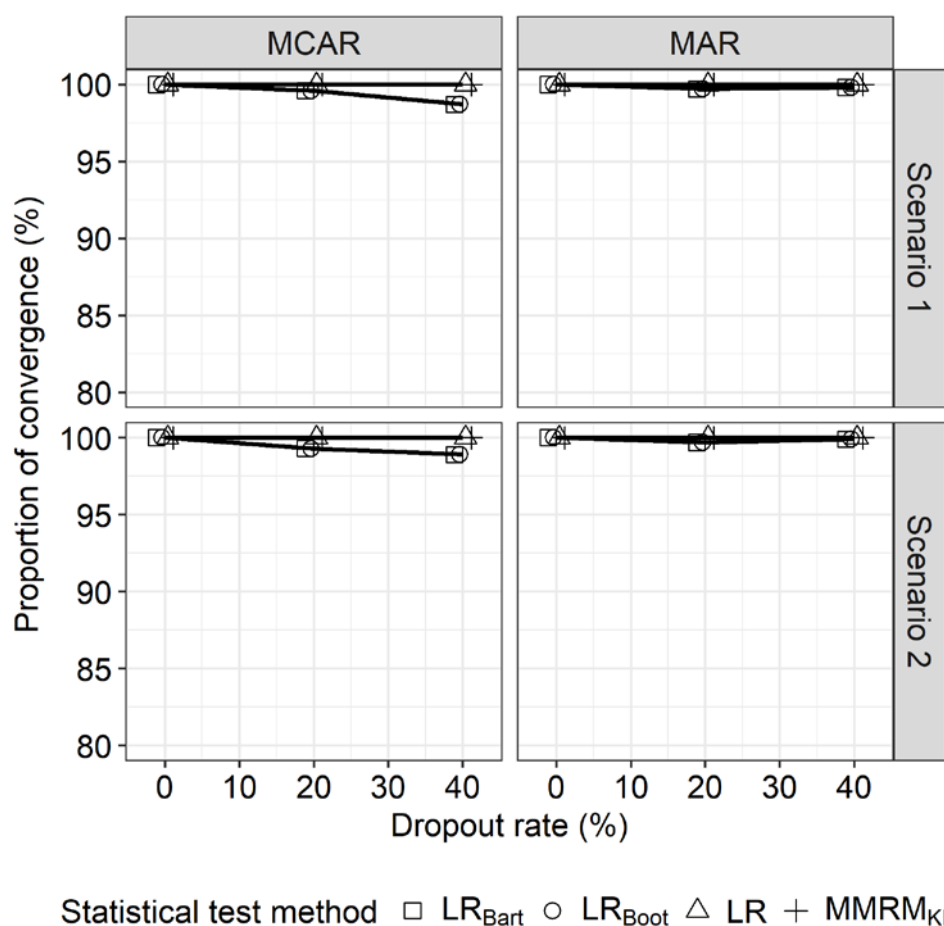


図 3.5-14 収束割合（%）のシミュレーション実験結果<sup>†</sup>（シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 20 例）

<sup>†</sup>LR<sub>Bart</sub>, Bartlett 補正による尤度比検定；LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定；LR, 従来の尤度比検定；MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定

### 3.5.8.4 収束割合（シナリオ 3, 4, 5 及び 6）

$N = 40$ （1 群 20 例）におけるシナリオ 3, 4, 5 及び 6 の収束割合の結果を、表 3.5-18、図 3.5-15 及び図 3.5-16 に示す。標本数 $N = 40$ （1 群 20 例）では、欠測のない完全データであっても、累積欠測割合が高い場合であっても、ほぼ 100% の収束が得られた。

表 3.5-18 収束割合（%）のシミュレーション実験結果<sup>†</sup>（シナリオ 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 20 例）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
3	LR <sub>Bart</sub>	100.0	99.8	99.4	100.0	99.8	99.9
	LR <sub>Boot</sub>	100.0	99.8	99.4	100.0	99.8	99.9
	LR	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0
4	LR <sub>Bart</sub>	100.0	99.9	99.7	100.0	99.9	100.0
	LR <sub>Boot</sub>	100.0	99.9	99.7	100.0	99.9	100.0
	LR	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0
5	LR <sub>Bart</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	LR <sub>Boot</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	LR	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0
6	LR <sub>Bart</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	LR <sub>Boot</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	LR	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0

<sup>†</sup>LR<sub>Bart</sub>, Bartlett 型補正による尤度比検定：LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定：LR, 従来の尤度比検定：MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定

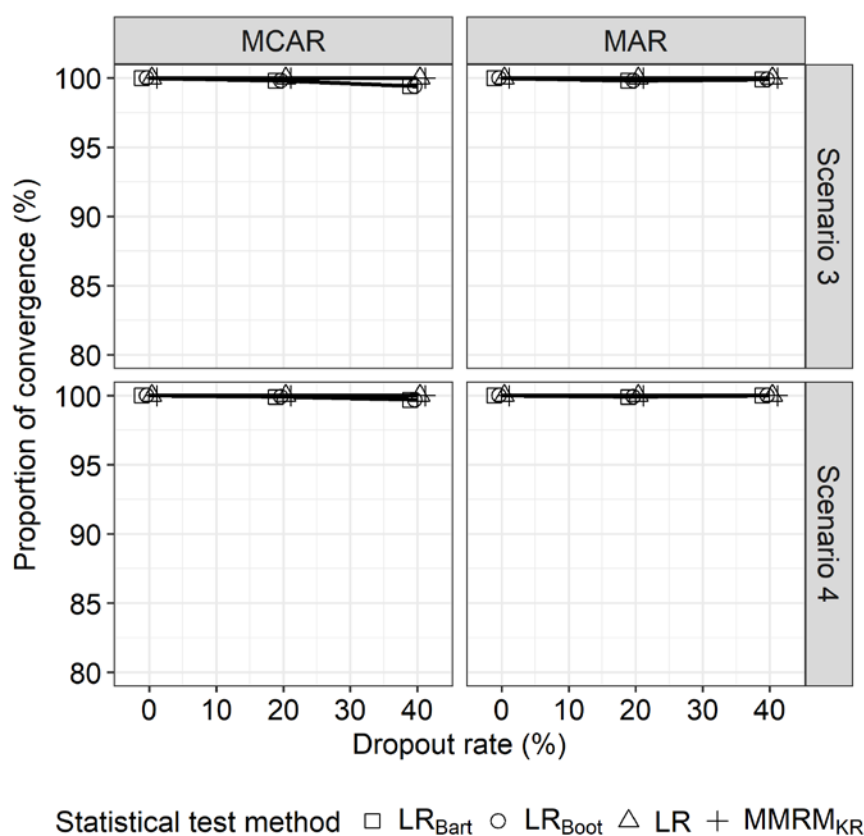


図 3.5-15 収束割合（％）のシミュレーション実験結果<sup>†</sup>（シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 20 例）

<sup>†</sup> LR<sub>Bart</sub>, Bartlett 補正による尤度比検定；LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定；LR, 従来の尤度比検定；MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定

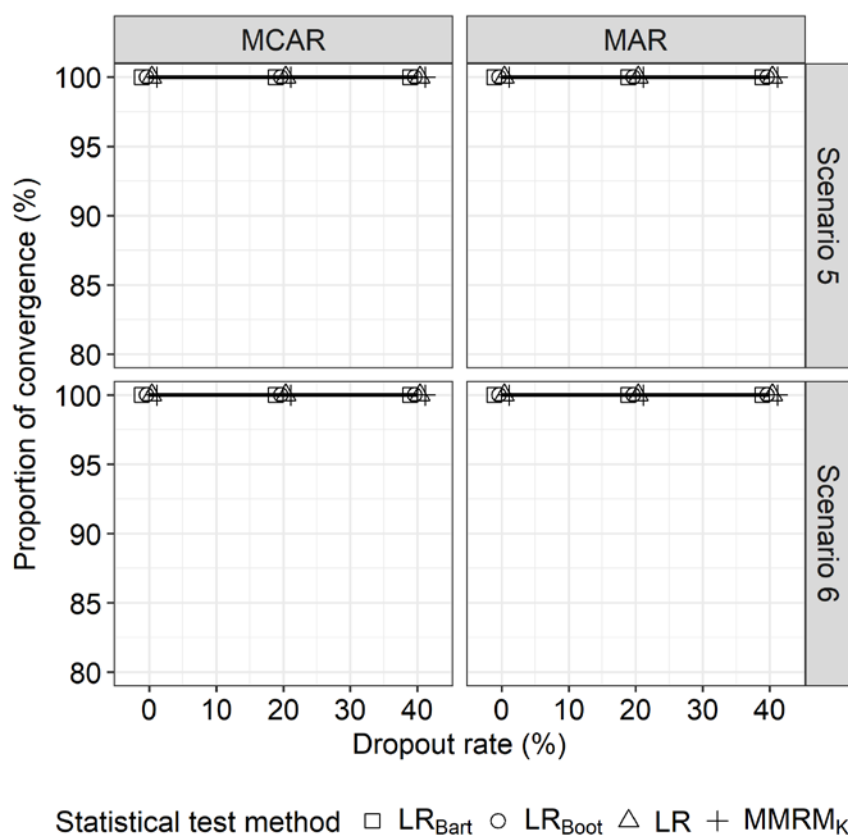


図 3.5-16 収束割合（％）のシミュレーション実験結果<sup>†</sup>（シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 20 例）

<sup>†</sup> LR<sub>Bart</sub>, Bartlett 補正による尤度比検定；LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定；LR, 従来の尤度比検定；MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定

### 3.6 補足的検討

#### 3.6.1 リサンプリング回数に関する検討

提案法ではブートストラップ法により帰無仮説下における尤度比検定統計量をリサンプリングしている。リサンプリング回数を増加させれば、アルゴリズム 1 で推定している Bartlett 補正項や、アルゴリズム 2 で利用する経験分布の推定精度が向上することは容易に予想できる。適切なリサンプリング回数について、Rocke は少なくとも 1000 回以上を推奨している (Rocke 1989)。さらに小西らは 500 回程度を目安として示している (小西, 越智, 大森 2008)。先行研究の報告を踏まえても、リサンプリング回数は概ね 1000 回以上程度あれば十分であろうと推測された。実際、本研究では 3000 回のリサンプリングとしており、十分な回数であると考えられた。

一方、ブートストラップ法は回数を増加させればその分だけ実行に時間を要するため、簡便さを求めるならばなるべくリサンプリング回数は少ない方がよい。

そこで、本セクションでは本研究のシナリオ 1 の完全データを想定した状況で、リサンプリング回数について検討した結果を示す。リサンプリング回数を 100, 200, 500, 1000, 2000 及び 3000 回としたときの Type I error rate を  $LR_{Bart}$  及び  $LR_{Boot}$  について算出した。反復回数は 1000 回と設定し、提案法の性能評価を実施した状況と同一条件とした。リサンプリング回数とは関係ないが、参考として同一条件下で LR についても Type I error rate を算出し参考値とした。表 3.6-1 及び図 3.6-1 に検討結果を示す。

表 3.6-1 リサンプリング回数と Type I error rate のシミュレーション実験結果<sup>†</sup> (シナリオ 1, 欠測なし)

標本数	方法 <sup>†</sup>	リサンプリング回数					
		100	200	500	1000	2000	3000
$N = 10$	$LR_{Bart}$	6.21	5.51	5.01	5.51	5.71	5.71
	$LR_{Boot}$	6.00	5.36	5.21	5.12	5.81	5.91
	LR	9.90	9.90	9.90	9.90	9.90	9.90
$N = 20$	$LR_{Bart}$	5.10	5.20	5.40	5.30	5.20	5.00
	$LR_{Boot}$	4.04	4.63	5.52	5.01	5.01	4.90
	LR	7.30	7.30	7.30	7.30	7.30	7.30
$N = 40$	$LR_{Bart}$	5.70	5.20	5.50	5.00	5.40	4.80
	$LR_{Boot}$	5.76	5.43	5.62	5.21	5.41	4.80
	LR	5.70	5.70	5.70	5.70	5.70	5.70

<sup>†</sup> $LR_{Bart}$ , Bartlett 型補正による尤度比検定:  $LR_{Boot}$ , ブートストラップ検定に基づく尤度比検定: LR, 従来の尤度比検定

シミュレーション結果より、500 回以上のリサンプリング回数であれば提案法は安定して Type I error rate が理論上の 5% に近い値となることが示された。なお、本研究で想定する状況は小標本かつ欠測が発生している経時測定データであり、安定した結果を与えるリサンプリング回数として 3000 回の設定は十分であると考えられた。

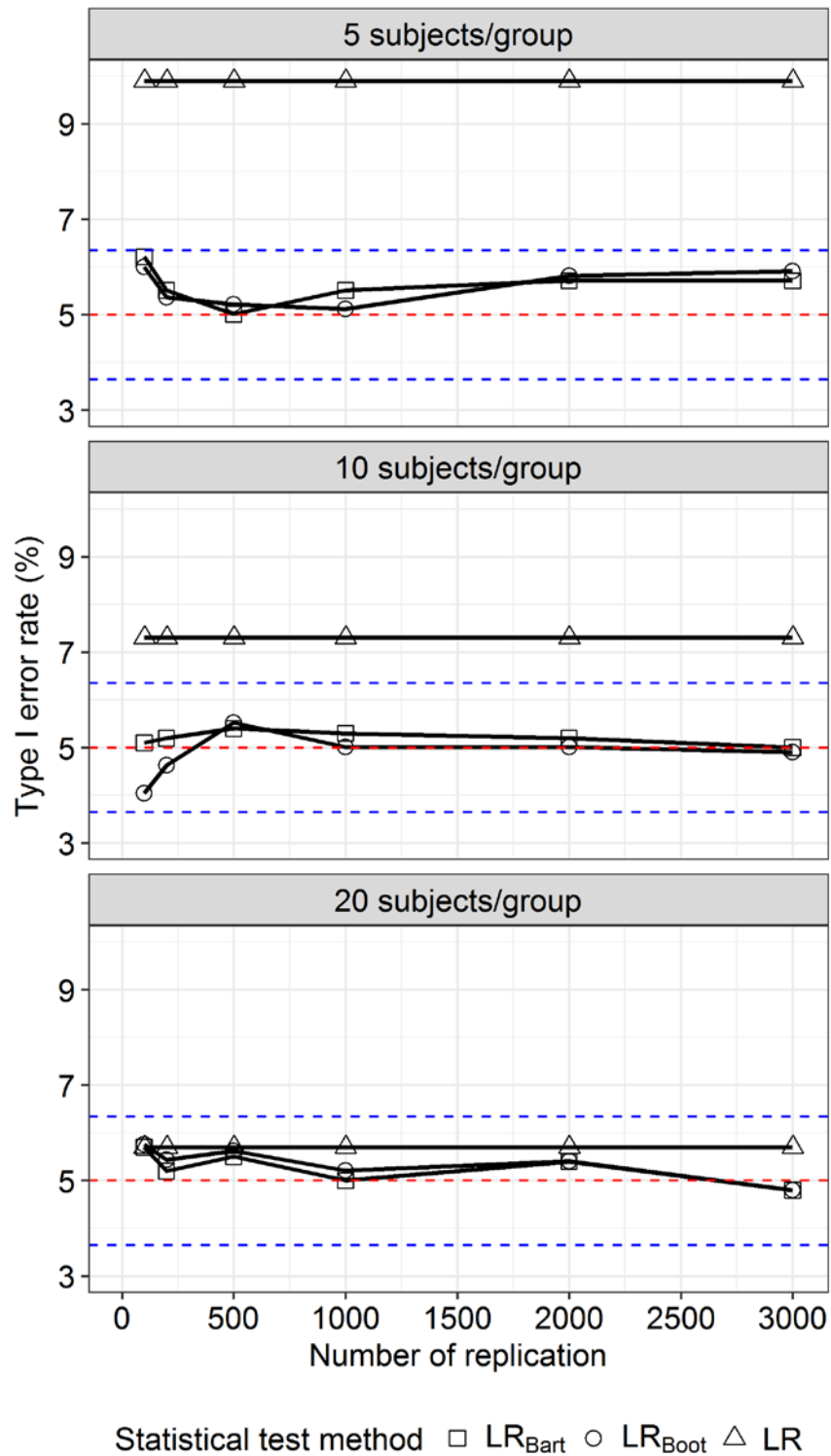


図 3.6-1 リサンプリング回数と Type I error rate のシミュレーション実験結果<sup>†</sup> (シナリオ 1, 欠測なし)

赤点線, 5%; 青点線, 95%信頼限界 (真の Type I error rate を 5% と仮定した場合の 1000 回反復によるモンテカルロ誤差); <sup>†</sup> LR<sub>Bart</sub>, Bartlett 補正による尤度比検定; LR<sub>Boot</sub>, ブートストラップ検定に基づく尤度比検定; LR, 従来の尤度比検定



### 3.6.2 ブートストラップ法による Bartlett 補正項の推定値

提案法 1 では、ブートストラップ法により帰無仮説下の尤度比検定統計量を 3000 回リサンプリングし、その期待値を求めて Bartlett 補正項を推定した。Type I error rate に関するシミュレーション実験結果で前述したように、小標本かつ欠測が発生する状況下においても提案法 1 は Type I error rate を 5%水準に維持していた。

本研究で設定している状況では、尤度比検定統計量は自由度 1 のカイ 2 乗分布に従うはずである。小標本かつ欠測が発生する状況下で構成される尤度比検定統計量の分布について確認し、ブートストラップ法を用いた提案法の有用性について検討する。

シミュレーションで推定した Bartlett 補正項について、平均値と分位点による 95% 区間 (2.5% 点及び 97.5% 点) を求めた (表 3.6-2)。Bartlett 補正項の分布を明らかにするために、推定した Bartlett 補正項についてヒストグラムを作成した。ヒストグラムについては付録に掲載した。さらに、Bartlett 補正項の分布の特徴量として、平均値と分位点 (2.5% 及び 97.5% 点) をプロットした (図 3.6-2, 図 3.6-3, 図 3.6-4, 図 3.6-5)。

図の番号，対象シナリオ及び 1 群当たりの例数の対応表

図の番号	対象シナリオ	1 群当たりの例数
図 3.6-2	1, 2, 3, 4	10
図 3.6-3	5, 6	10
図 3.6-4	1, 2, 3, 4	20
図 3.6-5	5, 6	20

表 3.6-2 Bartlett 補正項推定値の平均とその分位点（2.5%及び 97.5%点）のシミュレーション実験結果

標本数/ シナリオ	累積欠測割合					
	MCAR			MAR		
	0%	20%	40%	0%	20%	40%
5 例/群						
1	1.33 [1.23, 1.40]	---	---	1.33 [1.23, 1.40]	---	---
2	1.33 [1.23, 1.41]	---	---	1.33 [1.23, 1.41]	---	---
3	1.33 [1.23, 1.41]	---	---	1.33 [1.23, 1.41]	---	---
4	1.33 [1.23, 1.41]	---	---	1.33 [1.23, 1.41]	---	---
5	1.32 [1.20, 1.41]	---	---	1.32 [1.20, 1.41]	---	---
6	1.32 [1.18, 1.41]	---	---	1.32 [1.18, 1.41]	---	---
10 例/群						
1	1.19 [1.15, 1.22]	1.27 [1.17, 1.42]	1.38 [1.23, 1.65]	1.19 [1.15, 1.22]	1.27 [1.17, 1.41]	1.36 [1.22, 1.56]
2	1.19 [1.15, 1.22]	1.27 [1.17, 1.41]	1.37 [1.22, 1.60]	1.19 [1.15, 1.22]	1.27 [1.17, 1.41]	1.36 [1.22, 1.56]
3	1.19 [1.15, 1.22]	1.27 [1.18, 1.43]	1.38 [1.23, 1.63]	1.19 [1.15, 1.22]	1.27 [1.17, 1.41]	1.37 [1.22, 1.59]
4	1.19 [1.15, 1.22]	1.27 [1.17, 1.43]	1.38 [1.23, 1.63]	1.19 [1.15, 1.22]	1.27 [1.17, 1.40]	1.37 [1.23, 1.57]
5	1.19 [1.16, 1.21]	1.27 [1.18, 1.41]	1.37 [1.23, 1.59]	1.19 [1.16, 1.21]	1.26 [1.17, 1.39]	1.36 [1.22, 1.57]
6	1.19 [1.16, 1.21]	1.27 [1.18, 1.41]	1.37 [1.24, 1.59]	1.19 [1.16, 1.21]	1.26 [1.17, 1.38]	1.35 [1.22, 1.56]
20 例/群						
1	1.07 [1.05, 1.09]	1.10 [1.04, 1.17]	1.20 [1.09, 1.36]	1.07 [1.05, 1.09]	1.10 [1.05, 1.18]	1.19 [1.08, 1.33]
2	1.07 [1.05, 1.09]	1.10 [1.04, 1.16]	1.18 [1.09, 1.33]	1.07 [1.05, 1.09]	1.10 [1.05, 1.18]	1.19 [1.08, 1.34]
3	1.07 [1.05, 1.09]	1.10 [1.04, 1.17]	1.20 [1.09, 1.39]	1.07 [1.05, 1.09]	1.10 [1.04, 1.18]	1.20 [1.09, 1.34]
4	1.07 [1.05, 1.09]	1.10 [1.05, 1.16]	1.20 [1.09, 1.38]	1.07 [1.05, 1.09]	1.10 [1.04, 1.18]	1.20 [1.09, 1.34]
5	1.07 [1.06, 1.09]	1.10 [1.05, 1.18]	1.20 [1.09, 1.36]	1.07 [1.06, 1.09]	1.10 [1.05, 1.16]	1.18 [1.08, 1.31]
6	1.07 [1.06, 1.09]	1.10 [1.05, 1.17]	1.20 [1.10, 1.36]	1.07 [1.06, 1.09]	1.10 [1.05, 1.16]	1.17 [1.08, 1.30]

Bartlett 補正項推定値の平均値とその分位点[2.5%点, 97.5%点]

### 3.6.2.1 Bartlett 補正項推定値の平均値とその分位点（2.5%及び 97.5%点）のプロット

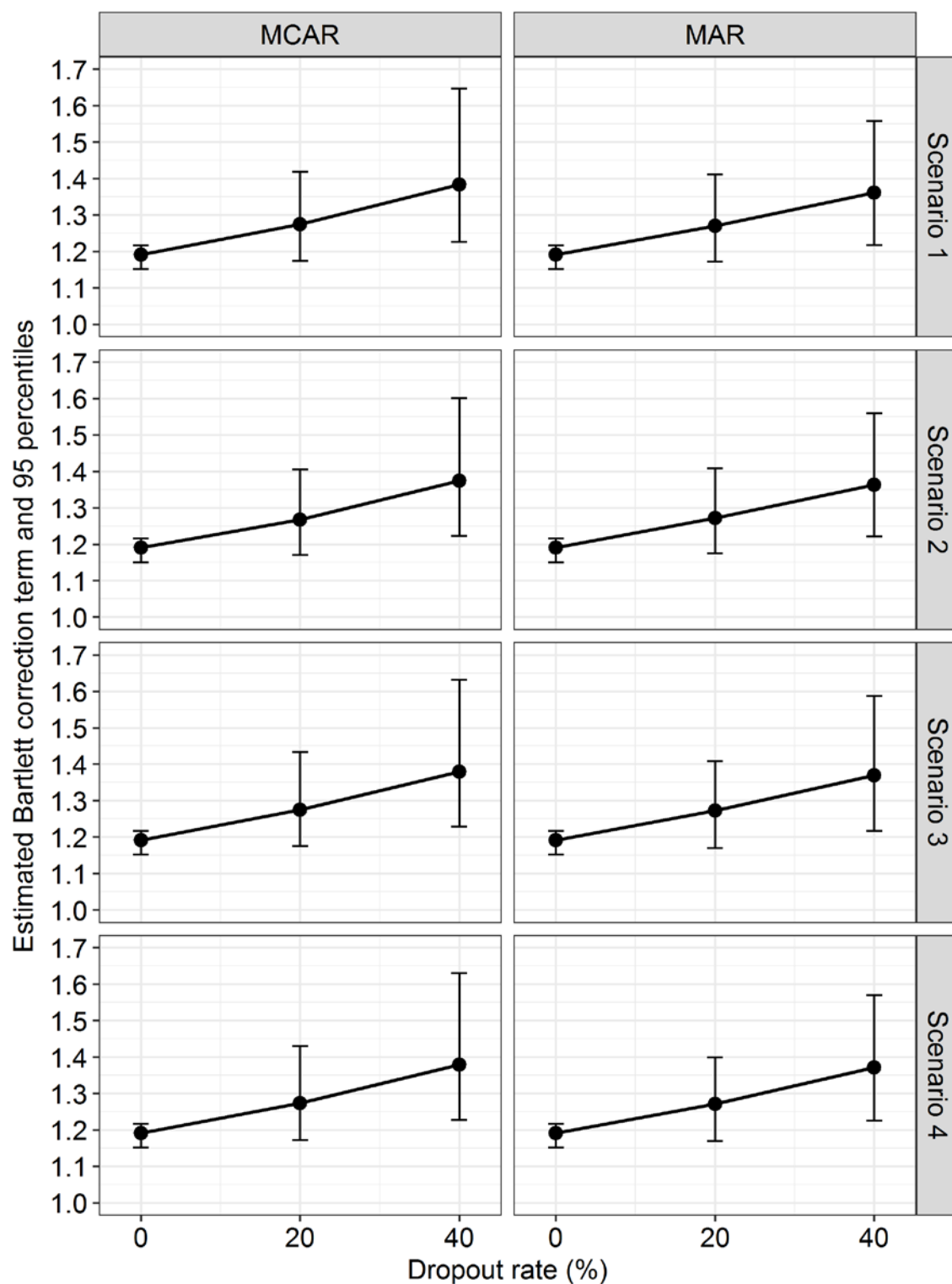


図 3.6-2 Bartlett 補正項推定値の平均とその分位点（2.5%及び 97.5%点）のシミュレーション実験結果（シナリオ 1, 2, 3 及び 4, MCAR 及び MAR 下, 1 群 10 例）

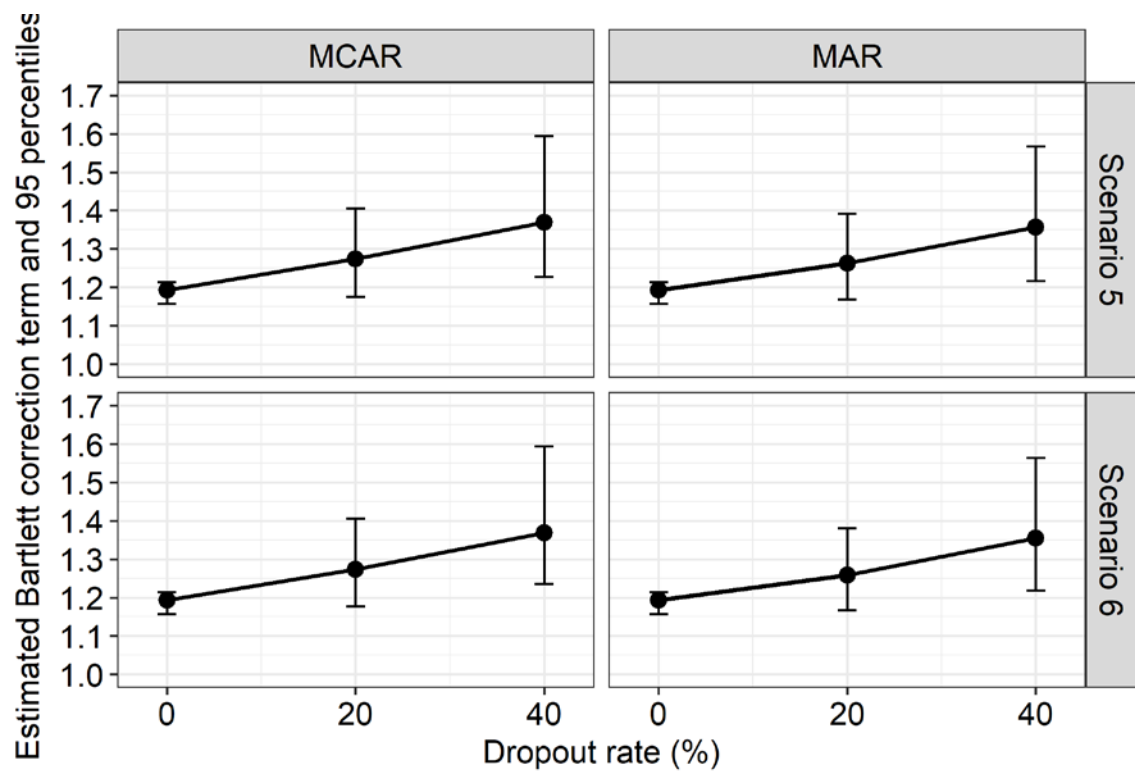


図 3.6-3 Bartlett 補正項推定値の平均とその分位点（2.5%及び 97.5%点）のシミュレーション実験結果（シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 10 例）

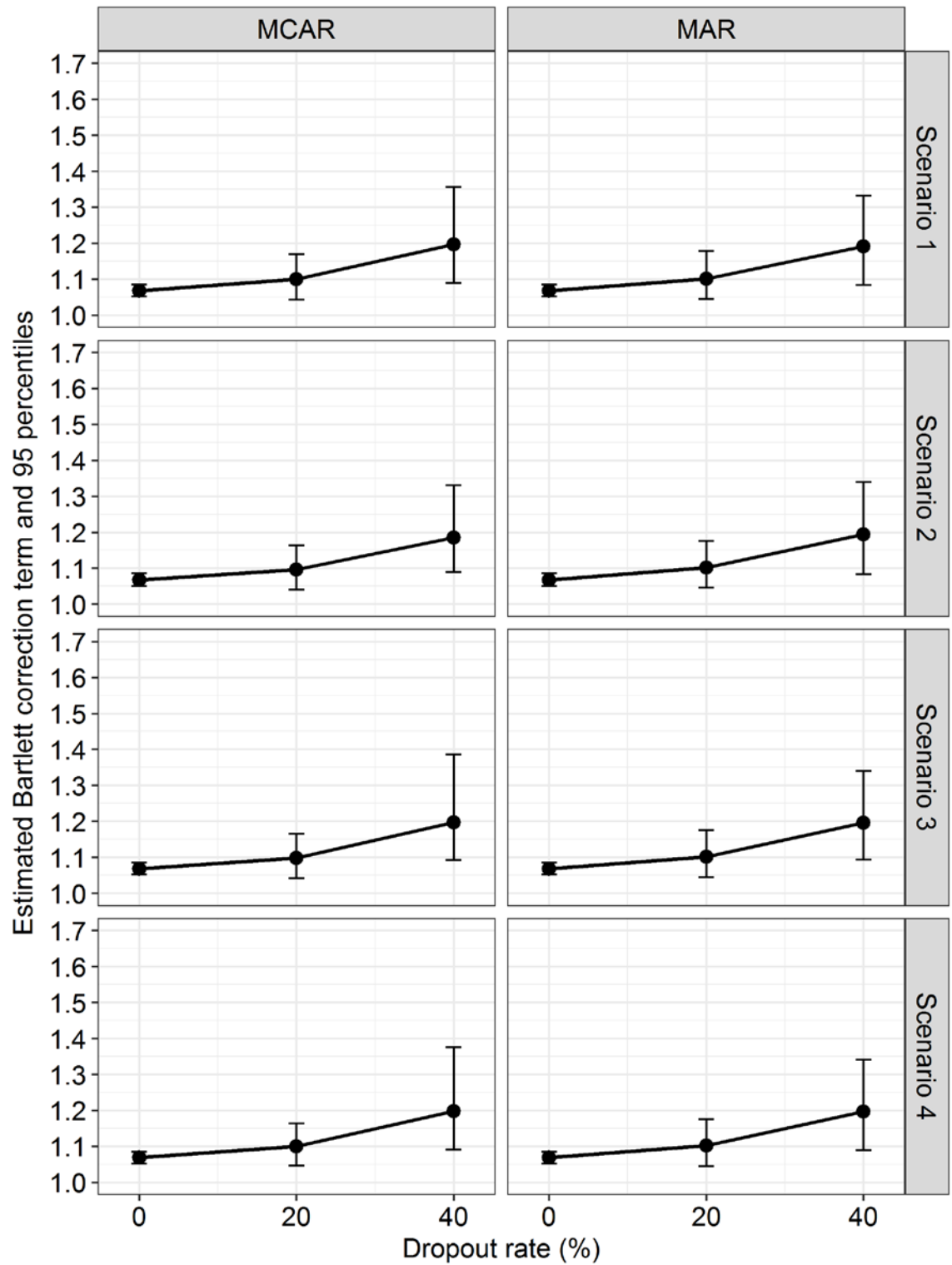


図 3.6-4 Bartlett 補正項推定値の平均とその分位点（2.5%及び 97.5%点）のシミュレーション実験結果（シナリオ 1, 2, 3 及び 4, MCAR 及び MAR 下, 1 群 20 例）

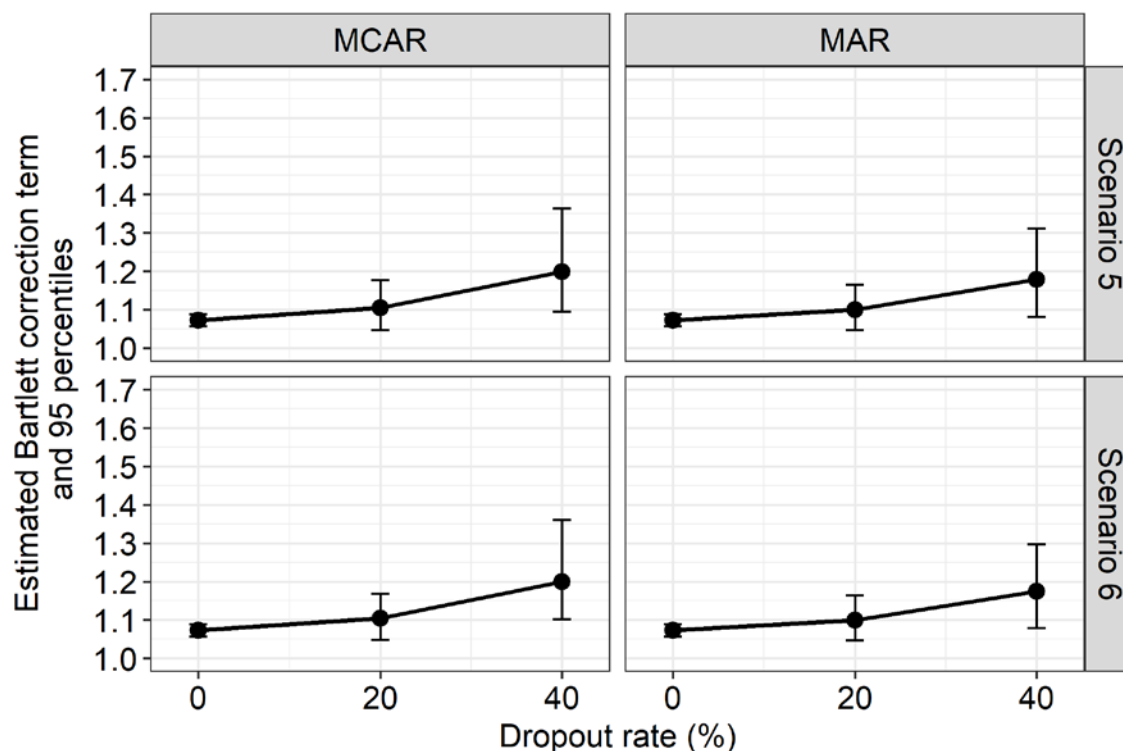


図 3.6-5 Bartlett 補正項推定値の平均とその分位点（2.5%及び 97.5%点）のシミュレーション実験結果（シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 20 例）

### 3.6.2.2 ブートストラップ法による Bartlett 補正項の推定値に関するまとめ

Bartlett 補正項の分布について、2 種類の観点から結果をまとめる。

第一に、同一対象者数の中で累積欠測割合別の Bartlett 補正項の分布について触れる。同一対象者数の中では累積欠測割合が増加すると、分布が広がる様子が明らかとなった。例えば、 $N = 20$ （1 群 10 例）でシナリオ 1 の MCAR 下では、Bartlett 補正項の分位点による 95% 区間は、累積欠測割合 0%, 20% 及び 40% でそれぞれ [1.15, 1.22], [1.17, 1.42], [1.23, 1.65] であった。同様に  $N = 20$ （1 群 10 例）でシナリオ 1 の MAR 下では、累積欠測割合 0%, 20% 及び 40% でそれぞれ [1.15, 1.22], [1.17, 1.41], [1.22, 1.56] であった。その他のシナリオでも同様な傾向が認められた。さらに  $N = 40$ （1 群 20 例）でも同様な傾向となっている。

次に、同一シナリオ及び同一累積欠測割合において対象者数別に Bartlett 補正項の分布について見てみる。対象者数が増加すると、分布が狭くなった。例えば、シナリオ 1 の MCAR 下における累積欠測割合 0% で 95% 区間は、 $N = 20$ （1 群 10 例）のとき [1.15, 1.22] であり、 $N = 40$ （1 群 20 例）のとき [1.05, 1.09] であった。その他のシナリオでも同様な傾向が認められた。

本検討から、解析対象となる対象者数が少ない場合、つまり小標本である場合やデータの欠測が多くなる場合では、尤度比検定統計量が従う自由度 1 のカイ 2 乗分布が平均 1 から大きくずれていることが分かった。また、ずれる程度を定量的に評価した。

### 3.6.3 Kenward and Roger 法の性能に関する検討

$N = 20$ （1 群 10 例）において、欠測メカニズムが MAR の下で、Type I error rate についてシミュレーションで検討した結果が、表 3.5-11 及び図 3.5-5 に示されている。Kenward and Roger 法（MMRM<sub>KR</sub>）は、累積欠測割合が増加すると過度に保守的な Type I error rate を示した。

Kenward and Roger（KR）法は、漸近理論が成立しづらい小標本下における対処法として、 $t$  検定における自由度を補正するだけでなく、固定効果の分散の推定値も補正する方法である。一般に KR 法は小標本下での性能が高いため、これに反する結果には説明が必要と考えられた。Type I error rate について KR 法が過度に保守的な結果となる理由は、自由度もしくは固定効果の分散の推定値の補正に問題があることで説明ができるはずである。そこで、この 2 点に着目して検討を進めることとした。

KR 法による自由度は、本シミュレーション設定では Satterthwaite 近似に一致する（Kenward and Roger 1997; Verbeke and Molenberghs 2009）。ただし、固定効果の分散の推定値は異なる。ここで、固定効果を REML 推定し自由度調整を Satterthwaite（Satterthwaite 1946; SAS Institute Inc. 2017）に変更した方法（MMRM<sub>SAT</sub>）を追加で設定した。1997 年に提案された KR 法は、固定効果の共分散構造のパラメータ設定が非線形である場合に、固定効果の分散に関する推定が適切に補正されないことが指摘されている。Kenward と Roger は補正時に適用していた近似を改善した方法を提案した（Kenward and Roger 2009; SAS Institute Inc. 2017）。本検討では KR2（MMRM<sub>KR2</sub>）として設定した。ただし、KR2 の結果は KR の結果と全く同様であったため、シミュレーション結果の掲載は省略した。また、経時測定データの解析における自由度調整法として標準的に使用される MMRM における自由度調整として Between-Within（BW）法（Schluchter and Elashoff 1990）が推奨されているため、検討に加えた。

評価指標は次に述べる 4 種類とする。シナリオ 1 及び 2 における Type I error rate（シナリオ 3 及び 4 における検出力）と収束割合を評価することに加えて、各方法における固定効果の標準誤差及び  $t$  検定の自由度を求め KR 法を評価することとした。

シミュレーション回数は、提案法の反復に相当な時間を要することから 1000 回としていた。問題となっている状況での結果を確認するため、まず 1000 回反復を実施しその上で結果の精度を向上させる目的で反復回数を 10 倍にした 10000 回反復も実施することとした（結果は付録に掲載）。なお、乱数発生時のシードは提案法及び比較方法の性能を評価したシミュレーションで使用したシードと同一とした。

#### 3.6.3.1 Type I error rate 及び検出力

1 群 5 例、10 例、20 例のパターンでそれぞれ Type I error rate 及び検出力を評価した。1 群 5 例の場合には欠測が発生すると極端に収束状況が悪化してしまうため、欠測がない状況下のみ検討結果を表示した。反復回数 1000 回の結果を本文に掲載した。反復回数を 10000 回とした場合についても検討し同様の結果を得ている。結果は付録に参考として掲載した。

表 3.6-3 Type I error rate (%)及び検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 1, 2, 3, 4, 5 及び 6, 欠測なし, 1 群 5 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	4.60	---	---	4.60	---	---
	MMRM <sub>KR</sub>	4.61	---	---	4.61	---	---
	MMRM <sub>BW</sub>	4.60	---	---	4.60	---	---
2	MMRM <sub>SAT</sub>	4.60	---	---	4.60	---	---
	MMRM <sub>KR</sub>	4.72	---	---	4.72	---	---
	MMRM <sub>BW</sub>	4.60	---	---	4.60	---	---
3	MMRM <sub>SAT</sub>	10.60	---	---	10.60	---	---
	MMRM <sub>KR</sub>	10.63	---	---	10.63	---	---
	MMRM <sub>BW</sub>	10.60	---	---	10.60	---	---
4	MMRM <sub>SAT</sub>	10.60	---	---	10.60	---	---
	MMRM <sub>KR</sub>	10.75	---	---	10.75	---	---
	MMRM <sub>BW</sub>	10.60	---	---	10.60	---	---
5	MMRM <sub>SAT</sub>	50.70	---	---	50.70	---	---
	MMRM <sub>KR</sub>	50.45	---	---	50.45	---	---
	MMRM <sub>BW</sub>	50.70	---	---	50.70	---	---
6	MMRM <sub>SAT</sub>	50.70	---	---	50.70	---	---
	MMRM <sub>KR</sub>	50.50	---	---	50.50	---	---
	MMRM <sub>BW</sub>	50.70	---	---	50.70	---	---

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

表 3.6-4 Type I error rate (%)及び検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 10 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	5.10	5.90	8.92	5.10	4.10	5.26
	MMRM <sub>KR</sub>	5.10	5.10	5.69	5.10	3.40	3.61
	MMRM <sub>BW</sub>	5.10	6.40	10.01	5.10	5.00	8.47
2	MMRM <sub>SAT</sub>	5.10	5.50	8.37	5.10	4.00	5.54
	MMRM <sub>KR</sub>	5.10	4.70	5.68	5.10	3.30	3.34
	MMRM <sub>BW</sub>	5.10	6.10	9.31	5.10	5.01	8.36
3	MMRM <sub>SAT</sub>	16.20	14.40	14.33	16.20	13.40	10.07
	MMRM <sub>KR</sub>	16.20	13.20	11.71	16.20	12.00	6.64
	MMRM <sub>BW</sub>	16.20	14.90	15.85	16.20	15.30	13.35
4	MMRM <sub>SAT</sub>	16.20	14.30	14.32	16.20	13.31	10.88
	MMRM <sub>KR</sub>	16.20	13.20	11.67	16.20	11.71	7.29
	MMRM <sub>BW</sub>	16.20	14.90	15.84	16.20	15.42	13.77
5	MMRM <sub>SAT</sub>	87.10	76.50	63.71	87.10	76.38	57.98
	MMRM <sub>KR</sub>	87.10	74.10	58.82	87.10	74.77	51.01
	MMRM <sub>BW</sub>	87.10	77.20	66.85	87.10	79.08	64.79
6	MMRM <sub>SAT</sub>	87.10	76.50	63.81	87.10	76.20	57.72
	MMRM <sub>KR</sub>	87.10	74.10	58.84	87.10	74.50	50.43
	MMRM <sub>BW</sub>	87.10	77.20	66.96	87.10	78.50	64.22

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定



表 3.6-5 Type I error rate (%)及び検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 20 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	4.50	4.90	6.40	4.50	5.50	6.60
	MMRM <sub>KR</sub>	4.50	4.70	5.00	4.50	5.10	4.40
	MMRM <sub>BW</sub>	4.50	4.90	7.20	4.50	6.20	7.80
2	MMRM <sub>SAT</sub>	4.50	5.70	7.20	4.50	5.50	6.80
	MMRM <sub>KR</sub>	4.50	5.30	6.10	4.50	4.80	5.30
	MMRM <sub>BW</sub>	4.50	5.70	7.60	4.50	5.80	7.40
3	MMRM <sub>SAT</sub>	31.70	27.20	24.40	31.70	26.70	21.30
	MMRM <sub>KR</sub>	31.70	26.40	21.60	31.70	25.30	18.10
	MMRM <sub>BW</sub>	31.70	27.50	24.90	31.70	27.60	23.60
4	MMRM <sub>SAT</sub>	31.70	27.20	24.40	31.70	26.90	20.60
	MMRM <sub>KR</sub>	31.70	26.40	21.60	31.70	25.40	17.80
	MMRM <sub>BW</sub>	31.70	27.50	24.90	31.70	27.80	22.90
5	MMRM <sub>SAT</sub>	99.50	97.70	92.20	99.50	97.50	92.60
	MMRM <sub>KR</sub>	99.50	97.60	91.10	99.50	97.30	91.70
	MMRM <sub>BW</sub>	99.50	97.80	92.50	99.50	97.80	94.20
6	MMRM <sub>SAT</sub>	99.50	97.70	92.20	99.50	97.50	92.80
	MMRM <sub>KR</sub>	99.50	97.60	91.10	99.50	97.30	91.50
	MMRM <sub>BW</sub>	99.50	97.80	92.50	99.50	98.10	93.90

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

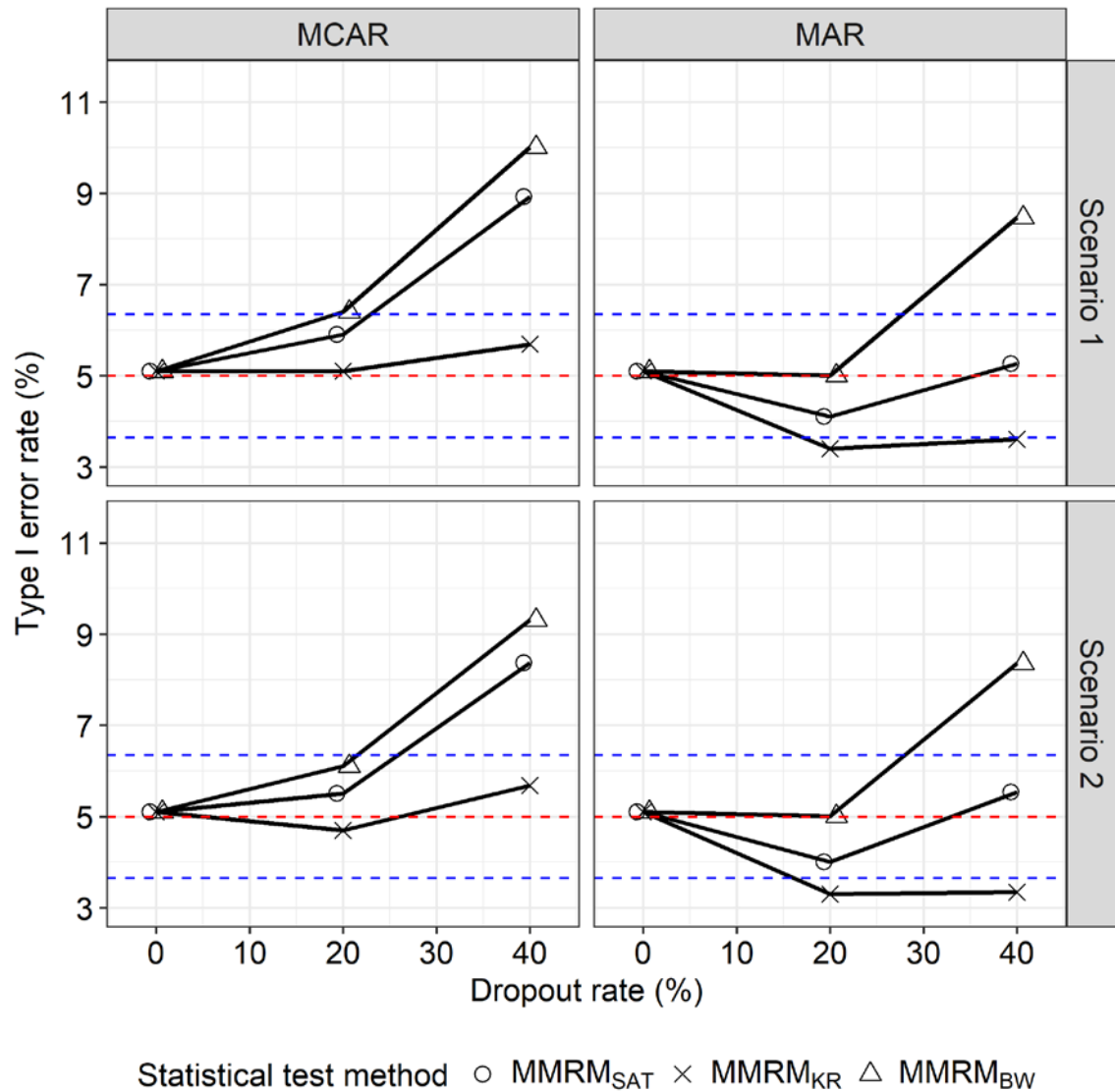


図 3.6-6 Type I error rate (%)のシミュレーション実験結果<sup>†</sup> (シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 10 例)

赤点線, 5% ; 青点線, 95%信頼限界 (真の Type I error rate を 5% と仮定した場合の 1000 回反復によるモンテカルロ誤差) ; <sup>†</sup> MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

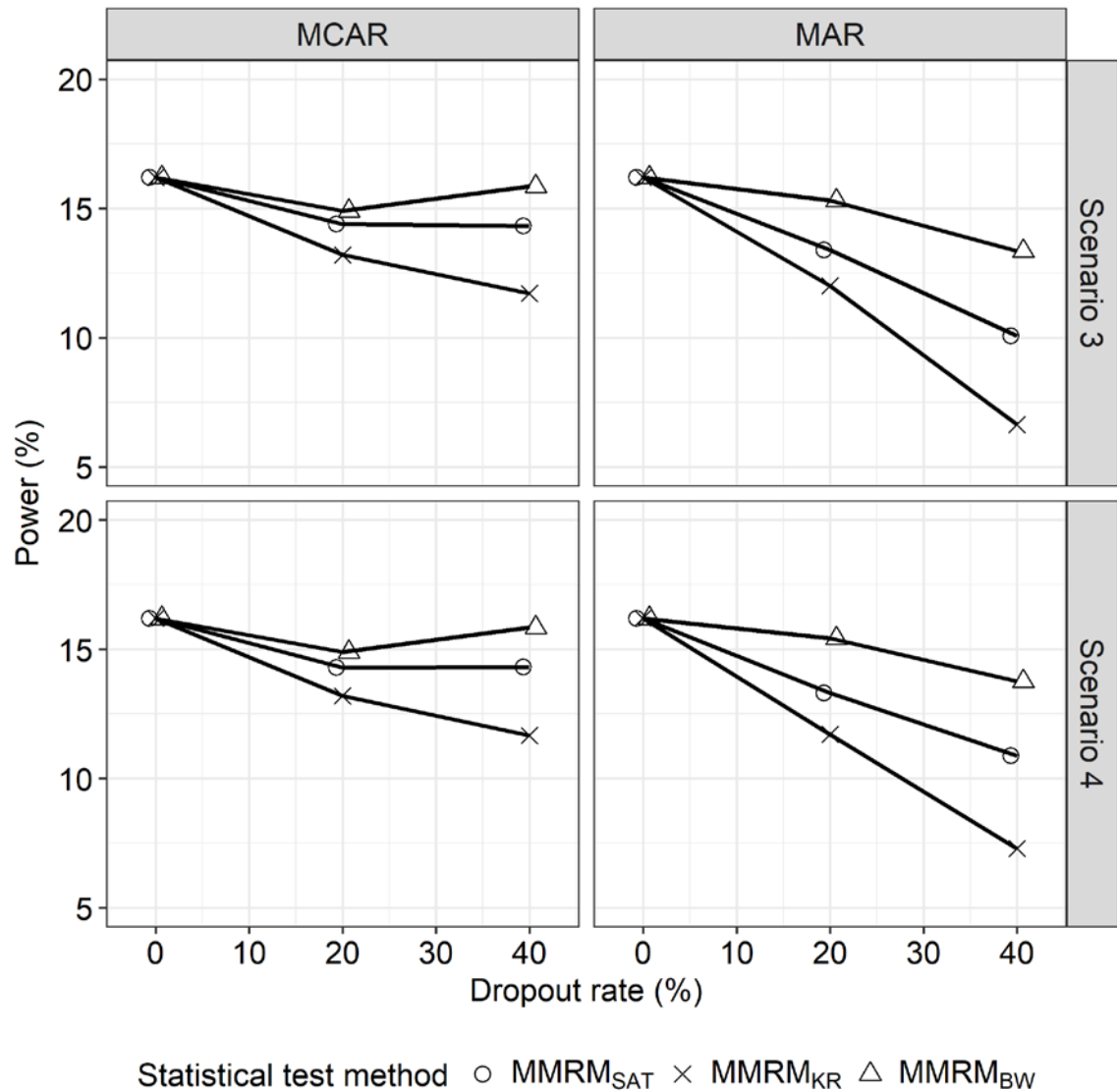


図 3.6-7 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 10 例)

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

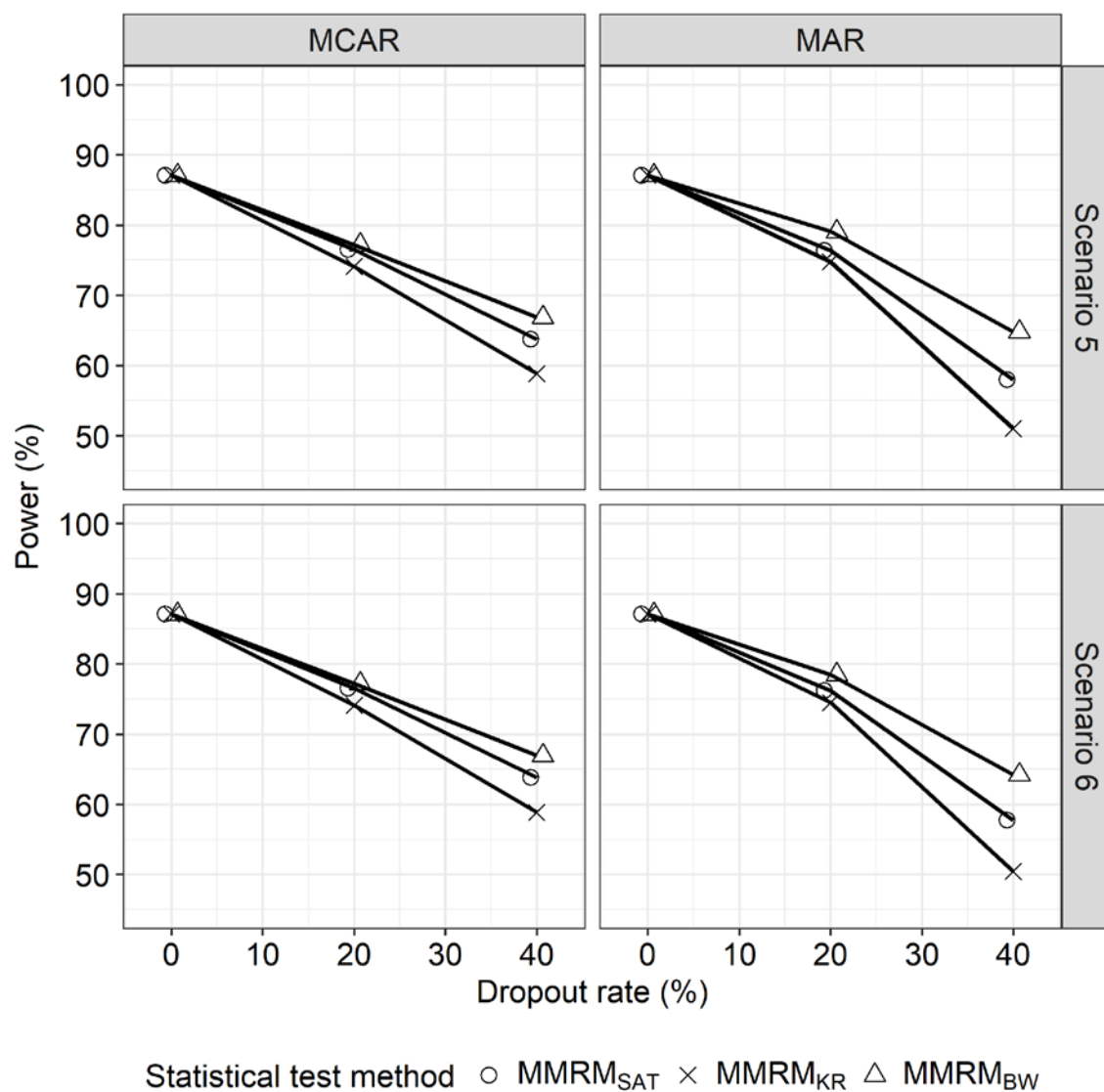


図 3.6-8 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 10 例)

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

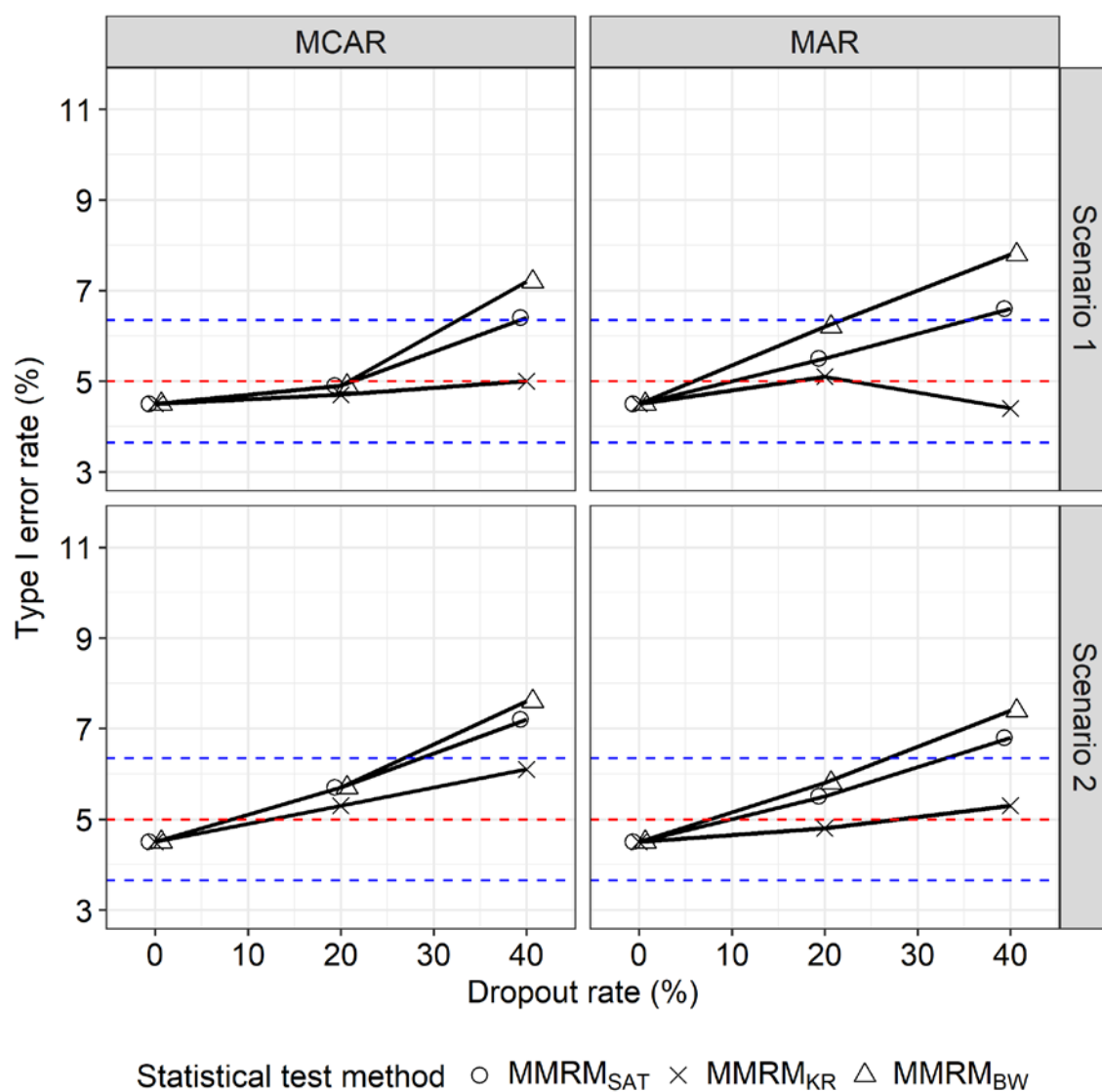


図 3.6-9 Type I error rate (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 20 例)

赤点線, 5% ; 青点線, 95%信頼限界 (真の Type I error rate を 5% と仮定した場合の 1000 回反復によるモンテカルロ誤差) ; <sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

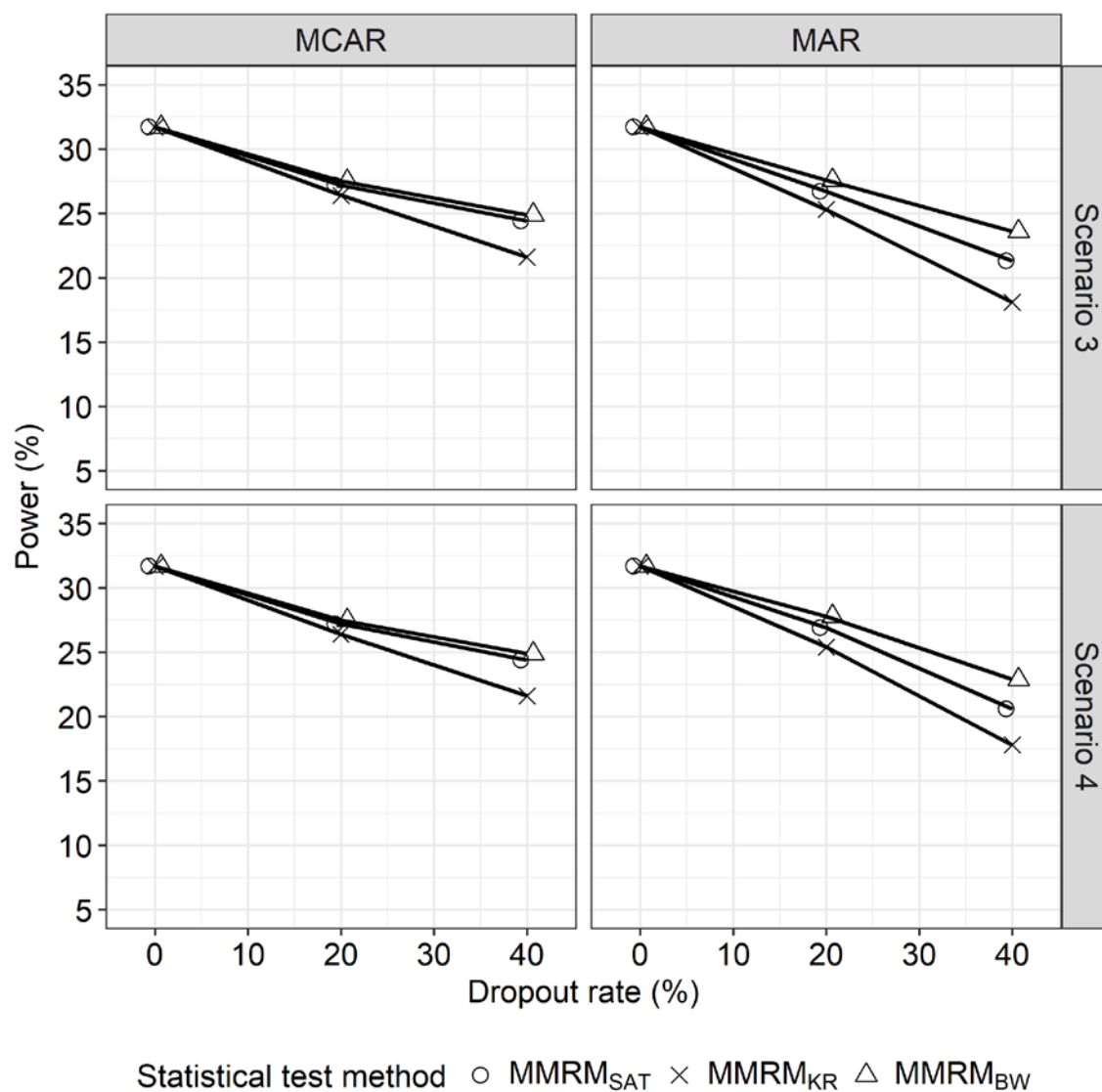


図 3.6-10 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 20 例)

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

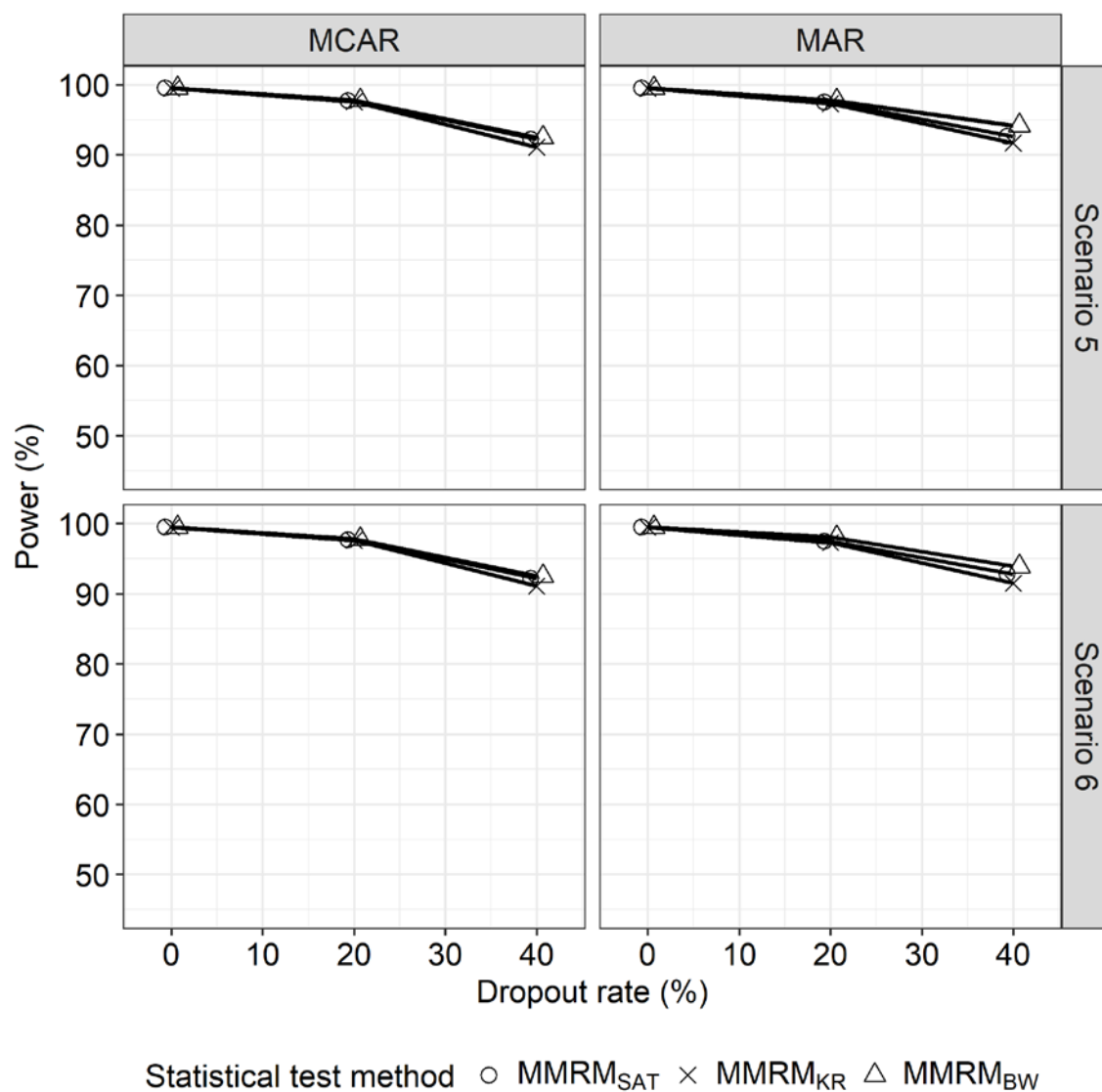


図 3.6-11 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 20 例)

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

### 3.6.3.2 収束割合

1 群 5 例, 10 例, 20 例のパターンでそれぞれ収束割合を評価した。1 群 5 例の場合には, 欠測が発生すると極端に収束状況が悪化してしまうため, 欠測がない状況下のみ検討結果を表示した。反復回数 1000 回の結果を本文に掲載した。反復回数を 10000 回とした場合についても検討し同様の結果を得ている。結果は付録に参考として掲載した。

1 群 5 例の場合には, 欠測がない状況 (累積欠測割合が 0%) であれば収束割合はすべての検討結果で 100%であった。1 群 10 例の場合, 図は付録に掲載した。また 1 群 20 例の場合には, 累積欠測割合 0%, 20% 及び 40% のすべての検討結果で収束割合は 100%であった。1 群 20 例の検討結果については, 結果が明確であるため図示化はしていない。

表 3.6-6 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 1, 2, 3, 4, 5 及び 6, 欠測なし, 1 群 5 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW</sub>	100.0	---	---	100.0	---	---
2	MMRM <sub>SAT</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW</sub>	100.0	---	---	100.0	---	---
3	MMRM <sub>SAT</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW</sub>	100.0	---	---	100.0	---	---
4	MMRM <sub>SAT</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW</sub>	100.0	---	---	100.0	---	---
5	MMRM <sub>SAT</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW</sub>	100.0	---	---	100.0	---	---
6	MMRM <sub>SAT</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW</sub>	100.0	---	---	100.0	---	---

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

表 3.6-7 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 10 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	100.0	100.0	91.9	100.0	100.0	87.4
	MMRM <sub>KR</sub>	100.0	100.0	91.9	100.0	100.0	87.4
	MMRM <sub>BW</sub>	100.0	100.0	91.9	100.0	100.0	87.4
2	MMRM <sub>SAT</sub>	100.0	100.0	95.6	100.0	99.9	84.9
	MMRM <sub>KR</sub>	100.0	100.0	95.6	100.0	99.9	84.9
	MMRM <sub>BW</sub>	100.0	100.0	95.6	100.0	99.9	84.9
3	MMRM <sub>SAT</sub>	100.0	100.0	92.1	100.0	100.0	88.4
	MMRM <sub>KR</sub>	100.0	100.0	92.1	100.0	100.0	88.4
	MMRM <sub>BW</sub>	100.0	100.0	92.1	100.0	100.0	88.4
4	MMRM <sub>SAT</sub>	100.0	100.0	92.2	100.0	99.9	86.4



	MMRM <sub>KR</sub>	100.0	100.0	92.2	100.0	99.9	86.4
	MMRM <sub>BW</sub>	100.0	100.0	92.2	100.0	99.9	86.4
5	MMRM <sub>SAT</sub>	100.0	100.0	92.3	100.0	99.9	94.0
	MMRM <sub>KR</sub>	100.0	100.0	92.3	100.0	99.9	94.0
	MMRM <sub>BW</sub>	100.0	100.0	92.3	100.0	99.9	94.0
6	MMRM <sub>SAT</sub>	100.0	100.0	92.3	100.0	100.0	93.9
	MMRM <sub>KR</sub>	100.0	100.0	92.3	100.0	100.0	93.9
	MMRM <sub>BW</sub>	100.0	100.0	92.3	100.0	100.0	93.9

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

表 3.6-8 収束割合（％）のシミュレーション実験結果<sup>†</sup>（シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 20 例）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>BW</sub>	100.0	100.0	100.0	100.0	100.0	100.0
2	MMRM <sub>SAT</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>BW</sub>	100.0	100.0	100.0	100.0	100.0	100.0
3	MMRM <sub>SAT</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>BW</sub>	100.0	100.0	100.0	100.0	100.0	100.0
4	MMRM <sub>SAT</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>BW</sub>	100.0	100.0	100.0	100.0	100.0	100.0
5	MMRM <sub>SAT</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>BW</sub>	100.0	100.0	100.0	100.0	100.0	100.0
6	MMRM <sub>SAT</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>BW</sub>	100.0	100.0	100.0	100.0	100.0	100.0

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

### 3.6.3.3（固定効果の）標準誤差

1 群 5 例, 10 例, 20 例のパターンでそれぞれ（固定効果の）標準誤差を評価した。1 群 5 例の場合には、欠測が発生すると極端に収束状況が悪化してしまうため、欠測がない状況下のみ検討結果を表示した。反復回数 1000 回の結果を本文に掲載した。反復回数を 10000 回とした場合についても検討し同様の結果を得ている。結果は付録に参考として掲載した。

表 3.6-9 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 1, 2, 3, 4, 5 及び 6, 欠測なし, 1 群 5 例）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	4.10	---	---	4.10	---	---

2	MMRM <sub>KR</sub>	4.15	---	---	4.15	---	---
	MMRM <sub>BW</sub>	4.10	---	---	4.10	---	---
	MMRM <sub>SAT</sub>	4.10	---	---	4.10	---	---
3	MMRM <sub>KR</sub>	4.16	---	---	4.16	---	---
	MMRM <sub>BW</sub>	4.10	---	---	4.10	---	---
	MMRM <sub>SAT</sub>	4.10	---	---	4.10	---	---
4	MMRM <sub>KR</sub>	4.17	---	---	4.17	---	---
	MMRM <sub>BW</sub>	4.10	---	---	4.10	---	---
	MMRM <sub>SAT</sub>	4.10	---	---	4.10	---	---
5	MMRM <sub>KR</sub>	4.11	---	---	4.11	---	---
	MMRM <sub>BW</sub>	4.10	---	---	4.10	---	---
	MMRM <sub>SAT</sub>	4.10	---	---	4.10	---	---
6	MMRM <sub>KR</sub>	4.13	---	---	4.13	---	---
	MMRM <sub>BW</sub>	4.10	---	---	4.10	---	---
	MMRM <sub>SAT</sub>	4.10	---	---	4.10	---	---

†MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

表 3.6-10 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 10 例）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	2.96	3.23	3.64	2.96	3.32	3.89
	MMRM <sub>KR</sub>	2.96	3.35	4.04	2.96	3.47	4.37
	MMRM <sub>BW</sub>	2.96	3.23	3.64	2.96	3.32	3.89
2	MMRM <sub>SAT</sub>	2.96	3.22	3.59	2.96	3.35	3.90
	MMRM <sub>KR</sub>	2.96	3.34	3.93	2.96	3.51	4.43
	MMRM <sub>BW</sub>	2.96	3.22	3.59	2.96	3.35	3.90
3	MMRM <sub>SAT</sub>	2.96	3.23	3.61	2.96	3.33	3.86
	MMRM <sub>KR</sub>	2.96	3.35	3.98	2.96	3.49	4.42
	MMRM <sub>BW</sub>	2.96	3.23	3.61	2.96	3.33	3.86
4	MMRM <sub>SAT</sub>	2.96	3.23	3.61	2.96	3.33	3.89
	MMRM <sub>KR</sub>	2.96	3.35	4.05	2.96	3.49	4.43
	MMRM <sub>BW</sub>	2.96	3.23	3.61	2.96	3.33	3.89
5	MMRM <sub>SAT</sub>	2.96	3.23	3.61	2.96	3.25	3.77
	MMRM <sub>KR</sub>	2.96	3.35	4.07	2.96	3.37	4.20
	MMRM <sub>BW</sub>	2.96	3.23	3.61	2.96	3.25	3.77
6	MMRM <sub>SAT</sub>	2.96	3.23	3.61	2.96	3.26	3.78
	MMRM <sub>KR</sub>	2.96	3.35	4.03	2.96	3.37	4.19
	MMRM <sub>BW</sub>	2.96	3.23	3.61	2.96	3.26	3.78

†MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

表 3.6-11 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 20 例）

シナリオ	方法 <sup>†</sup>	累積欠測割合					

		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	2.10	2.29	2.55	2.10	2.32	2.64
	MMRM <sub>KR</sub>	2.10	2.33	2.69	2.10	2.37	2.81
	MMRM <sub>BW</sub>	2.10	2.29	2.55	2.10	2.32	2.64
2	MMRM <sub>SAT</sub>	2.10	2.28	2.53	2.10	2.33	2.68
	MMRM <sub>KR</sub>	2.10	2.32	2.66	2.10	2.38	2.86
	MMRM <sub>BW</sub>	2.10	2.28	2.53	2.10	2.33	2.68
3	MMRM <sub>SAT</sub>	2.10	2.29	2.56	2.10	2.31	2.64
	MMRM <sub>KR</sub>	2.10	2.33	2.70	2.10	2.36	2.83
	MMRM <sub>BW</sub>	2.10	2.29	2.56	2.10	2.31	2.64
4	MMRM <sub>SAT</sub>	2.10	2.29	2.56	2.10	2.32	2.67
	MMRM <sub>KR</sub>	2.10	2.33	2.70	2.10	2.37	2.85
	MMRM <sub>BW</sub>	2.10	2.29	2.56	2.10	2.32	2.67
5	MMRM <sub>SAT</sub>	2.10	2.29	2.56	2.10	2.29	2.61
	MMRM <sub>KR</sub>	2.10	2.33	2.70	2.10	2.33	2.76
	MMRM <sub>BW</sub>	2.10	2.29	2.56	2.10	2.29	2.61
6	MMRM <sub>SAT</sub>	2.10	2.29	2.56	2.10	2.29	2.62
	MMRM <sub>KR</sub>	2.10	2.33	2.70	2.10	2.33	2.76
	MMRM <sub>BW</sub>	2.10	2.29	2.56	2.10	2.29	2.62

†MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

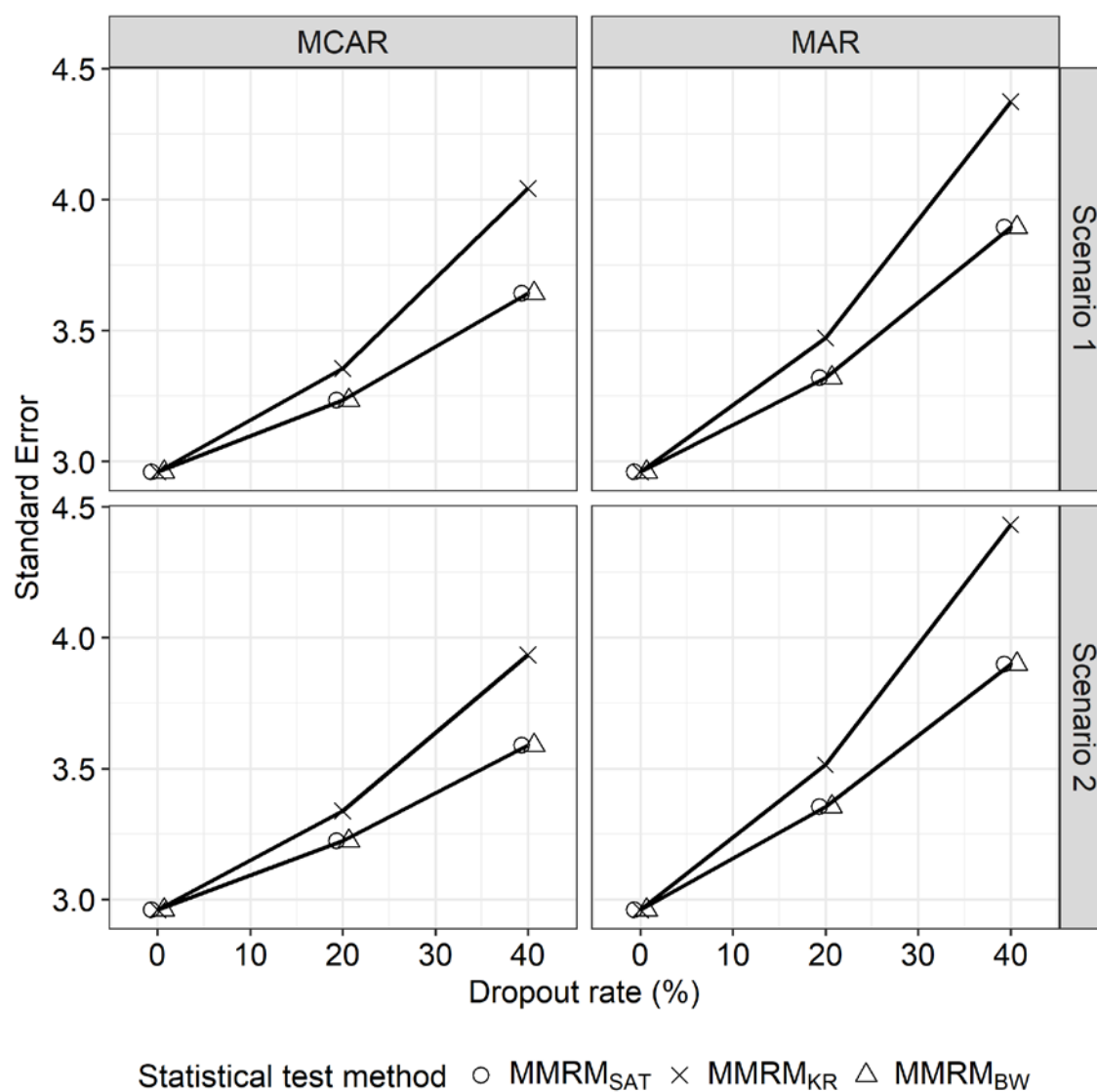


図 3.6-12 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 10 例）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

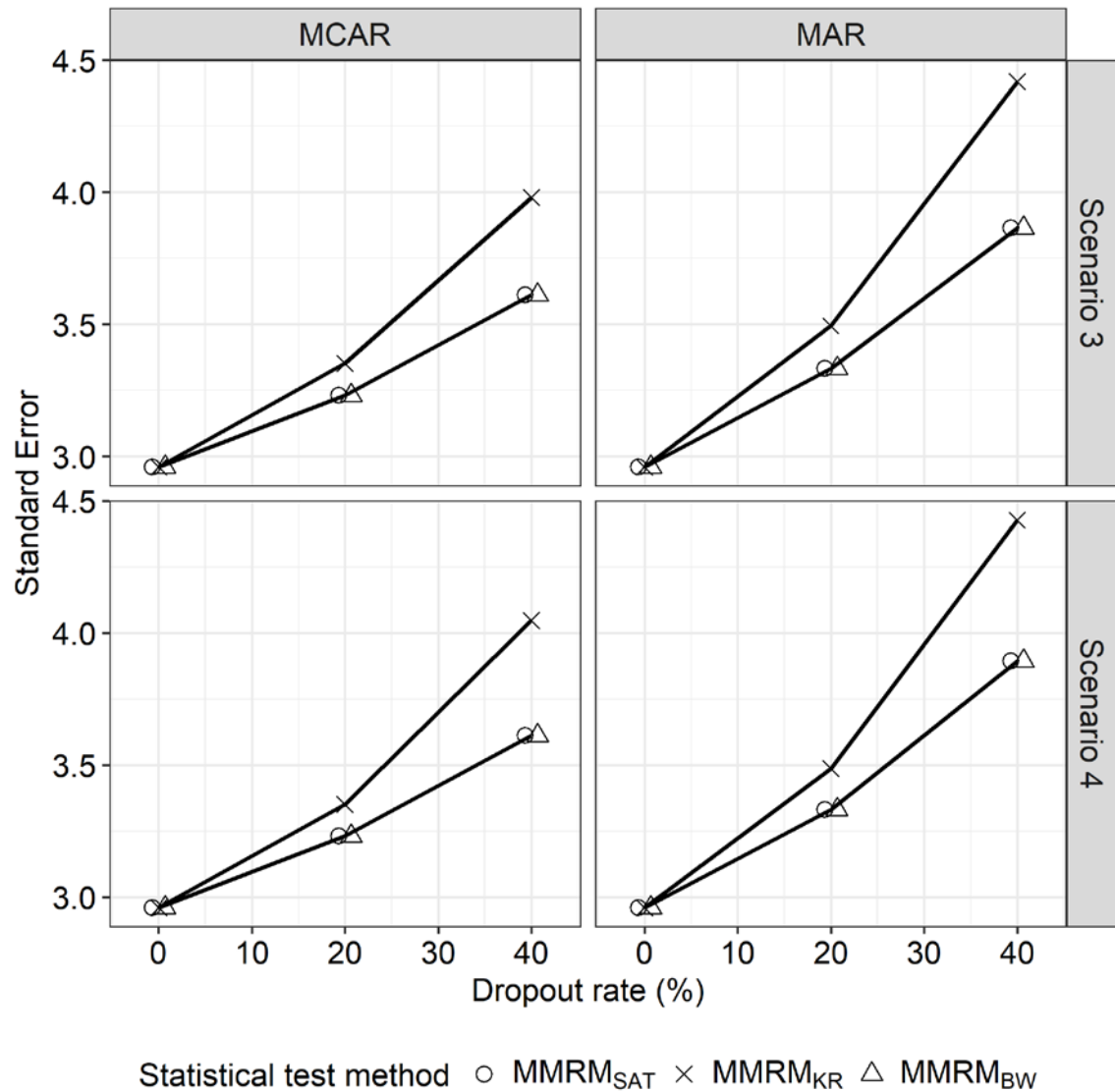


図 3.6-13 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 10 例）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

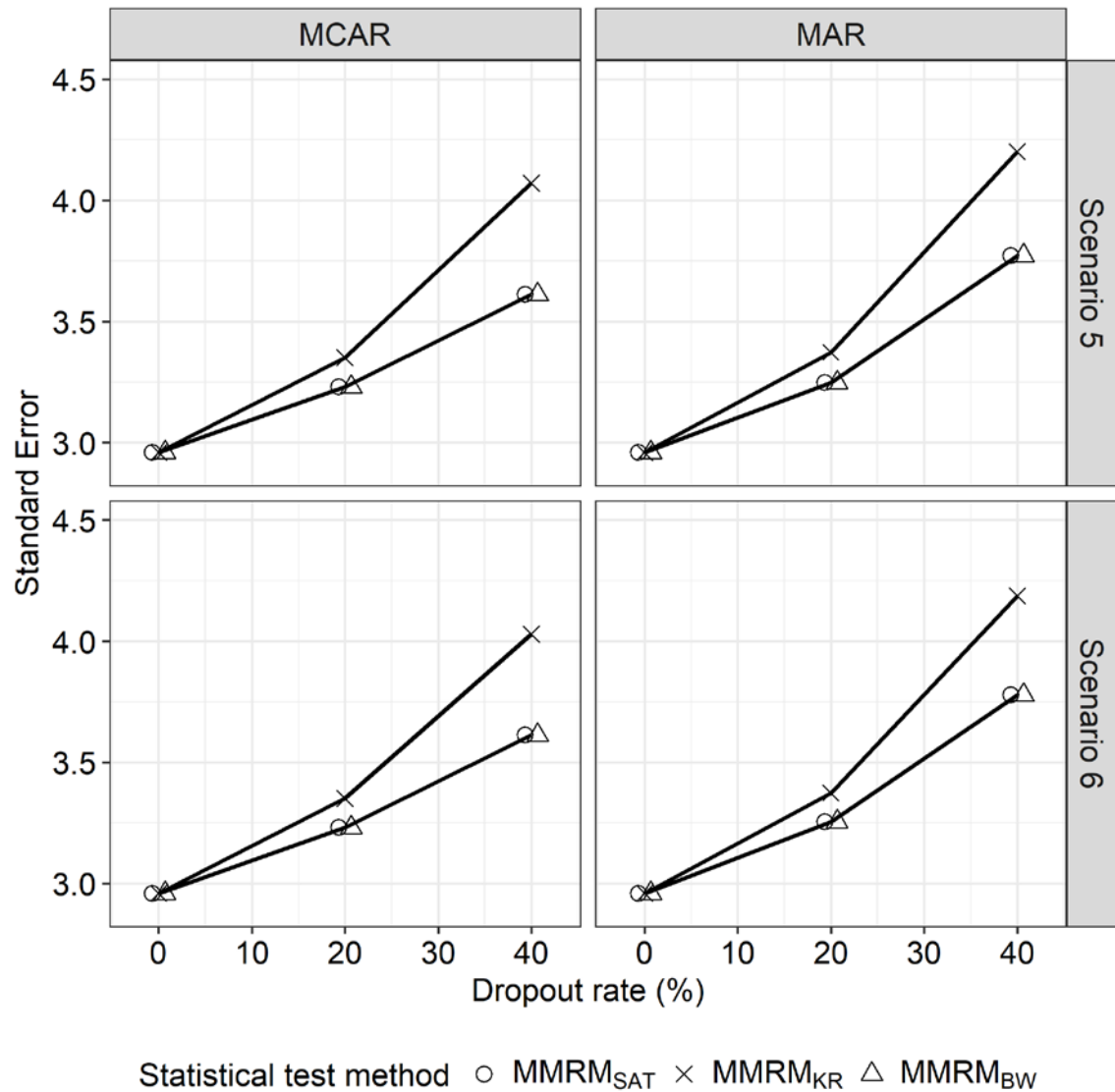


図 3.6-14 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 10 例）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

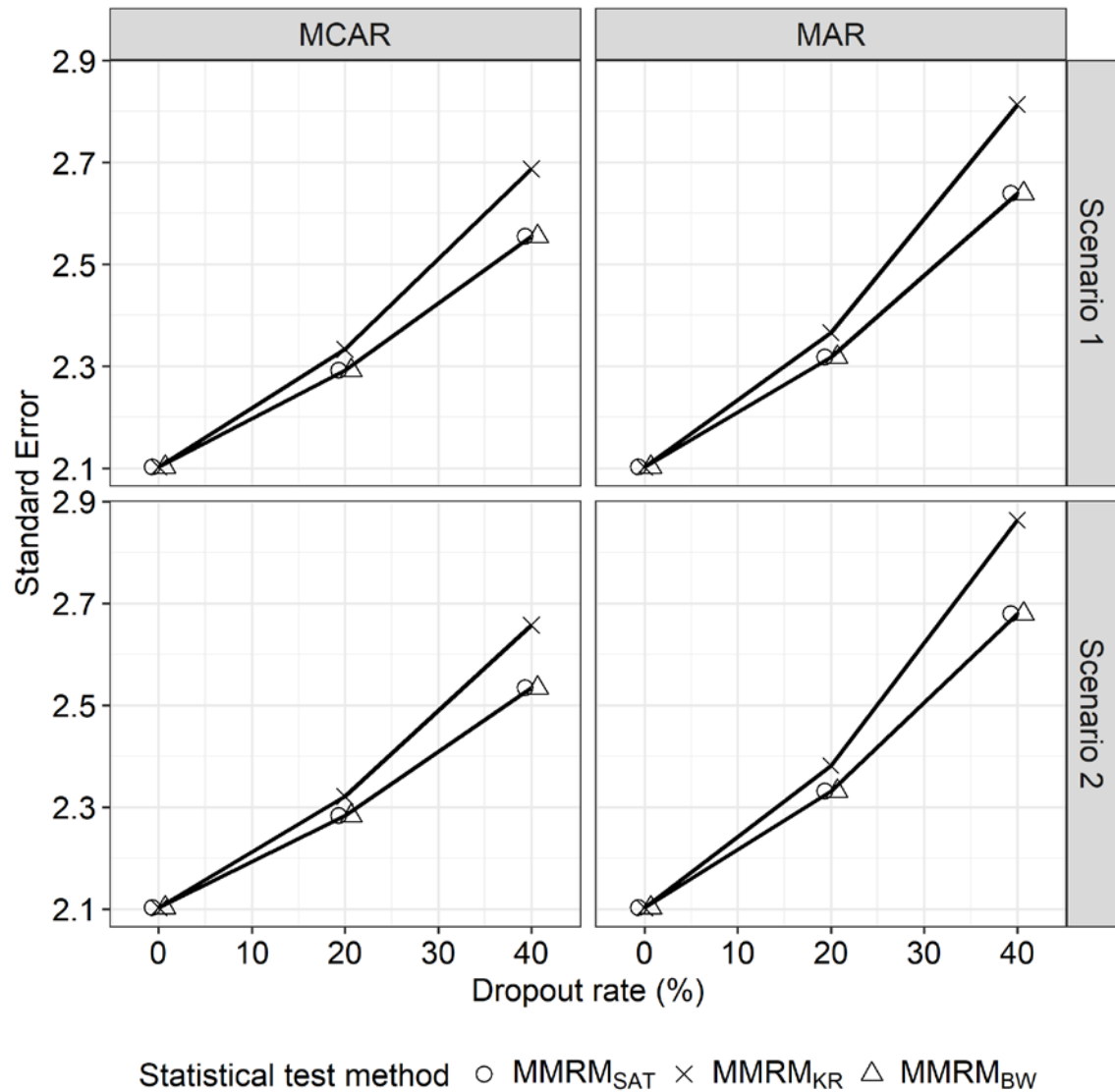


図 3.6-15 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 20 例）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

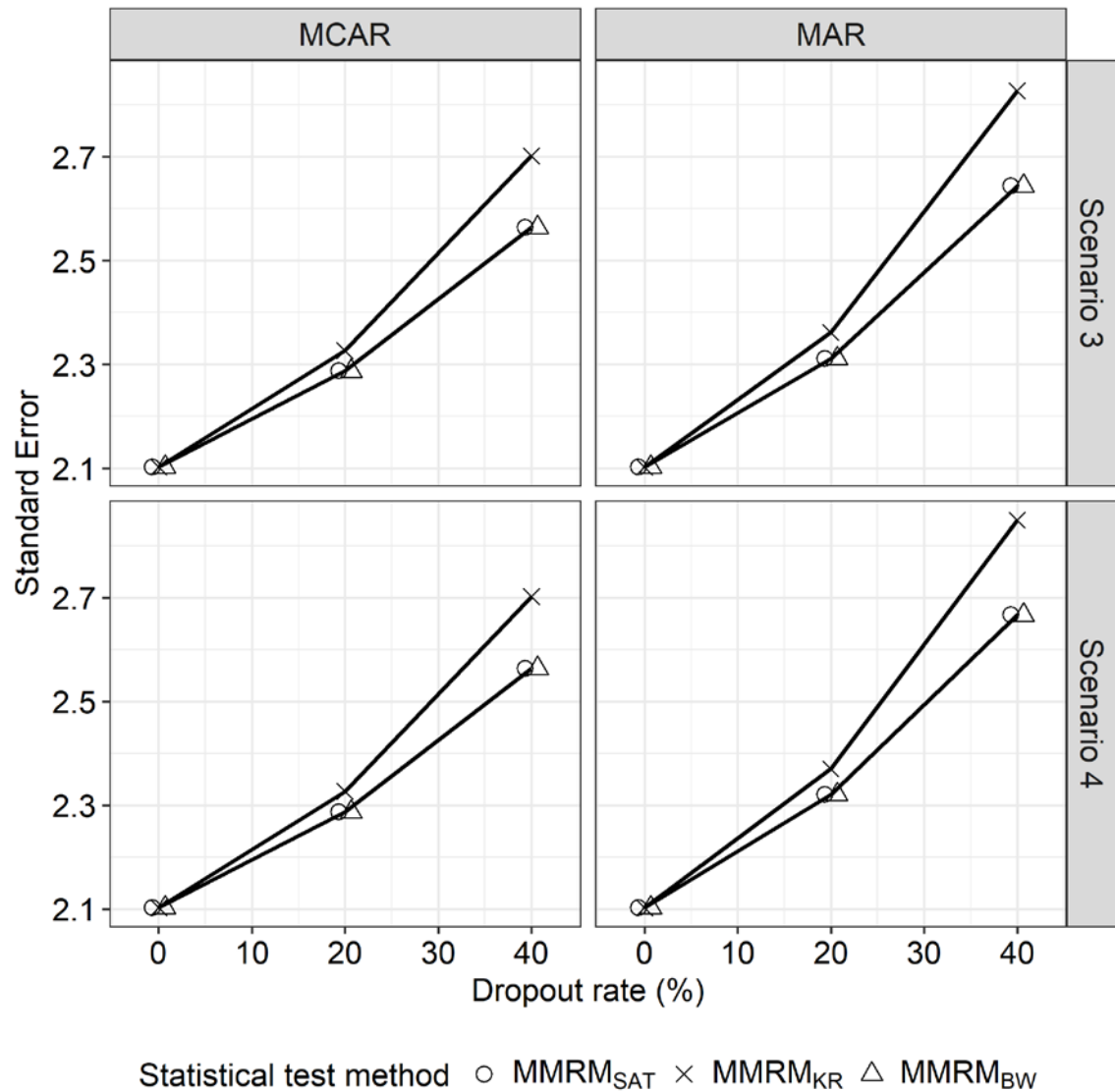


図 3.6-16 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 20 例）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定



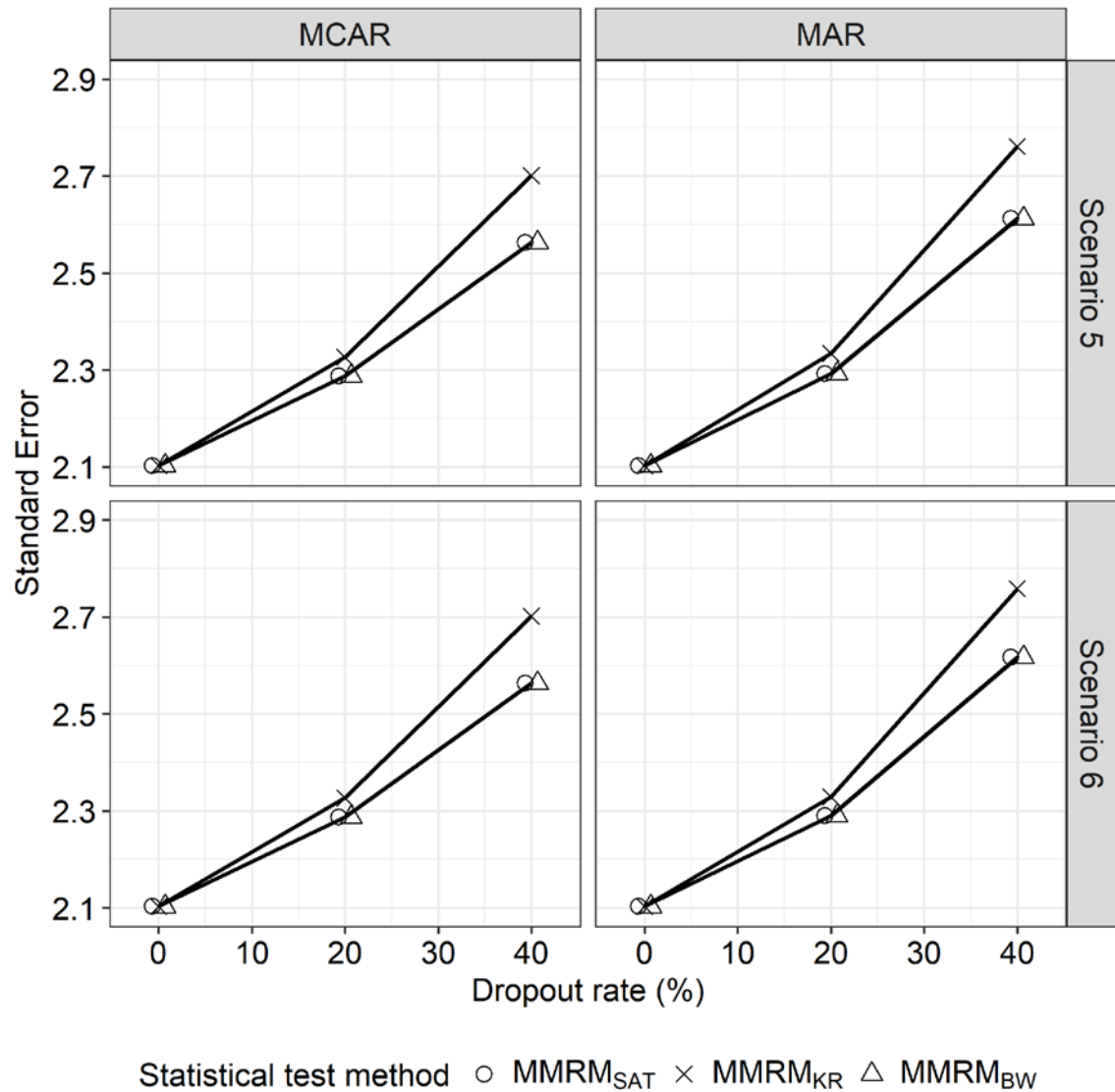


図 3.6-17 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 20 例）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

### 3.6.3.4 (*t*検定の) 自由度

1 群 5 例, 10 例, 20 例のパターンでそれぞれ (*t*検定の) 自由度を評価した。1 群 5 例の場合には, 欠測が発生すると極端に収束状況が悪化してしまうため, 欠測がない状況下のみ検討結果を表示した。反復回数 1000 回の結果を本文に掲載した。なお, 反復回数 1000 回の図は付録に掲載した。反復回数を 10000 回とした場合についても検討し同様の結果を得ている。結果は付録に参考として掲載した。

表 3.6-12 *t*検定の自由度のシミュレーション実験結果<sup>†</sup> (シナリオ 1, 2, 3, 4, 5 及び 6, 欠測なし, 1 群 5 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	8.00	---	---	8.00	---	---
	MMRM <sub>KR</sub>	7.99	---	---	7.99	---	---
	MMRM <sub>BW</sub>	8.00	---	---	8.00	---	---
2	MMRM <sub>SAT</sub>	8.00	---	---	8.00	---	---
	MMRM <sub>KR</sub>	7.99	---	---	7.99	---	---
	MMRM <sub>BW</sub>	8.00	---	---	8.00	---	---
3	MMRM <sub>SAT</sub>	8.00	---	---	8.00	---	---
	MMRM <sub>KR</sub>	7.99	---	---	7.99	---	---
	MMRM <sub>BW</sub>	8.00	---	---	8.00	---	---
4	MMRM <sub>SAT</sub>	8.00	---	---	8.00	---	---
	MMRM <sub>KR</sub>	7.98	---	---	7.98	---	---
	MMRM <sub>BW</sub>	8.00	---	---	8.00	---	---
5	MMRM <sub>SAT</sub>	8.00	---	---	8.00	---	---
	MMRM <sub>KR</sub>	7.98	---	---	7.98	---	---
	MMRM <sub>BW</sub>	8.00	---	---	8.00	---	---
6	MMRM <sub>SAT</sub>	8.00	---	---	8.00	---	---
	MMRM <sub>KR</sub>	7.98	---	---	7.98	---	---
	MMRM <sub>BW</sub>	8.00	---	---	8.00	---	---

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

表 3.6-13 *t*検定の自由度のシミュレーション実験結果<sup>†</sup> (シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 10 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	18.00	14.96	11.82	18.00	12.03	8.60
	MMRM <sub>KR</sub>	18.00	14.96	11.80	18.00	12.03	8.53
	MMRM <sub>BW</sub>	18.00	18.00	18.00	18.00	18.00	18.00
2	MMRM <sub>SAT</sub>	18.00	15.07	12.04	18.00	11.82	8.59
	MMRM <sub>KR</sub>	18.00	15.07	12.02	18.00	11.82	8.53
	MMRM <sub>BW</sub>	18.00	18.00	18.00	18.00	18.00	18.00
3	MMRM <sub>SAT</sub>	18.00	15.04	11.92	18.00	12.06	8.71
	MMRM <sub>KR</sub>	18.00	15.04	11.90	18.00	12.06	8.69
	MMRM <sub>BW</sub>	18.00	18.00	18.00	18.00	18.00	18.00
4	MMRM <sub>SAT</sub>	18.00	15.04	11.92	18.00	12.07	8.65
	MMRM <sub>KR</sub>	18.00	15.04	11.90	18.00	12.07	8.56
	MMRM <sub>BW</sub>	18.00	18.00	18.00	18.00	18.00	18.00
5	MMRM <sub>SAT</sub>	18.00	15.04	12.00	18.00	13.06	9.47

	MMRM <sub>KR</sub>	18.00	15.04	11.98	18.00	13.06	9.47
	MMRM <sub>BW</sub>	18.00	18.00	18.00	18.00	18.00	18.00
6	MMRM <sub>SAT</sub>	18.00	15.04	12.06	18.00	13.14	9.46
	MMRM <sub>KR</sub>	18.00	15.04	11.96	18.00	13.14	9.44
	MMRM <sub>BW</sub>	18.00	18.00	18.00	18.00	18.00	18.00

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

表 3.6-14  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup>（シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 20 例）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	38.00	32.66	26.19	38.00	25.76	18.27
	MMRM <sub>KR</sub>	38.00	32.66	26.19	38.00	25.76	18.27
	MMRM <sub>BW</sub>	38.00	38.00	38.00	38.00	38.00	38.00
2	MMRM <sub>SAT</sub>	38.00	32.93	26.82	38.00	25.41	17.74
	MMRM <sub>KR</sub>	38.00	32.93	26.82	38.00	25.41	17.74
	MMRM <sub>BW</sub>	38.00	38.00	38.00	38.00	38.00	38.00
3	MMRM <sub>SAT</sub>	38.00	32.85	26.34	38.00	25.82	18.32
	MMRM <sub>KR</sub>	38.00	32.85	26.34	38.00	25.82	18.32
	MMRM <sub>BW</sub>	38.00	38.00	38.00	38.00	38.00	38.00
4	MMRM <sub>SAT</sub>	38.00	32.85	26.33	38.00	25.78	18.13
	MMRM <sub>KR</sub>	38.00	32.85	26.33	38.00	25.78	18.13
	MMRM <sub>BW</sub>	38.00	38.00	38.00	38.00	38.00	38.00
5	MMRM <sub>SAT</sub>	38.00	32.85	26.35	38.00	27.71	20.55
	MMRM <sub>KR</sub>	38.00	32.85	26.35	38.00	27.71	20.55
	MMRM <sub>BW</sub>	38.00	38.00	38.00	38.00	38.00	38.00
6	MMRM <sub>SAT</sub>	38.00	32.85	26.33	38.00	27.93	20.68
	MMRM <sub>KR</sub>	38.00	32.85	26.33	38.00	27.93	20.68
	MMRM <sub>BW</sub>	38.00	38.00	38.00	38.00	38.00	38.00

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

### 3.6.4 収束に関する検討

本セクションでは収束割合について考察する。収束割合はシミュレーションにより検討したが、これを理論的に計算する方法を考える。考えやすいように、 $N = 20$ （1 群 10 例）で MCAR もしくは MAR で欠測が発生する場合を例にする。シミュレーションでは、最終時点での累積欠測割合が 0%、20% 及び 40% となるように設定していた。仮定した全 7 時点の最終時点（第 7 時点）における収束条件を満たす標本数とはどんな標本数となるであろうか。

第 7 時点は第 1 時点から 6 時点までにデータの条件付き分布となっている。すなわち、推定すべきパラメータと自由度は次のようになっている。

- ✓ 第 7 時点における 2 群の平均パラメータに対して自由度 2
  - ✓  $7 \times 7$  の分散共分散行列の 7 行目の 6 個の非対角成分をパラメータ推定するため自由度 6
- したがって、最終時点である第 7 時点では標本数が 8 以下ではパラメータ推定ができない状況となっている。逆に標本数が 9 以上あれば、パラメータ推定可能であり収束が得られると言

える。

$N = 20$  (1 群 10 例) で累積欠測割合 40% の状況を考える。サイズ 20, 確率 60% の 2 項分布を仮定し, サイズが 9 以上となる確率を計算すると約 94.3% となる。表 3.6-7 より, 累積欠測割合 40% のときの収束割合はほぼ 90% 以上であり理論上の数値と同等な結果となっていた。シナリオ 1, 2, 3 及び 4 で収束割合が 90% を下回っており, 理論上の数値と合致していないが, その原因は究明できなかった。

### 3.7 実データへの適用

提案法及びシミュレーションによる検討で対象とした比較方法を含めて, 実データへ適用しその有用性を検討した。小標本かつ欠測を伴う経時測定データとして, 産後うつ臨床試験データ及び認知行動療法臨床試験データを対象とした。

提案法のモチベーションは, 小標本かつ欠測を伴う経時測定データに対する有用な統計的推測方法の開発であるが, 本提案法の応用としてクロスオーバー試験データにも適用可能である。そこで, 気管支喘息治療薬のクロスオーバー試験データを対象に検討した結果についても記載する。

#### 3.7.1 産後うつ臨床試験

産後うつに対する治療には抗うつ剤やカウンセリングなどの治療が行われているが, エストロゲンの経皮投与による治療が有効である可能性がある。そこで, Gregoire et al. (1996) は 61 名の産後 3 か月以内の女性を対象に二重盲検プラセボ対照比較試験を実施した。試験の計画では 100 例の被験者登録を見込んでいたが, 実際には想定よりも登録が進捗せずに 61 例の女性がプラセボ群 (27 例) もしくはエストロゲン群 (34 例) にランダム割付された。本臨床試験の有効性に関するデータが書籍に掲載されており, 実データへの適用を検討するために用いることとした (Everitt 2001)。有効性データの被験者個別値一覧を付録に掲載した。

本試験は 1 か月ごとにうつ症状の指標である Edinburgh postnatal depression scale (EPDS) スコア等の調査を行い 8 か月目までフォローアップしている。被験者は, ランダム化前までに 2 回 EPDS スコアを測定し, その後 6 ヶ月間フォローアップされた。つまり, 1 か月目及び 2 か月目の EPDS スコアはランダム化前の測定値となる。EPDS スコアは, スコアが高いほどうつ症状の重症度が高いことを示す指標である。

小標本かつ欠測を伴う経時測定データを検討するために, 部分集団を定義する。先行研究では EPDS スコア 19 以上が臨床的に意味のある重症うつ症状との報告がある (McCabe-Beane JE et al. 2016)。報告されている臨床的に意味があるとされているカットオフ値ではないが, ベースライン EPDS スコアが 21 を超える集団をうつ症状が重症な部分集団と定義する。部分集団の対象例数は 30 例 (プラセボ群及びエストロゲン群それぞれ 15 例) であり, 8 ヶ月目には 9 例が脱落し, 21 例 (プラセボ群 9 例及びエストロゲン群 12 例) の EPDS スコアが得られている。対象とする例数は, シミュレーションで検討した標本数とほぼ同一であり, 提案法の有用性について検討可能と考えられた。

8 ヶ月目の最終評価時点における全体集団では, プラセボ群で 37.0% (10/27), エストロゲン群で 17.6% (6/34) の被験者に EPDS スコアの欠測が認められた。また部分集団では, プラセボ群で 40.0% (6/15), エストロゲン群で 20.0% (3/15) の被験者で EPDS スコアが欠測となっ

ていた。すべての欠測は単調な欠測であり、シミュレーションで用いた欠測パターンと同一である。各評価時点における投与群ごとの欠測データの割合を表 3.7-1 に示した。

表 3.7-1 EPDS スコアの欠測割合（測定時点ごと）

Time	Placebo	Estrogen	Total
<Overall>	N = 27	N = 34	N = 61
Baseline	0.0 (0/27)	0.0 (0/34)	0.0 (0/61)
3 months	0.0 (0/27)	0.0 (0/34)	0.0 (0/61)
4 months	18.5 (5/27)	8.8 (3/34)	13.1 (8/61)
5 months	37.0 (10/27)	14.7 (5/34)	24.6 (15/61)
6 months	37.0 (10/27)	17.6 (6/34)	26.2 (16/61)
7 months	37.0 (10/27)	17.6 (6/34)	26.2 (16/61)
8 months	37.0 (10/27)	17.6 (6/34)	26.2 (16/61)
<部分集団>	N = 15	N = 15	N = 30
Baseline	0.0 (0/15)	0.0 (0/15)	0.0 (0/30)
3 months	0.0 (0/15)	0.0 (0/15)	0.0 (0/30)
4 months	20.0 (3/15)	13.3 (2/15)	16.7 (5/30)
5 months	40.0 (6/15)	20.0 (3/15)	30.0 (9/30)
6 months	40.0 (6/15)	20.0 (3/15)	30.0 (9/30)
7 months	40.0 (6/15)	20.0 (3/15)	30.0 (9/30)
8 months	40.0 (6/15)	20.0 (3/15)	30.0 (9/30)

ベースライン値の定義：1 ヶ月目及び2 ヶ月目の平均

部分集団の定義：EPDS スコアのベースライン値>21 である被験者集団

EPDS スコアのベースラインは1 ヶ月目及び2 ヶ月目の EPDS スコアの平均値と定義された。原著論文では、3 ヶ月目以降 8 ヶ月目までの EPDS スコアを平均して結果変数として扱っている。平均化した EPDS スコアを結果変数とし、ベースライン EPDS スコア及び投与群を説明変数とした分散分析が主要な解析として位置付けられていた。付録に掲載した被験者個別値一覧に基づき主要な解析を再現したところ、ベースライン EPDS スコアの平均値（標準偏差）は、それぞれプラセボ群 21.26（3.11）及びエストロゲン群 21.59（3.06）であった（表 3.7-2）。

表 3.7-2 EPDS スコアのベースライン値の要約統計量

Time	Placebo	Estrogen	Total
<Overall>			
N	27	34	61
Mean (SD)	21.26 (3.11)	21.59 (3.06)	21.44 (3.06)
Median	21.5	21.0	21.0
IQ Range	[18.5, 23.0]	[19.5, 24.0]	[19.5, 23.5]
Min, Max	[15.5, 27.5]	[17.0, 27.5]	[15.5, 27.5]
<部分集団>			
N	15	15	30
Mean (SD)	23.37 (2.09)	24.40 (1.91)	23.88 (2.04)
Median	22.5	24.5	23.5
IQ Range	[21.5, 26.0]	[23.0, 26.0]	[22.0, 26.0]
Min, Max	[21.5, 27.5]	[21.5, 27.5]	[21.5, 27.5]

ベースライン値の定義：1 ヶ月目及び2 ヶ月目の平均

部分集団の定義：EPDS スコアのベースライン値>21 である被験者集団

結果変数に関する群間差とその 95%信頼区間は、4.38 [2.87, 6.88]であった。これらの解析結果は原著論文に記載されている数値とほぼ一致した。

EPDS スコアの経時推移を投与群別に確認するため、各測定時点における投与群ごとの EPDS

要約統計量を求めた。解析結果を表 3.7-3 に示す。ランダム化前の 1 ヶ月目及び 2 ヶ月目はプラセボ群及びエストロゲン群ともにほぼ同様なスコアを示した。ランダム化後の 3 ヶ月目以降では、プラセボ群の EPDS スコアと比較してエストロゲン群の EPDS スコアは低値を示した。エストロゲン群ではプラセボ群よりもうつ症状が軽減されていることが EPDS スコアから示唆された。

表 3.7-3 EPDS スコアの要約統計量（測定時点ごと、投与群ごと）

Item	Overall		部分集団	
	Placebo	Estrogen	Placebo	Estrogen
<1 month>				
N	26	34	15	15
Mean (SD)	21.92 (3.19)	21.94 (3.25)	23.67 (2.38)	24.53 (2.23)
Median	22.0	22.0	24.0	24.0
IQ Range	[20.0, 24.0]	[20.0, 24.0]	[22.0, 25.0]	[23.0, 27.0]
Min, Max	[15.0, 28.0]	[15.0, 28.0]	[18.0, 28.0]	[21.0, 28.0]
<2 month>				
N	27	34	15	15
Mean (SD)	20.78 (3.95)	21.24 (3.55)	23.07 (3.56)	24.27 (2.31)
Median	21.0	21.0	22.0	24.0
IQ Range	[17.0, 24.0]	[18.0, 24.0]	[21.0, 26.0]	[23.0, 26.0]
Min, Max	[15.0, 28.0]	[15.0, 28.0]	[16.0, 28.0]	[20.0, 28.0]
<3 month>				
N	27	34	15	15
Mean (SD)	16.48 (5.28)	13.35 (5.55)	17.33 (5.78)	13.80 (6.09)
Median	17.0	13.5	18.0	14.0
IQ Range	[13.0, 19.0]	[8.0, 16.0]	[13.0, 20.0]	[11.0, 19.0]
Min, Max	[7.0, 26.0]	[1.0, 27.0]	[7.0, 26.0]	[1.0, 24.0]
<4 month>				
N	22	31	12	13
Mean (SD)	15.86 (6.14)	11.71 (6.56)	18.08 (6.46)	13.85 (7.09)
Median	14.5	12.0	20.0	14.0
IQ Range	[12.0, 21.0]	[7.0, 16.0]	[13.0, 22.5]	[9.0, 18.0]
Min, Max	[4.0, 27.0]	[1.0, 27.0]	[4.0, 27.0]	[2.0, 27.0]
<5 month>				
N	17	29	9	12
Mean (SD)	14.12 (5.00)	9.10 (5.48)	15.22 (5.47)	11.25 (6.52)
Median	15.0	9.0	17.0	10.5
IQ Range	[9.0, 17.0]	[5.0, 12.0]	[15.0, 18.0]	[7.0, 15.5]
Min, Max	[4.0, 22.0]	[1.0, 24.0]	[4.0, 22.0]	[1.0, 24.0]
<6 month>				
N	17	28	9	12
Mean (SD)	12.18 (5.90)	8.79 (4.67)	12.56 (4.98)	9.92 (5.79)
Median	13.0	9.0	13.0	10.0
IQ Range	[8.0, 17.0]	[6.0, 12.0]	[12.0, 14.0]	[7.5, 12.5]
Min, Max	[2.0, 23.0]	[0.0, 22.0]	[4.0, 21.0]	[0.0, 22.0]
<7 month>				
N	17	28	9	12
Mean (SD)	11.35 (4.50)	7.29 (5.75)	12.33 (4.61)	9.08 (7.08)
Median	12.0	7.0	13.0	8.5
IQ Range	[8.0, 15.0]	[2.5, 11.0]	[11.0, 16.0]	[3.5, 12.5]
Min, Max	[3.0, 18.0]	[0.0, 24.0]	[3.0, 18.0]	[0.0, 24.0]
<8 month>				
N	17	28	9	12
Mean (SD)	10.82 (4.71)	6.46 (4.74)	11.67 (5.39)	8.50 (6.17)
Median	11.0	6.0	10.0	7.0
IQ Range	[7.0, 13.0]	[3.0, 7.5]	[9.0, 15.0]	[3.5, 11.5]

Min, Max	[3.0, 20.0]	[1.0, 23.0]	[3.0, 20.0]	[2.0, 23.0]
部分集団の定義：EPDS スコアのベースライン値>21 である被験者集団				

さらに、経時推移を視覚的に確認するために、全体集団と部分集団別に投与群ごとの EPDS スコアの推移を示した（図 3.7-1 及び図 3.7-2）。ランダム化後の 3 ヶ月目以降、エストロゲン群がプラセボ群よりも平均値で常に低値となっており、うつ症状の改善が示唆された。また、EPDS スコア平均値の推定精度を評価するために、標準誤差についても併せて示した。部分集団では全体集団よりも対象例数が減少するため、推定精度は低下していることが標準誤差のバーの長さから確認できた。

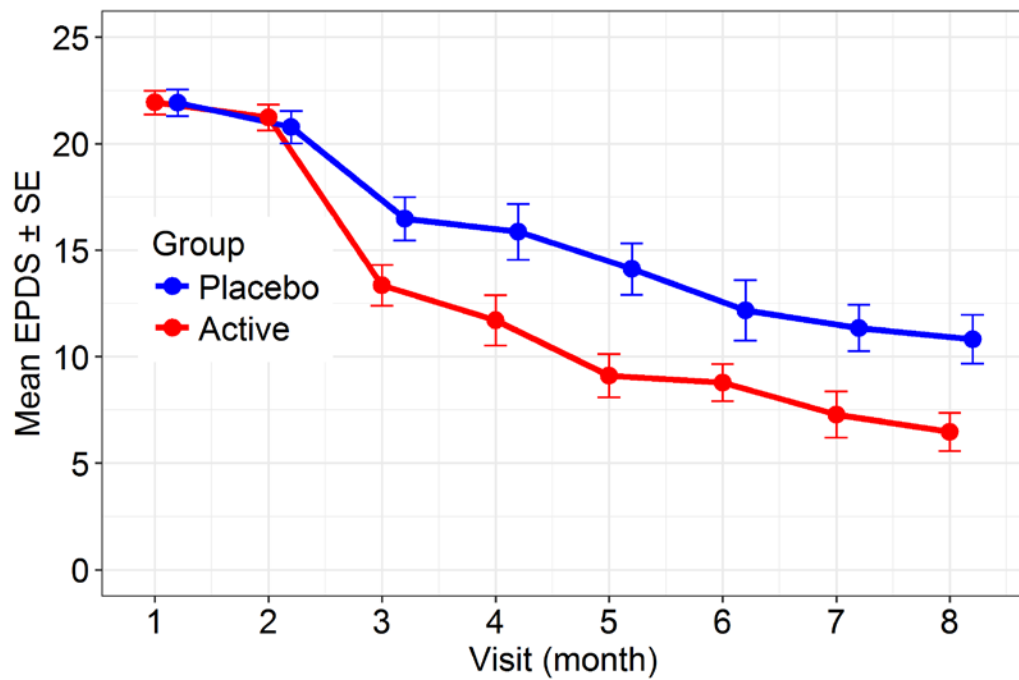


図 3.7-1 EPDS 平均値の推移（測定時点ごと，投与群ごと）[全体]

EPDS スコアの要約統計量をプロット

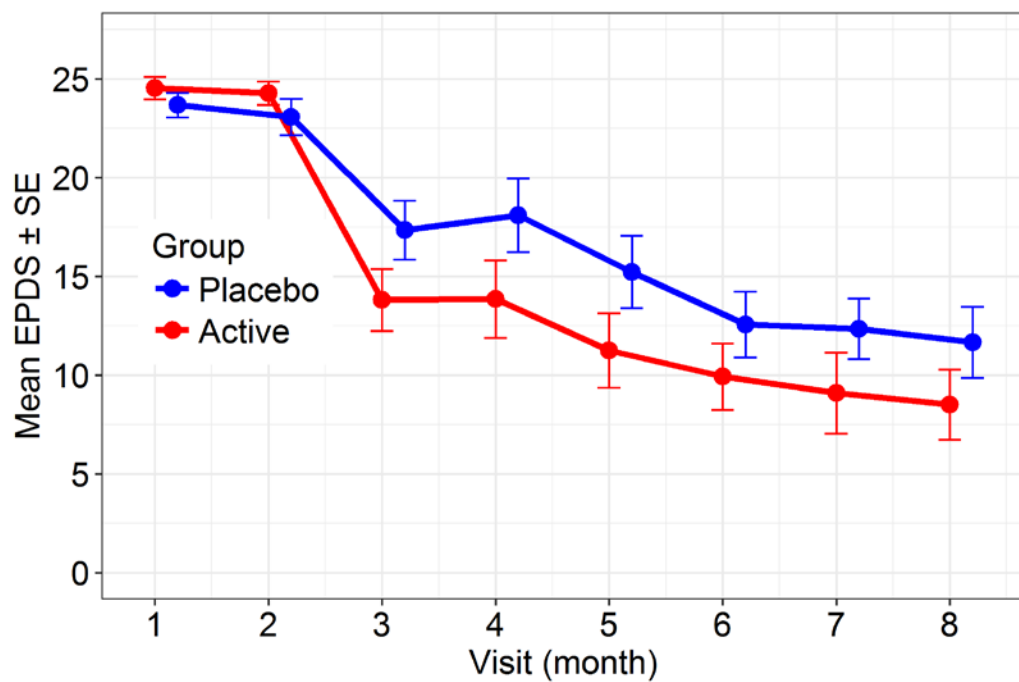


図 3.7-2 EPDS 平均値の推移（測定時点ごと，投与群ごと）[部分集団]

EPDS スコアの要約統計量をプロット

提案法及び比較方法の解析では，被験者を変量効果とした混合効果モデルを用いた。

$$y_{ij} = \beta_0 + \beta_1 G_i + \beta_2 t_{ij} + \beta_3 G_i t_{ij} + b_{0i} + \varepsilon_{ij}$$

被験者は， $i = 1, 2, \dots, 61$ であり，各被験者ごとに最終測定時点が異なっているため， $j =$



1, 2, ...,  $n_i$ となる。また，被験者の変量効果を $b_{0i} \sim N(0, \sigma_b^2)$ とし，誤差を $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ とした。最終測定時点 $j = 8$ における $\beta_{3j}$ が主要評価項目である EPDS スコアの治療効果に相当しており，用いた方法ごとに 95%信頼区間と $p$ 値を算出した。さらに，参考としてランダム化以降の測定時点についても，同様に解析結果を提示した（表 3.7-4）。

全体集団及び部分集団のいずれにおいても，LR<sub>Bart</sub>及びLR<sub>Boot</sub>の $p$ 値は従来法であるLRよりも保守的な値となった。特に最終測定時点である8か月目では，全体集団の解析結果ではLR<sub>Bart</sub>やLR<sub>Boot</sub>の $p$ 値（それぞれ，0.0031及び0.0050）はLRの $p$ 値（0.0019）よりも大きく提案法は保守的な結果を与える傾向がある。こうした結果はシミュレーションによる検討で得られた結果と一貫していた。

表 3.7-4 産後うつ臨床試験における群間差及びその95%信頼区間（測定時点ごと）

Item	Overall		部分集団	
	Estimate [95% CI]	$p$ -value	Estimate [95% CI]	$p$ -value
<3 month>				
LR <sub>Bart</sub>	3.13 [0.56, 6.35]	0.0202	3.53 [0.25, 8.88]	0.0392
LR <sub>Boot</sub>	3.13 [0.60, 6.31]	0.0193	3.53 [0.31, 8.83]	0.0367
LR	3.13 [0.64, 6.28]	0.0173	3.53 [0.56, 8.57]	0.0269
MMRM <sub>KR</sub>	3.13 [0.33, 5.93]	0.0293	3.53 [-0.90, 7.97]	0.1138
$t$ -test	3.13 [0.33, 5.93]	0.0293	3.53 [-0.90, 7.97]	0.1141
<4 month>				
LR <sub>Bart</sub>	3.91 [1.02, 7.44]	0.0109	3.80 [0.03, 9.55]	0.0488
LR <sub>Boot</sub>	3.91 [0.99, 7.47]	0.0103	3.80 [0.08, 9.50]	0.0483
LR	3.91 [1.14, 7.32]	0.0082	3.80 [0.40, 9.20]	0.0340
MMRM <sub>KR</sub>	3.91 [0.43, 7.38]	0.0283	3.80 [-1.48, 9.08]	0.1509
$t$ -test	4.15 [0.58, 7.73]	0.0237	4.24 [-1.39, 9.87]	0.1331
<5 month>				
LR <sub>Bart</sub>	5.09 [2.55, 8.29]	0.0004	4.60 [1.20, 10.10]	0.0152
LR <sub>Boot</sub>	5.09 [2.58, 8.26]	0.0003	4.60 [1.24, 10.05]	0.0173
LR	5.09 [2.67, 8.17]	0.0002	4.60 [1.66, 9.63]	0.0072
MMRM <sub>KR</sub>	5.09 [2.04, 8.15]	0.0015	4.60 [-0.18, 9.38]	0.0584
$t$ -test	5.01 [1.75, 8.28]	0.0034	3.97 [-1.66, 9.61]	0.1564
<6 month>				
LR <sub>Bart</sub>	3.28 [0.21, 6.99]	0.0382	3.27 [-0.10, 8.72]	0.0552
LR <sub>Boot</sub>	3.28 [0.30, 6.90]	0.0337	3.27 [-0.06, 8.68]	0.0543
LR	3.28 [0.38, 6.82]	0.0294	3.27 [0.28, 8.34]	0.0373
MMRM <sub>KR</sub>	3.28 [0.13, 6.42]	0.0414	3.27 [-1.17, 7.71]	0.1420
$t$ -test	3.39 [0.19, 6.59]	0.0383	2.64 [-2.40, 7.68]	0.2871
<7 month>				
LR <sub>Bart</sub>	3.99 [1.09, 7.54]	0.0100	3.97 [-0.03, 10.12]	0.0514
LR <sub>Boot</sub>	3.99 [1.06, 7.57]	0.0130	3.97 [0.03, 10.07]	0.0497
LR	3.99 [1.24, 7.39]	0.0070	3.97 [0.47, 9.61]	0.0321
MMRM <sub>KR</sub>	3.99 [0.79, 7.20]	0.0158	3.97 [-1.25, 9.18]	0.1296
$t$ -test	4.07 [0.77, 7.37]	0.0169	3.25 [-2.44, 8.94]	0.2462
<8 month>				
LR <sub>Bart</sub>	4.34 [1.67, 7.66]	0.0031	3.93 [-0.19, 10.24]	0.0586
LR <sub>Boot</sub>	4.34 [1.70, 7.64]	0.0050	3.93 [-0.16, 10.21]	0.0583
LR	4.34 [1.81, 7.52]	0.0019	3.93 [0.29, 9.74]	0.0383
MMRM <sub>KR</sub>	4.34 [1.45, 7.23]	0.0040	3.93 [-1.23, 9.09]	0.1288
$t$ -test	4.36 [1.43, 7.29]	0.0045	3.17 [-2.24, 8.57]	0.2349

部分集団の定義：EPDS スコアのベースライン値>21である被験者集団

図 3.7-3 に最終測定時点である 8 か月目において、パラメトリックブートストラップ法によりリサンプリングした尤度比検定統計量の経験分布に関するヒストグラムを示した。リサンプリングは 3000 回であり、上側が全体集団、下側が部分集団のヒストグラムとなっている。それぞれのヒストグラムの垂直な青点線は経験分布の平均値を示しており、全体集団及び部分集団でそれぞれ 1.09 及び 1.20 であった。また、ヒストグラムの垂直な赤点線は経験分布における 95% 点であり、全体集団及び部分集団でそれぞれ 4.14 及び 4.55 であった。

本事例では、帰無仮説下で尤度比検定統計量は漸近的に自由度 1 のカイ 2 乗分布に従うことから、経験分布の理論的な平均は自由度と等しく 1、95% 点は 3.84 となるはずである。実際にリサンプリングされた経験分布の平均値及び 95% 点と比較してみると、平均値や 95% 点は理論値よりも大きくなっている。これらの結果から、欠測を伴う小標本の経時測定データにおける尤度比検定統計量に従う分布は理論上のカイ 2 乗分布よりも平均が大きくなる方向にシフトしていることが分かる。パラメトリックブートストラップ法による Bartlett 補正は、このようなシフトを補正するために有効な方法と考えられる。

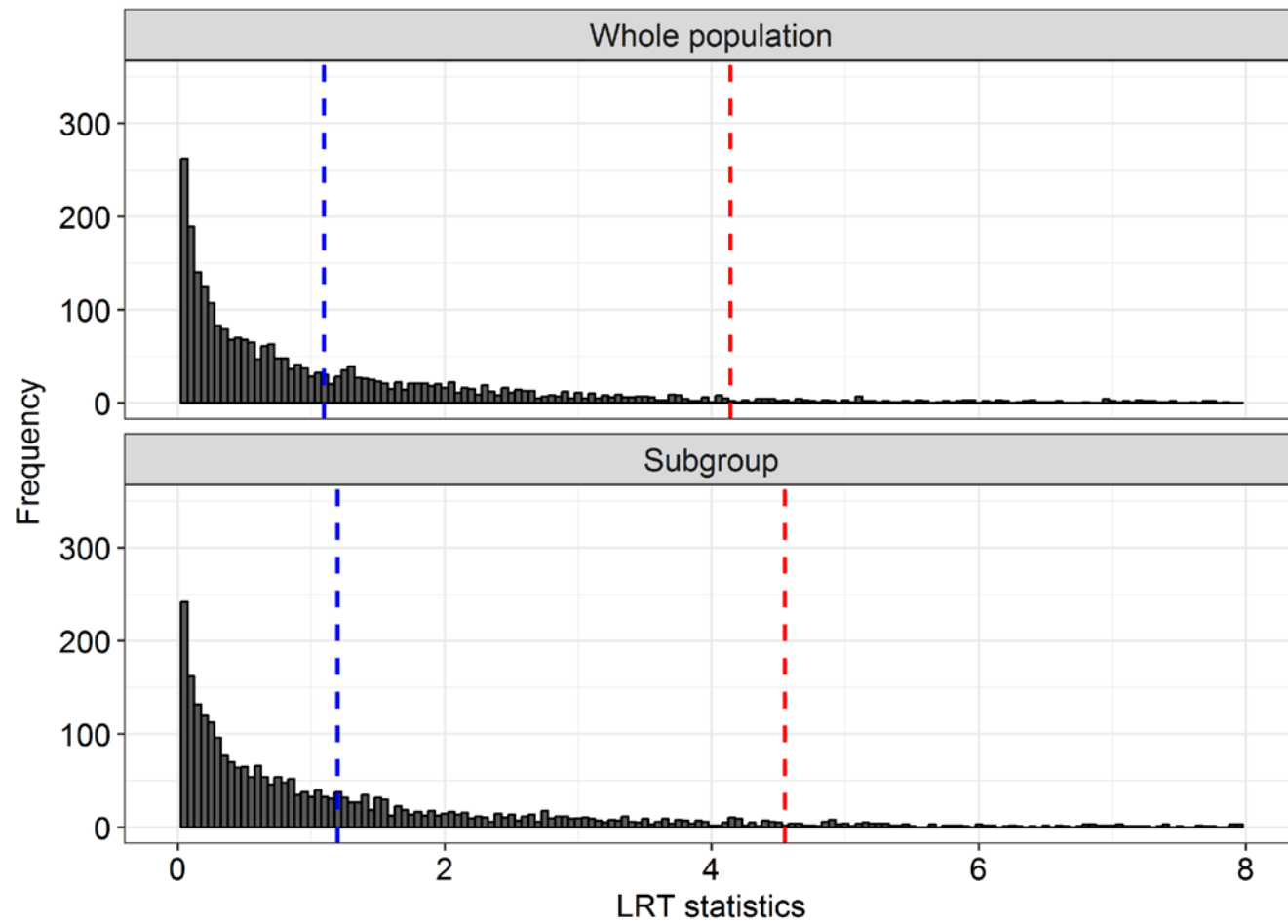


図 3.7-3 産後うつ臨床試験におけるパラメトリックブートストラップ法による尤度比検定統計量の経験分布<sup>†</sup>（8 ヶ月時点）

<sup>†</sup> 青点線，経験分布の平均（全体集団では 1.09，部分集団では 1.20）；赤点線，経験分布の 95% 点（全体集団では 4.14，部分集団では 4.55）

### 3.7.2 認知行動療法の臨床試験

不安やうつ症状を持つ患者に対して、認知行動療法が効果的であることが知られている (DeRubeis et al. 1999)。開発初期の認知行動療法は、特別なトレーニングを受けた医療従事者が直接患者と対話する方法が主流であり、他の治療方法と比較して人的リソースの負担が大きい点が問題であった。そこで、抑うつ症状のある患者に対してマルチメディア技術によるインタラクティブなプログラムを利用した認知行動療法が開発された。この療法は **BtB** (Beating the Blues : 憂鬱を追い払う) と呼ばれている。

Proudfoot et al. (2003)は、**BtB** の有用性を評価するためにイギリスの一次医療施設で治療を受けている患者を対象に臨床試験を実施した。臨床試験では 167 例の患者が **BtB** もしくは従来の投薬による治療 (Treatment as usual: TAU) にランダム化された。167 例のうち 100 例の実データが書籍に掲載されており、本研究では 100 例の実データを対象に提案法の適用を検討することとした (Everitt and Horthorn 2011)。本検討に用いた 100 例の被験者個別値一覧を付録に掲載した。

臨床試験では、うつ症状の測定に Beck Depression Inventory (BDI) が用いられた (Beck et al. 1996)。BDI のスコアは、うつ症状の重症度を反映している。測定は、治療開始前及び開始後の合計 5 回 (Pre-treatment, 2 か月後, 3 か月後, 5 か月後, 8 か月後) 実施された。さらに、ランダム化の対象となった患者に対して、抗うつ剤の治療歴 (治療有無) 及び過去の抑うつ症状の期間 ( $\leq 6$  か月,  $>6$  か月) についても調査されている。

Proudfoot らは、167 例のデータを対象に BDI を結果変数とし、ベースライン値、測定時期、治療群、抗うつ剤の治療歴有無及び過去の抑うつ症状の期間を固定効果に、被験者に関するランダム切片及びランダム傾きを変量効果に仮定した線形混合効果モデルによる解析結果を報告している。最終測定時点であるランダム化後 8 か月目の BDI に関する群間差及びその 95% 信頼区間は、-5.21 [-8.60, -1.82] と推定された。**BtB** により TAU よりも約 5 ポイントの BDI スコア減少が確認され、抑うつ症状の改善が示唆された。

論文では、全体集団に対する **BtB** の治療効果に加えて、抗うつ剤の治療歴有無や過去の抑うつ症状の期間に対する **BtB** の治療効果についても臨床的に興味があるものとされている。そこで、抑うつ症状の期間が 6 か月を超える患者集団を興味ある部分集団と定義した。

書籍に掲載されている 100 例の実データについて、100 例全体と抑うつ症状の期間が 6 か月を超える部分集団について、各評価時点における治療群ごとの BDI スコアの欠測割合を表 3.7-5 にまとめた。

全体集団では、8 ヶ月目の最終評価時点で TAU 群では 47.9% (23/48)、**BtB** 群では 48.1% (25/52) の被験者で BDI スコアが欠測であった。また、部分集団では TAU 群で 36.0% (9/25)、**BtB** 群で 42.3% (11/26) の被験者で BDI スコアが欠測であった。これらの欠測はすべて単調な欠測であり、シミュレーションで用いた欠測パターンと同一である。

表 3.7-5 BDI スコアの欠測割合 (測定時点ごと)

Time	TAU	BtB	Total
<Overall>	N = 48	N = 52	N = 100
Baseline	0.0 (0/48)	0.0 (0/52)	0.0 (0/100)
2m	6.3 (3/48)	0.0 (0/52)	3.0 (3/100)
3m	25.0 (12/48)	28.8 (15/52)	27.0 (27/100)

5m	39.6 (19/48)	44.2 (23/52)	42.0 (42/100)
8m	47.9 (23/48)	48.1 (25/52)	48.0 (48/100)
<部分集団>	<i>N</i> = 25	<i>N</i> = 26	<i>N</i> = 51
Baseline	0.0 (0/25)	0.0 (0/26)	0.0 (0/51)
2m	0.0 (0/25)	0.0 (0/26)	0.0 (0/51)
3m	8.0 (2/25)	26.9 (7/26)	17.6 (9/51)
5m	24.0 (6/25)	42.3 (11/26)	33.3 (17/51)
8m	36.0 (9/25)	42.3 (11/26)	39.2 (20/51)

ベースライン値の定義：pre-treatment の値

部分集団の定義：罹病期間>6 ヶ月である被験者集団

BDI スコアのベースラインは治療前値（Pre-treatment）と定義されている。BDI スコアのベースライン値に関する要約統計量を、表 3.7-6 にまとめた。ベースラインスコアの平均値（標準偏差）は、全体集団では TAU 群 24.19（9.82）及び BtB 群 22.54（11.74）であった。一方、部分集団では、TAU 群 25.40（10.37）及び BtB 群 24.77（10.40）であった。治療群間でベースライン値に大きなスコアの違いは認められない。

表 3.7-6 BDI スコアのベースライン値の要約統計量

Time	TAU	BtB	Total
<Overall>			
N	48	52	100
Mean (SD)	24.19 (9.82)	22.54 (11.74)	23.33 (10.84)
Median	23.0	20.5	22.0
IQ Range	[16.5, 30.5]	[13.5, 31.0]	[15.0, 30.5]
Min, Max	[7.0, 47.0]	[2.0, 49.0]	[2.0, 49.0]
<部分集団>			
N	25	26	51
Mean (SD)	25.40 (10.37)	24.77 (10.40)	25.08 (10.29)
Median	27.0	21.5	23.0
IQ Range	[20.0, 33.0]	[16.0, 32.0]	[17.0, 33.0]
Min, Max	[7.0, 47.0]	[11.0, 46.0]	[7.0, 47.0]

ベースライン値の定義：pre-treatment の値

部分集団の定義：罹病期間>6 ヶ月である被験者集団

BDI スコアの経時推移を治療群別に確認するため、各測定時点における治療群ごとの BDI スコア要約統計量を算出した。解析結果を表 3.7-7 に示す。臨床試験の最終測定時点である 8 か月目において、全体集団の BDI スコア平均値（標準偏差）は、TAU 群 13.60（11.47）及び BtB 群 8.85（6.09）であった。また、部分集団では TAU 群 18.00（11.49）及び BtB 群 10.73（5.99）であった。全体集団及び部分集団のいずれの集団においても、TAU 群と比較して BtB 群で BDI スコアの改善が認められた。

表 3.7-7 BDI スコアの要約統計量（測定時点ごと、投与群ごと）

Item	Overall		部分集団	
	TAU	BtB	TAU	BtB
<Baseline>				
N	48	52	25	26
Mean (SD)	24.19 (9.82)	22.54 (11.74)	25.40 (10.37)	24.77 (10.40)
Median	23.0	20.5	27.0	21.5
IQ Range	[16.5, 30.5]	[13.5, 31.0]	[20.0, 33.0]	[16.0, 32.0]
Min, Max	[7.0, 47.0]	[2.0, 49.0]	[7.0, 47.0]	[11.0, 46.0]
<2 month>				

N	45	52	25	26
Mean (SD)	19.47 (11.08)	14.71 (10.12)	22.36 (11.51)	14.88 (8.90)
Median	20.0	12.5	22.0	14.5
IQ Range	[9.0, 27.0]	[7.0, 21.0]	[17.0, 29.0]	[9.0, 20.0]
Min, Max	[0.0, 48.0]	[0.0, 40.0]	[0.0, 48.0]	[1.0, 36.0]
<3 month>				
N	36	37	23	19
Mean (SD)	17.67 (12.66)	12.03 (10.37)	19.83 (13.99)	13.89 (12.10)
Median	15.5	10.0	19.0	12.0
IQ Range	[7.0, 24.0]	[5.0, 16.0]	[8.0, 28.0]	[5.0, 20.0]
Min, Max	[2.0, 49.0]	[0.0, 53.0]	[2.0, 49.0]	[0.0, 53.0]
<5 month>				
N	29	29	19	15
Mean (SD)	16.28 (12.79)	9.24 (7.99)	20.53 (12.93)	10.93 (7.31)
Median	19.0	8.0	23.0	10.0
IQ Range	[3.0, 24.0]	[3.0, 12.0]	[9.0, 27.0]	[5.0, 15.0]
Min, Max	[0.0, 47.0]	[0.0, 30.0]	[0.0, 47.0]	[0.0, 29.0]
<8 month>				
N	25	27	16	15
Mean (SD)	13.60 (11.47)	8.85 (6.09)	18.00 (11.49)	10.73 (5.99)
Median	13.0	9.0	18.0	11.0
IQ Range	[2.0, 20.0]	[3.0, 13.0]	[13.0, 22.0]	[6.0, 14.0]
Min, Max	[0.0, 40.0]	[0.0, 23.0]	[0.0, 40.0]	[2.0, 23.0]

ベースライン値の定義：pre-treatment の値

部分集団の定義：罹病期間>6 ヶ月である被験者集団

さらに、BDI スコアの経時推移を視覚的に確認するために、全体集団と部分集団別に治療群ごとの BDI スコアの推移をグラフで示した（図 3.7-4 及び図 3.7-5）。ランダム化後、BtB 群が TAU 群よりも平均値で常に低値となっており、抑うつ症状の改善が示唆された。また、BDI スコア平均値の推定精度を評価するために、標準誤差についても併せて示した。部分集団では全体集団よりも対象例数が減少するため、推定精度は低下していることが標準誤差のバーの長さから確認できた。

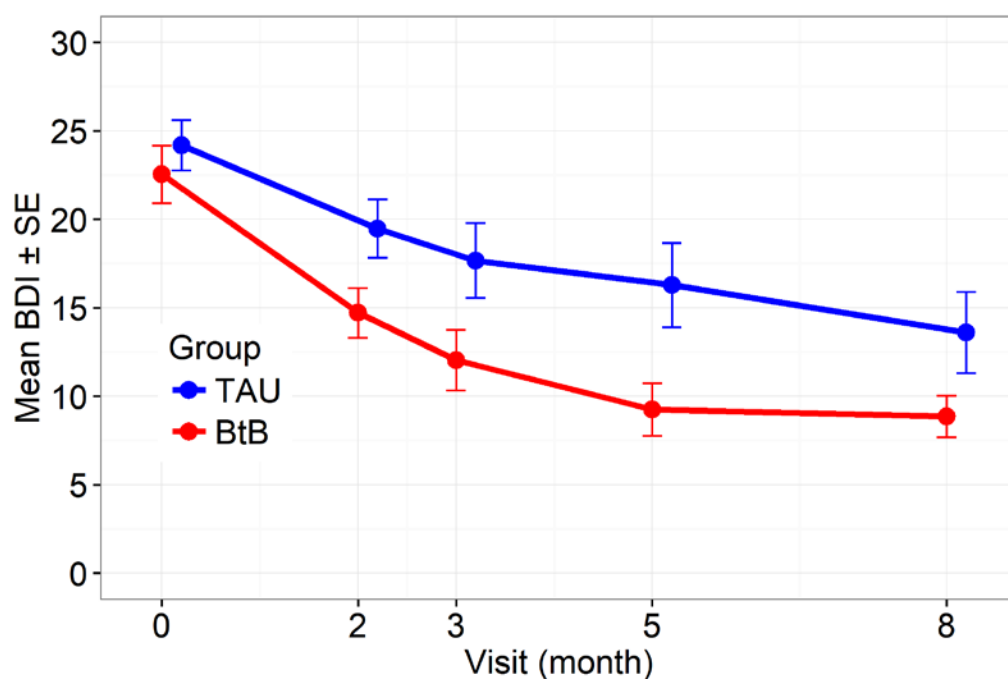


図 3.7-4 BDI 平均値の推移（測定時点ごと，投与群ごと）[全体]

BDI スコアの要約統計量をプロット

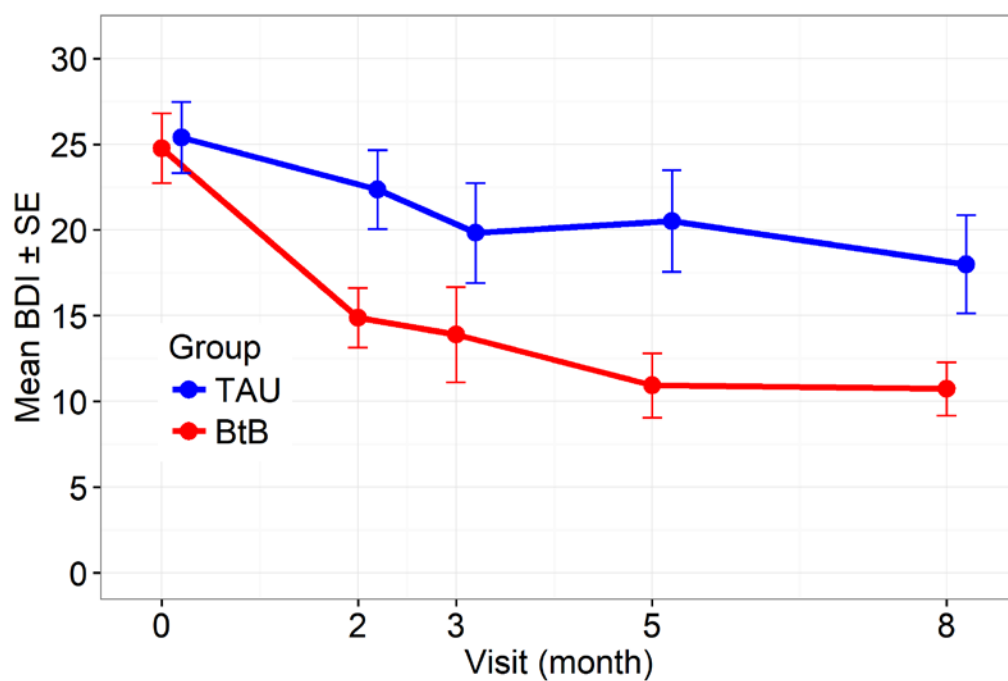


図 3.7-5 BDI 平均値の推移（測定時点ごと，投与群ごと）[部分集団]

BDI スコアの要約統計量をプロット

提案法及び比較方法の解析では，被験者を変量効果とした混合効果モデルを用いた。部分集団はうつ症状の罹病期間が6か月を超える集団と定義しているため， $\gamma_{1i}$ を除外したモデルにより解析を実施した。

$$y_{ij} = \beta_0 + \beta_1 G_i + \beta_2 t_{ij} + \beta_3 G_i t_{ij} + b_{0i} + \gamma_{0i} + \gamma_{1i} + \varepsilon_{ij}$$

被験者は、 $i = 1, 2, \dots, 100$ であり、各被験者ごとに最終測定時点が異なっているため、 $j = 1, 2, \dots, n_i$ となる。また、被験者の変量効果を $b_{0i} \sim N(0, \sigma_b^2)$ とし、誤差を $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ とした。 $\gamma_{0i}$ 及び $\gamma_{1i}$ はそれぞれ抗うつ剤の治療歴有無及び症状の期間（ $\leq 6$ か月、 $>6$ か月）とする。最終測定時点 $j = 8$ における $\beta_{3j}$ が主要評価時点であり、用いた方法ごとに95%信頼区間と $p$ 値を算出した。さらに、参考としてランダム化以降の測定時点についても、同様に解析結果を提示した（表 3.7-8）。

全体集団及び部分集団のいずれにおいても、 $LR_{Bart}$ 及び $LR_{Boot}$ の $p$ 値は従来法である $LR$ よりも保守的な値を示す傾向が認められた。また、 $MMRM_{KR}$ による $p$ 値よりも $LR_{Bart}$ 及び $LR_{Boot}$ の $p$ 値は保守的であり、シミュレーションで得られた結果とも一致する。3か月目の解析結果に注目すると、 $MMRM_{KR}$ では、全体集団及び部分集団のいずれの解析結果も $p$ 値が有意水準5%未満であり慣用的には有意な結果と解釈できる。一方、 $LR_{Bart}$ 、 $LR_{Boot}$ 、 $LR$ では $p$ 値は5%以上であり、仮に5%有意水準で治療効果に関する群間差の仮説検定を実施した場合には結論が異なることになる。

表 3.7-8 認知行動療法の臨床試験における群間差及びその95%信頼区間（測定時点ごと）

Item	Overall		部分集団	
	Estimate [95% CI]	$p$ -value	Estimate [95% CI]	$p$ -value
<2 month>				
$LR_{Bart}$	-5.93 [-7.06, 0.51]	0.0898	-8.85 [-13.02, -0.67]	0.0306
$LR_{Boot}$	-5.93 [-7.09, 0.55]	0.0916	-8.85 [-12.97, -0.72]	0.0290
$LR$	-5.93 [-7.00, 0.45]	0.0844	-8.85 [-13.00, -0.69]	0.0301
$MMRM_{KR}$	-5.93 [-10.37, -1.48]	0.0095	-8.84 [-14.98, -2.71]	0.0056
$t$ -test	-4.76 [-9.03, -0.48]	0.0296	-7.48 [-13.25, -1.70]	0.0123
<3 month>				
$LR_{Bart}$	-5.49 [-7.21, 1.61]	0.2073	-8.08 [-12.76, 0.73]	0.0790
$LR_{Boot}$	-5.49 [-7.17, 1.57]	0.2083	-8.08 [-12.80, 0.77]	0.0796
$LR$	-5.49 [-7.11, 1.50]	0.1966	-8.08 [-12.71, 0.68]	0.0769
$MMRM_{KR}$	-5.49 [-10.78, -0.20]	0.0422	-8.08 [-15.61, -0.54]	0.0363
$t$ -test	-5.64 [-11.03, -0.25]	0.0407	-5.93 [-14.18, 2.32]	0.1541
<5 month>				
$LR_{Bart}$	-4.58 [-6.51, 2.79]	0.4155	-6.79 [-11.92, 2.63]	0.1979
$LR_{Boot}$	-4.58 [-6.59, 2.87]	0.4119	-6.79 [-12.01, 2.73]	0.1923
$LR$	-4.58 [-6.39, 2.66]	0.4034	-6.79 [-11.76, 2.46]	0.1881
$MMRM_{KR}$	-4.57 [-9.95, 0.80]	0.0942	-6.78 [-14.27, 0.71]	0.0747
$t$ -test	-7.03 [-12.65, -1.42]	0.0150	-9.59 [-17.22, -1.97]	0.0153
<8 month>				
$LR_{Bart}$	-3.37 [-5.63, 4.23]	0.7725	-5.60 [-11.37, 4.33]	0.3620
$LR_{Boot}$	-3.37 [-5.66, 4.26]	0.7771	-5.60 [-11.46, 4.42]	0.3692
$LR$	-3.37 [-5.42, 3.00]	0.7627	-5.60 [-11.03, 3.98]	0.3414
$MMRM_{KR}$	-3.34 [-8.02, 1.33]	0.1577	-5.57 [-12.45, 1.32]	0.1100
$t$ -test	-4.75 [-9.81, 0.31]	0.0654	-7.27 [-14.07, -0.46]	0.0372

ベースライン値の定義：pre-treatment の値

部分集団の定義：罹病期間>6ヶ月である被験者集団



図 3.7-6 に最終測定時点である 8 か月目において、パラメトリックブートストラップ法によりリサンプリングした尤度比検定統計量の経験分布に関するヒストグラムを示した。リサンプリングは 3000 回であり、上側が全体集団、下側が部分集団のヒストグラムとなっている。それぞれのヒストグラムの垂直な青点線は経験分布の平均値を示しており、全体集団及び部分集団でいずれも 1.09 であった。また、垂直な赤点線は経験分布の 95% 点を示しており、全体集団及び部分集団で、それぞれ 4.24 及び 4.27 であった。

本事例では、帰無仮説下で尤度比検定統計量は漸近的に自由度 1 のカイ 2 乗分布に従うことから、経験分布の理論的な平均は自由度と等しく 1、95% 点は 3.84 となるはずである。実際にリサンプリングされた経験分布の平均値及び 95% 点と比較してみると、平均値や 95% 点は理論値よりも大きくなっている。これらの結果から、欠測を伴う小標本の経時測定データにおける尤度比検定統計量に従う分布は理論上のカイ 2 乗分布よりも平均が大きくなる方向にシフトしていることが分かる。パラメトリックブートストラップ法による Bartlett 補正は、このようなシフトを補正するために有効な方法と考えられる。

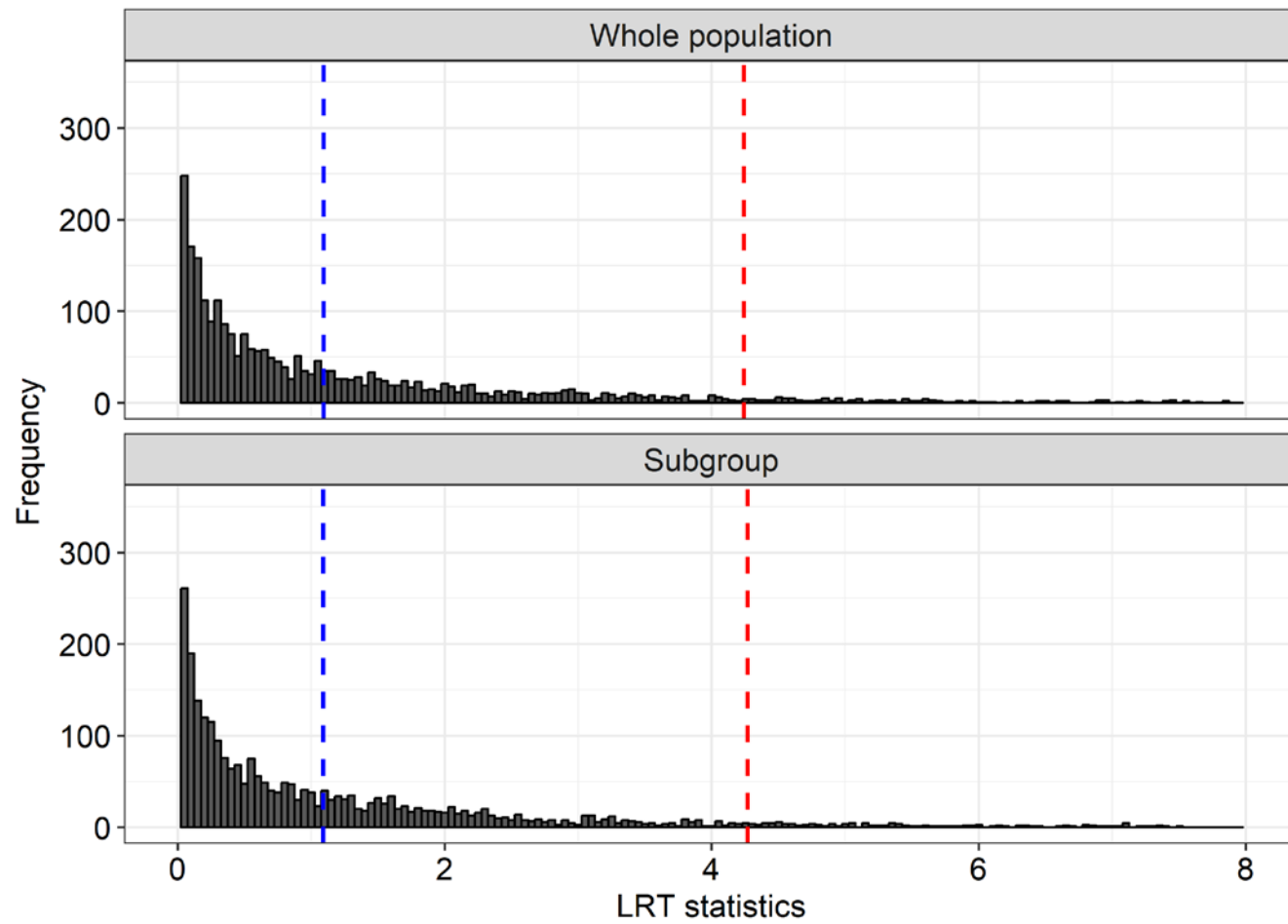


図 3.7-6 認知行動療法の臨床試験におけるパラメトリックブーストラップ法による尤度比検定統計量の経験分布<sup>†</sup>（8 ヶ月時点）

<sup>†</sup> 青点線，経験分布の平均（全体集団では 1.09，部分集団では 1.09）；赤点線，経験分布の 95% 点（全体集団では 4.24，部分集団では 4.27）

### 3.7.3 クロスオーバー試験

#### 3.7.3.1 クロスオーバー試験とは

2種類の治療 A 及び B の効果を比較する状況を考える。通常の臨床試験では、被験者をランダムに治療群 A または B に割り付けて比較試験を実施する。しかし、比較試験を選択した場合、対象とするアウトカムに関する全体の変動は被験者内変動に被験者間変動を加えたものとなり必要症例数が増大する。そこで、各被験者で治療 A と B の両方を行い、それぞれの効果を観察するクロスオーバー試験が実施されることがある。ここでは、最も簡単な AB/BA (2 剤 2 期) クロスオーバー試験を考えることにする。

この試験デザインでは、試験期間 ( $k = 1, 2$ ) として期間 1 と期間 2 を設定し各被験者を群 1 ( $n_1$  例) 及び群 2 ( $n_2$  例) へ割り付ける。群 1 とは期間 1 に治療 A、期間 2 に治療 B を受ける群であり、群 2 では順番が逆になり期間 1 で治療 B、期間 2 で治療 A を受ける群となる。2×2 クロスオーバー試験デザインを表 3.7-9 にまとめた。

表 3.7-9 クロスオーバー試験の記法

群 ( $i$ )	例数	結果	期間 ( $k = 1$ )		期間 ( $k = 2$ )	
			治療	結果	治療	結果
1	$n_1$	$Y_{1jk}$	A	$Y_{1j1}$	B	$Y_{1j2}$
2	$n_2$	$Y_{2jk}$	B	$Y_{2j1}$	A	$Y_{2j2}$

治療に対する結果を  $Y_{ijk}$  とおく。ここで、 $i$  は群 ( $i = 1, 2$ )、 $j$  は被験者 (群 1 :  $j = 1, 2, \dots, n_1$ , 群 2 :  $j = 1, 2, \dots, n_2$ )、そして  $k$  は期間 ( $k = 1, 2$ ) を表すものとする。

結果を  $Y_{ijk}$  について以下の線形混合効果モデルを想定する。

$$Y_{ijk} = \mu + \alpha_i + b_{ij} + \beta_k + \tau_{ik} + \varepsilon_{ijk} \quad (*)$$

$\alpha_i$  は  $i$  番目の群の固定効果 ( $i = 1, 2$ )、 $b_{ij}$  は  $i$  番目の群、 $j$  番目の被験者の変量効果であり、 $N(0, \sigma_b^2)$  に従うものとする。ここで、 $\sigma_b^2$  は被験者間分散である。 $\beta_k$  は  $k$  番目の期間の固定効果 ( $k = 1, 2$ ) とする。 $\tau_{ik}$  は  $i$  番目の群、 $k$  番目の期間に対する薬剤の固定効果である。例えば、表 3.7-9 に示した例では  $\tau_{11} = A$ 、 $\tau_{12} = B$  となる。 $\varepsilon_{ijk}$  は被験者内誤差 (変動) であり、 $N(0, \sigma_\omega^2)$  に従うものとする。ここで、 $\sigma_\omega^2$  は被験者内分散である。

結果変数  $Y_{ijk}$  が従う周辺モデルは、 $Y_{ijk}$  が (\*) に示した線形混合効果モデルであることを考えると、次のような多変量正規分布となる。

$$\begin{pmatrix} Y_{ij1} \\ Y_{ij2} \end{pmatrix} \sim MVN \left[ \begin{pmatrix} \mu + \alpha_i + \beta_1 + \tau_{i1} \\ \mu + \alpha_i + \beta_2 + \tau_{i2} \end{pmatrix}, \begin{pmatrix} \sigma_b^2 + \sigma_\omega^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \sigma_\omega^2 \end{pmatrix} \right]$$

産後うつ臨床試験や認知行動療法臨床試験では、結果変数が従う周辺分布の分散共分散構造には恣意性を排除するため、unstructured (無構造) を仮定した。一方、クロスオーバー試験で結果変数  $Y_{ijk}$  が従う周辺分布の分散共分散構造は、compound symmetry (複合対称構造) を仮定する方が自然であることが分かる。

したがって、シミュレーション実験や産後うつ臨床試験及び認知行動療法臨床試験の線形混合効果モデルの分散共分散構造で仮定した unstructured (無構造) ではなく、本セクションで紹介するクロスオーバー試験の解析では、compound symmetry (複合対称構造) を適用して事例の解析を行う。

### 3.7.3.2 気管支拡張剤のクロスオーバー試験

Senn and Auclair (1990)は気管支拡張剤のクロスオーバー試験データの提示方法について論文で個別値と掲載し報告している。Senn and Auclair (1990)が報告しているクロスオーバー試験データの一部を用いて、提案法及び比較法に関する解析を行った。

本データは、13 人の 7 歳から 14 歳までの中等度または重症喘息の小児に対して、肺機能の測定値である最大呼吸流量（Peak expiratory flow: PEF）を測定した 2 剤 2 期のクロスオーバー試験データである。当時、既に薬効が確立されていた気管支拡張剤であるサルブタモール 200  $\mu$ g の吸引 1 回と新規に開発された薬剤であるホルモテロール 12  $\mu$ g を比較する試験であった。子供達は、2 つの群のうちの 1 つにランダム割り付けられた。1 つの群では、早朝にホルモテロールを投与され、投与後 8 時間目まで試験実施施設で経時的に PEF が測定された。投与後 8 時間目の測定後、子供は一時帰宅し自宅で両親が簡易測定器を用いて投与後 10, 11 及び 12 時間目で PEF を測定した。次に少なくとも 1 日間のウォッシュアウト（投与された薬剤の効果を減弱させるための期間）を置き、再び試験実施施設に来院した。来院後サルブタモール投与を 1 回受け、その後同様に試験実施施設及び自宅で PEF を測定した。もう一方の群では、最初にサルブタモールの投与を受けた後、ウォッシュアウトを経てホルモテロールの投与を受ける以外は 1 つ目の群と手順は同一である。クロスオーバー試験の模式図を図 3.7-7 に示す。

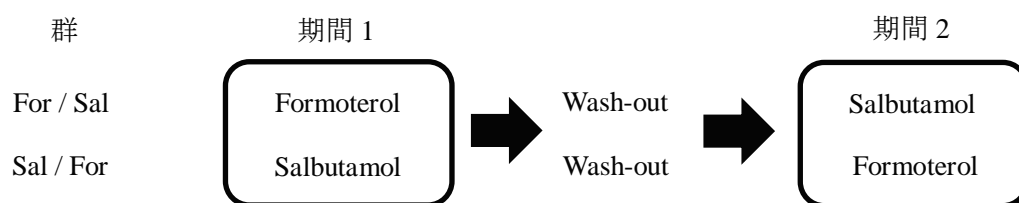


図 3.7-7 気管支拡張剤クロスオーバー試験の模式図

For, ホルモテロール; Sal, サルブタモール

解析では、各期間における投与後 8 時間目の PEF を抽出して用いることとした。投与後 8 時間目の PEF に関する被験者ごとの個別値を表 3.7-10 に示す。

表 3.7-10 各期間における投与後 8 時間目の PEF（L/min）個別値一覧

群	被験者番号	期間 1	期間 2
For / Sal	1	310	270
	4	310	260
	6	370	300
	7	410	390
	10	250	210
	11	380	350
	14	330	365
Sal / For	2	370	385
	3	310	400
	5	380	410
	9	290	320
	12	260	340
	13	90	220

群ごとに各被験者の PEF 測定値をプロットして図 3.7-8 に示した。プロットからサルブタモールと比較してホルモテロールで最大呼気流量が改善している様子が認められた。

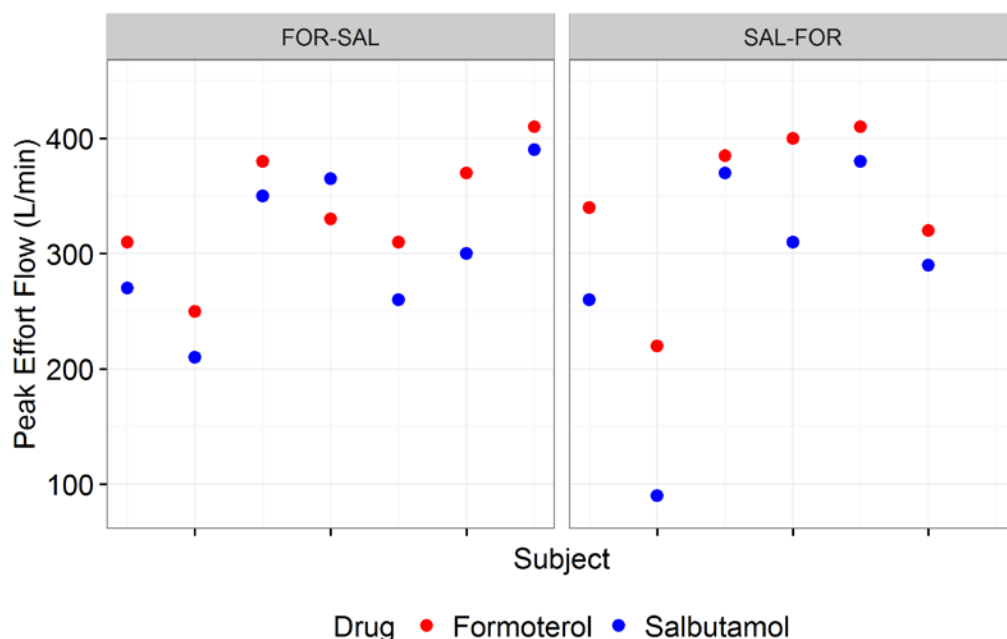


図 3.7-8 群ごとの PEF 個別値プロット

気管支拡張剤のクロスオーバー試験データを提案法及び比較方法により解析した。解析では、3.7.3.1 で説明した線形混合効果モデルを用いた。また、測定値の周辺分布が従う多変量正規分布の分散共分散構造には、compound symmetry（複合対称構造）を仮定した。解析結果を表 3.7-11 及び表 3.7-12 に示す。

従来法の LR と比較して、LR<sub>Bart</sub> 及び LR<sub>Boot</sub> では  $p$  値が保守的な値を示した。MMRM<sub>KR</sub> の  $p$  値と比較しても LR<sub>Bart</sub> 及び LR<sub>Boot</sub> の  $p$  値は保守的な値を与える傾向が認められており、これはシミュレーション実験による結果と一貫していた。

表 3.7-11 PEF に関する薬剤間差とその 95%信頼区間

解析方法	PEF (L/min)		
	薬剤間差 ( <i>Formoterol</i> – <i>Salbutamol</i> )	薬剤間差の 95% 信頼区間	$p$ -value
LR <sub>Bart</sub>	46.61	[23.06, 70.16]	0.0011
LR <sub>Boot</sub>	46.61	[22.73, 70.49]	0.0013
LR	46.61	[25.66, 67.56]	0.0003
MMRM <sub>KR</sub>	46.61	[22.89, 70.33]	0.0012
$t$ -test	45.38	[20.85, 69.91]	0.0017

周辺分布の分散共分散構造に仮定した構造、Compound symmetry (複合対称構造)  
リサンプリング回数、3000 回； $t$ -test は対応のある  $t$  検定

表 3.7-12 被験者間分散及び被験者内分散

解析方法	被験者間分散	被験者内分散
REML	4846.54	750.41
ML	4100.92	634.96

### 3.8 考察

#### 3.8.1 欠測の単調性/非単調性

シミュレーションでは、解析対象となるデータに単調な欠測を仮定した。MAR の下で MMRM は妥当なパラメータ推定を行えるため、提案法は非単調な欠測であっても適用可能であり同様な性能を発揮することが期待できる。つまり、対象とするデータの欠測が単調か非単調であるかは問題ではない。このことは、MAR の下で MMRM によりパラメータがどのように推定されるかを理解することで容易に予想できる。

被験者  $i$  ( $i = 1, 2, \dots, N$ ) の結果変数を  $Y_i$  とする。  $Y_i$  は観測データ  $Y_i^o$  と欠測データ  $Y_i^m$  から構成されるので、  $Y_i = (Y_i^o, Y_i^m)$  とできる。被験者  $i$  に対する欠測識別変数を  $R_i$ 、結果変数の分布のパラメータを  $\theta$ 、欠測識別変数の分布のパラメータを  $\varphi$  とすると、観測データの尤度関数  $L(\theta, \varphi)$  は、次の式になる。

$$L(\theta, \varphi) = \prod_{i=1}^N \int f(Y_i^o, Y_i^m, R_i | \theta, \varphi) dY_i^m$$

ところで、MAR とは欠測メカニズムが結果変数  $Y_i$  が与えられた下での確率に基づいていると考えることなので、

$$\text{欠測メカニズム} = f(R_i | Y_i, \varphi)$$

と書ける。また、これは欠測メカニズムが結果変数のパラメータ  $\theta$  に依存していないことも同時に仮定している。

結果変数  $Y_i$  の確率密度関数  $f(Y_i^o, Y_i^m, R_i | \theta, \varphi)$  が  $f(Y_i | \theta)$  と  $f(R_i | Y_i, \varphi)$  の積に分解できるとすれば、

$$L(\theta, \varphi) = \prod_{i=1}^N \int f(Y_i^o, Y_i^m | \theta) f(R_i | Y_i^o, Y_i^m, \varphi) dY_i^m$$

MAR の定義から、  $f(R_i | Y_i, \varphi) = f(R_i | Y_i^o, Y_i^m, \varphi) = f(R_i | Y_i^o, \varphi)$  なので、

$$L(\theta, \varphi) = \prod_{i=1}^N \int f(Y_i^o, Y_i^m | \theta) f(R_i | Y_i^o, \varphi) dY_i^m$$

$$L(\theta, \varphi) = \prod_{i=1}^N \int f(Y_i^o | \theta) f(R_i | Y_i^o, \varphi) dY_i^m$$

したがって、観測データの対数尤度は、

$$l(\theta, \varphi) = \sum_{i=1}^N \log f(Y_i^o | \theta) + \sum_{i=1}^N \log f(R_i | Y_i^o, \varphi)$$

となる (Matsuyama 2004)。MAR の下では、興味のあるパラメータ  $\theta$  の推定は、欠測過程  $f(R_i | Y_i^o, \varphi)$  を含めない  $f(Y_i^o | \theta)$  に基づいて行うことができる。これが直接尤度 (Direct Likelihood: DL) であり、パラメータ  $\theta$  の推定に欠測メカニズムは関与していない。そのため、欠測パターンが単調か非単調であるかもパラメータ  $\theta$  の推定に全く関係ない。このことは、欠測メカニズム MCAR の場合にも当てはまるので、MCAR と MAR は無視可能な欠測メカニズムと呼ばれる (Laird 1988)。したがって、シミュレーションでは単調な欠測のみを仮定し非単調な欠測について検討していないが、仮に欠測パターンが非単調であったとしても単調な場合と同様な性能が期待できると考えられた (Peace 2018)。

#### 3.8.2 KR 法が過度に保守的となる原因

帰無仮説下における MCAR を仮定した場合には、Kenward and Roger 法や提案法は、ほぼ同

様な Type I error rate を示しており、名目有意水準の 5% に近い値であった。しかしながら、同じ帰無仮説下での MAR を仮定した場合では、累積脱落割合が増加するにしたがい KR 法は提案法よりも保守的な Type I error rate を示した。本来 KR 法は、小標本における漸近近似を改善した方法として知られていたため、本研究ではさらに現象の原因を検討した。

KR 法は小標本下における統計的推測の問題を改善するために提案されており、 $t$  検定における自由度を補正するだけでなく、固定効果の分散の推定値も補正する方法である。KR 法 (Kenward and Roger 1997) に基づく固定効果パラメータの検定サイズは、実践的な条件下で、Satterthwaite 法と比べてより名義の水準に近いことが指摘されている (Gosho et al. 2017)。本研究におけるシミュレーション結果では、欠測メカニズムが MAR で累積欠測割合が 40% となる状況において、KR 法は過度に保守的な Type I error rate を示した。

本検討の結果から、Type I error rate の大きさは以下に示す順番であった。

$$\text{MMRM}_{\text{BW}} > \text{MMRM}_{\text{SAT}} > \text{MMRM}_{\text{KR}}$$

$\text{MMRM}_{\text{SAT}}$  は検討した状況下では、名義水準 5% を超える Type I error rate であった。一方、 $\text{MMRM}_{\text{KR}}$  は MAR の下で名義水準 5% を下回り、過度に保守的な結果を示した。本検討における Satterthwaite 法と KR 法の違いは固定効果の分散の推定だけである。したがって、これらの結果から KR 法では固定効果の分散の推定値を適切に補正できていない可能性が考えられた。

欠測がない状況での検討ではあるものの、本検討と類似した設定での研究結果が報告されている (Arnau et al. 2009)。Arnau et al. (2009) によると、1 群 10 例の経時測定データを模したシミュレーション結果より Type I error rate の大きさは、 $\text{MMRM}_{\text{BW}} > \text{MMRM}_{\text{SAT}} > \text{MMRM}_{\text{KR}}$  となっている。この結果は、本検討の結果と一致していることが確認できた。

固定効果の分散の推定について、KR 法が提案された論文で導出されている補正式の主要な部分を以下にレビューする。詳しい記法は、各論文を参照することとする。

固定効果パラメータ  $\hat{\beta}$  の分散  $V(\hat{\beta})$  は Kackar and Harville (1984) より 2 つの項に分解できる。

$$V(\hat{\beta}) = \Phi + \Lambda$$

$\Phi$  はモデルによる  $V(\hat{\beta})$  の推定量、 $\Lambda$  は  $V(\hat{\beta})$  の漸近的な分散共分散行列が分散を過小推定している程度を表している。別の表現をすれば、 $\Lambda$  を  $\Phi$  に加えることにより過小推定された分散を本来あるべき分散の大きさに戻している。

Kenward と Roger は  $\hat{\Phi}$  について次のようなテイラー展開を考えた。

$$\hat{\Phi} \cong \Phi + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial \Phi}{\partial \sigma_i} + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2 \Phi}{\partial \sigma_i \partial \sigma_j}$$

この期待値を取ると、

$$E[\hat{\Phi}] \cong E[\Phi] + \sum_{i=1}^r E\{(\hat{\sigma}_i - \sigma_i)\} \frac{\partial \Phi}{\partial \sigma_i} + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r \omega_{ij} \frac{\partial^2 \Phi}{\partial \sigma_i \partial \sigma_j} + O(N^{-3})$$

となる。ここで、 $\omega_{ij} = E\{(\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j)\}$  としている。

$\hat{\sigma}_i$  は与えられたデータから REML 推定により不偏推定量を求め、求めた値をプラグインすることになる。1997 年当時は  $E\{(\hat{\sigma}_i - \sigma_i)\} = 0$  として  $E[\hat{\Phi}]$  の第 2 項を無視していた。

したがって、

$$E[\hat{\Phi}] = \Phi + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r \omega_{ij} \frac{\partial^2 \Phi}{\partial \sigma_i \partial \sigma_j}$$

を補正式の導出に用いていた。

実際には、小標本における $V(\hat{\beta})$ の推定において第2項が無視できない誤差を生み出していることが次第に明らかとなった。2009年、Kenward と Roger は当初無視していた第2項 $E\{(\hat{\sigma}_i - \sigma_i)\}$ を無視せずに固定効果パラメータの分散の推定値を改善した補正式を提案した(Kenward and Roger 2009)。改善した補正式を KR2 と呼ぶことにする。

補正式を比較するため、KR と KR2 の式を以下に並べて示す。

$$\Phi_A^{OLD} = \hat{\Phi} + 2\hat{\Phi} \left\{ \sum_{i=1}^r \sum_{j=1}^r \omega_{ij} \left( Q_{ij} - P_i \hat{\Phi} P_j - \frac{1}{4} R_{ij} \right) \right\} \hat{\Phi}$$

$$\Phi_A^{NEW} = \hat{\Phi} + 2\hat{\Phi} \left\{ \sum_{i=1}^r \sum_{j=1}^r \omega_{ij} \left( Q_{ij} - P_i \hat{\Phi} P_j - \frac{1}{4} R_{ij} \right) \right\} \hat{\Phi} - \frac{1}{4} \sum_{s,t=1}^r \omega_{st} V_t \Phi P_s \Phi$$

KR2 では KR よりも $V(\hat{\beta})$ が小さく推定されるため、Type I error rate は MMRM<sub>KR</sub> よりも大きくなり過度に保守的な傾向は、改善されるはずである。

しかしながら、Kenward and Roger が提案した補正式 (KR2) はそもそも分散共分散行列のパラメータが非線形である場合にバイアスの補正を目指したものであり、線形である場合には当初提案された KR と等価であることが示されている (Kenward and Roger 2009)。Kenward and Roger の定義によれば、分散共分散行列が無構造 (Unstructured) である場合に線形と分類している。つまり、本検討では分散共分散行列を無構造としているため、KR2 と KR の結果は変わらないはずである。念のため、KR2 を追加し 1000 回反復のシミュレーションにより KR と等価であることを確認した (KR と結果は変わらないため、図表に結果は表示していない)。

本検討で、Kenward and Roger が提案した KR 及び KR2 のいずれの方法を用いたとしても、分散共分散行列を無構造としてパラメータ推定した場合には、小標本かつ欠測を伴う状況下において固定効果の推定値に関する標準誤差を過大推定している可能性が示唆された。KR 法が名目有意水準を維持できない状況下であっても、提案法は適切に有意水準を維持しており有用性が認められた。

### 3.8.3 t検定の MAR 下における性能

群間で異なる脱落割合となる MAR 下において、t検定は MMRM と比較して検出力が低下することが指摘されている (Ashbeck and Bell 2016)。1 群 10 例のシミュレーション結果では、MMRM<sub>KR</sub> と t検定の検出力はほぼ同等であり、t検定の明確な検出力の低下は確認できなかった。一方、1 群 20 例のシミュレーション結果では、MMRM<sub>KR</sub> よりも t検定の検出力は低下しており、既に指摘されている知見を確認することができた。

1 群 10 例の設定で既知の知見が確認できなかった原因として、脱落割合が大きい場合に MMRM<sub>KR</sub> は小標本であるためパラメータの収束が得られないケースが多いことが考えられた。また、シミュレーション設定は Gosho et al. (2017)を参考にしているが、設定に強く依存した結果である可能性も否定できない。原因に関する検討は、今後の検討課題である。

### 3.8.4 リサンプリング回数

提案法のアルゴリズム 1 では、帰無仮説下における検定統計量の期待値を推定するためにブートストラップ法を用いている。一方、アルゴリズム 2 では、ブートストラップ法により帰無仮説下の検定統計量が構成する分布の上側分位点 (実際には 95%点) を推定している。いずれ



のアルゴリズムにおいても、推定精度を向上させるためにはリサンプリング回数を十分多くしておく必要がある (Cordeiro and Cribari-Neto 2014)。本研究のシミュレーション実験では、リサンプリング回数を 3000 回としたが、別途リサンプリング回数の適切性について検討した。リサンプリング回数を変えて提案法の Type I error rate の挙動を確認した結果、概ね 1000 回以上のリサンプリング回数であれば少なくとも本研究の状況設定では大きな問題にはならないと考えられた。

### 3.8.5 ブートストラップ法による Bartlett 補正項の推定値

欠測が発生していない状況では、ブートストラップ法によりほぼ大きなばらつきなしに Bartlett 補正項を推定できた。同一被験者数において累積欠測割合が増加すると推定された Bartlett 補正項のばらつきは増加した。これは平均的には累積欠測割合だけ欠測が発生するが、シミュレーションの各回で対象者数は異なっているため、ブートストラップ法により推定される Bartlett 補正項にもばらつきが発生しているためと考えられた。

また、対象者数が増加するとブートストラップ法により推定された Bartlett 補正項のばらつきは小さくなる傾向が認められた。これは、対象者数の増加により観測値の数も増加することで推定精度が向上したためと考えられた。

### 3.8.6 収束条件

セクション 3.6.4 で、収束が得られる最終測定時点における最低標本数について触れた。最低標本数は対象データの測定回数により異なるが、収束が得られない場合に尤度比検定統計量の計算に必要な対立仮説下の対数尤度が求まらないため、尤度比検定ができないことになる。これは提案法の限界でもあるが、同時に尤度比検定の限界でもある。

本研究では結果変数 $\mathbf{Y}$ の分散共分散構造を無構造 (unstructured) とした。無構造は、すべての時点間における分散及び共分散が異なると仮定した柔軟な構造であり、解析者の恣意性が入りにくいと考えられている。また、分散共分散構造を誤特定する可能性が低いことから多くの研究者も無構造の使用を推奨している (Mallinckrodt et al. 2008)。

しかしながら、無構造は未知のパラメータを多く含むため、小標本及び時点数が多くさらに欠測により情報量が減少している場合に、収束値が得られないことがある。収束が得られない場合には、より単純な分散共分散構造を仮定して解析を進める。単純な分散共分散構造を指定すれば、収束の問題は回避しやすいが分散共分散構造を誤特定してしまう可能性は高まる。実際、誤った分散共分散構造を指定すると名義の有意水準を維持できない問題が発生する (Gosho et al. 2017)。

### 3.8.7 提案法の限界

本研究では、 $N = 10$  (1 群 5 例) の欠測がない完全データに対して同様にシミュレーション実験を行った。KR 法の性能が最良であり、提案法は 2 番目に良い結果であった。 $N = 10$  で欠測が発生した場合に、経時測定データの標本数が少なすぎるため MMRM によるパラメータ推定がほとんど収束せず、結果として推定がほとんど行えない状況となる。したがって、提案法のアルゴリズムを適用することができない。これは、パラメータ推定に反復計算を採用している多くの従来法及び提案法での限界である。

恣意性を排除することを目的に分散共分散構造に無構造を仮定することが多いが、収束の間

題が発生した場合に、パラメータ数が少ない構造（例えば、複合対称）へ変更することを検討するのも一案と考える。ただし、医薬品開発の臨床試験では収束の問題が発生した場合の対処方法について事前に定義して計画を文書化することが必要である。

### 3.8.8 結論

MMRM におけるブートストラップ法を用いて尤度比検定統計量の経験分布を推定し、高次漸近理論に基づく近似推測手法を提案した。提案法が、欠測を伴う小標本の経時測定データに対して、帰無仮説の下で第 1 種の過誤確率を適切に維持し、また対立仮説の下で従来法とほぼ同等の検出力を示すことをシミュレーション実験により確認を行った。欠測メカニズムが MAR の下では、小標本における漸近近似を用いた改善法である Kenward and Roger 法よりも性能が良い可能性を示した。提案法を産後うつ臨床試験、認知行動療法臨床試験及び気管支喘息治療薬のクロスオーバー試験データに適用し、比較群間に差がない場合に従来の尤度比検定よりも保守的な結果となる傾向を示した。また、比較群間に差がある場合に適切に差を検出する結果となる傾向を示した。

## 4. 並べ替え検定による推測手法

### 4.1 背景と目的

本節より MMRM に並べ替え検定を適用した方法について述べる。MMRM は時点間相関を柔軟にモデリング可能であり臨床試験の経時測定データ解析では現在標準的な解析方法として普及している。しかしながら、大標本近似に基づく推定方法であるため、小標本及び欠測が発生する状況下で Type I error rate のインフレーションが認められている点がデメリットである。

小標本下における統計的推測の有効なアプローチとして、古くから並べ替え法に基づく推測手法が広く用いられてきた (Good 2005; Mielke and Berry 2007)。並べ替え検定の起源は、1935 年 Fisher R. A. の「The Design of Experiments」にさかのぼる (Fisher 1935)。並べ替え検定は観察された値に基づいており、有限標本下であっても第 1 種の過誤確率を名義水準にコントロールする正確検定である。並べ替え検定は小標本であっても対象データを並べ替える手順により検定統計量の経験分布を推定する。すべての並べ替えが作成可能であれば、正確な検定となる点がメリットである。標本数が多くなると並べ替える組み合わせ数が膨大となり計算コストが増大する。事実上すべての並べ替えを作成することが不可能となるため、実践的にはランダムに発生した適当な数の並び替えに基づいて Monte Carlo 検定を行う。

並べ替え検定は、データがランダム化されていること及びデータの交換可能性が仮定できることが必要となる。交換可能性とは、例えば観測値ベクトルを  $\mathbf{Y}$  とし、その並べ替えたベクトルを  $\mathbf{Y}^*$  とした時に、 $\mathbf{Y}^*$  が  $\mathbf{Y}$  と同じ分布を持つことである (Commenges 2003)。これまでに固定効果のみを仮定した線形モデルに並べ替え検定を適用する方法については多数の先行研究が発表されている (Anderson and Robinson 2001)。しかし、経時測定データの解析におけるマルチレベルモデルを対象とした並べ替え検定に関する研究は少数である。さらに経時測定データでほぼ確実に発生する欠測について扱っている研究はほとんどない。単純に線形モデルにより回帰係数の検定統計量を計算したのでは、欠測が発生している場合に大きなバイアスが入る可能性がある。一方、欠測メカニズムが MAR の下で妥当な統計的推測を与える MMRM を用いれば、欠測データに対しても対応可能となる。

これまでに固定効果のみを仮定した線形モデルに並べ替え検定を適用する方法について、主に 3 つの方法が提案されている。

1 つ目は、興味ある変数 (ランダム化臨床試験では投与群を表す変数) を並べ替える方法で、Fisher R. A. の考え方を応用した考え方となっている。並べ替え検定の前提となっているランダム化及びデータの交換可能性が成立していれば、線形モデルでも問題なく適用可能な考え方である。

2 つ目は、結果変数を並べ替える方法で、Manly (1997) により提案された。しかしながら、本方法は共変量に外れ値が存在する場合に検定が不安定となり結果として第 1 種の過誤確率が増加することが指摘されている (Kennedy and Cade 1996; Anderson and Legendre 1999)。

3 つ目は、線形モデルの残差を並べ替える方法である。残差を並べ替える方法はさらに大きく 2 つの方法があり、興味ある変数を除きそれ以外の変数を共変量としたモデル (帰無仮説のもとでの回帰モデル) の推定値により得られる残差を利用する方法 (Freedman and Lane 1983)、及び線形モデルのすべての変数を共変量としたフルモデルの推定値により得られる残差を利用する方法 (Ter Braak 1992) がある。

すべての並べ替えを与えることができれば、並べ替え検定は正確な検定である。しかしながら、データによっては並べ替えの組み合わせ数が膨大となり、すべての並べ替えを与えることが計算上実現困難となることがある。この時、並べ替えをランダムサンプリングによって行う Monte Carlo 検定が利用できる (Dwass 1957; Marozzi 2004; Pesarin and Salmaso 2010)。並べ替え検定は、漸近的に名義上の水準 (nominal size) を有しておりパラメトリックな検定と同程度の検出力を持つとされている (Hoeffding 1952)。これらの理由から、漸近的なパラメトリック検定よりも並べ替え検定は、小標本に適しているとされている。

本研究では、欠測を伴う小標本の経時測定データに対して並べ替え検定の考え方を適用した MMRM による解析方法を 2 種類開発した。これら 2 種類の方法は、興味ある回帰係数の  $t$  検定統計量が従う経験分布を Monte Carlo 法により推定する。MMRM は、欠測メカニズムが MAR の下で妥当な統計的推測を与える。これらの方法は、並べ替えたデータセットを従来の MMRM により解析するだけで実現できる簡便な方法である。

さらに、これらの方法を標準的な解析方法である結果変数が従う周辺分布の分散共分散構造に無構造を仮定した Kenward-Roger (KR) 法 (Kenward and Roger 1997) 及び自由度調整を Between-within 法 (Schluchter and Elashoff 1990) とした方法により比較した。実践的な有用性を評価するために産後うつ臨床試験データ (Gregoire et al. 1996) にこれら解析方法を適用した結果についても評価した。

## 4.2 並べ替え検定による提案法

### 4.2.1 並べ替え検定による推測手法における記法

MMRM では、切片、投与群、時点及び投与群と時点の交互作用項でモデリングする。対象者  $i$  における投与群を表すベクトルを  $\mathbf{X}_{ig}$ 、時点をあらわす  $n_i$  次元ベクトルを  $\mathbf{X}_{it_u}$  とする。

固定効果の計画行列  $\mathbf{X}_i$  は、

$$\mathbf{X}_i = (\mathbf{1}, \mathbf{X}_{ig}, \mathbf{X}_{it_1}, \mathbf{X}_{it_2}, \dots, \mathbf{X}_{it_u}, \mathbf{X}_{igt_1}, \mathbf{X}_{igt_2}, \dots, \mathbf{X}_{igt_n})$$

固定効果パラメータベクトル  $\hat{\boldsymbol{\beta}}$  は、

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_g, \hat{\beta}_{t_1}, \hat{\beta}_{t_2}, \dots, \hat{\beta}_{t_n}, \hat{\beta}_{gt_1}, \hat{\beta}_{gt_2}, \dots, \hat{\beta}_{gt_n})^T$$

となる。

対比ベクトルを  $\mathbf{h}^T = (0, 1, 0, \dots, 0, 0, \dots, 1)$  とすると、最終時点の 2 群間の平均値の差は、

$$\mathbf{h}^T \boldsymbol{\beta} = \hat{\beta}_g + \hat{\beta}_{gt_n}$$

となる。MMRM における  $t$  検定統計量  $T(\beta_1^{null})$  を、

$$T(\beta_1^{null}) = \frac{\mathbf{h}^T \hat{\boldsymbol{\beta}}}{\sqrt{\mathbf{h}^T \text{Cov}(\hat{\boldsymbol{\beta}}) \mathbf{h}}}$$

とおく。

並べ替え検定による推測手法における帰無仮説下の  $t$  検定統計量  $T(\beta_1^{null})$  を、上式により定義しこれを基本統計量とする。本研究では、 $T(\beta_1^{null})$  を計算するために REML 推定値を用いた。例えば、本研究のシミュレーション研究での設定である  $N = 20$  で欠測が発生しない場合は、最終測定時点の群間差に関する  $t$  検定統計量は自由度 18 の  $t$  分布に従うこととなる。

## 4.2.2 既知のコンセプト

並べ替え検定には、異なる複数のコンセプトが提案されているが、大きく以下の4種類に整理できる (Anderson and Legendre 1999)。

コンセプト 1	投与群の指示変数を並べ変える方法
コンセプト 2	結果変数を並べ変える方法
コンセプト 3	帰無仮説のもとでの回帰モデルにより推定した残差を並べ替える方法
コンセプト 4	フルモデルにより推定した線形モデルの残差を並べ替える方法

## 4.2.3 既知のコンセプトに対する問題点

コンセプト 2 及び 4 の問題点について言及する。

### 4.2.3.1 コンセプト 2：結果変数の並べ替えの問題点

Manly は、あたかも帰無仮説のもとで共変量の組み合わせに対して結果変数 $Y$ をランダムに割り当てられるように扱うことを提案した (Manly 1997)。すなわち、結果変数 $Y$ をランダム化する操作である。線形モデルの結果変数 $Y$ をランダムに並び替える操作は、帰無仮説のもとで興味ある固定効果パラメータ $\beta_1$ 以外のすべての回帰係数をゼロとみなすことと等価である。帰無仮説のもとで考える点については問題ないが、Manly の方法では線形モデルの変数間の相関を適切に扱えていない点が欠点となる (Ottoboni 2018)。言い換えれば、 $\beta_1$ 以外に非ゼロの回帰係数が存在した場合に誤差 $\epsilon$ と結果変数 $Y$ が全く異なる分布を持つこととなる。結果として推定する $\beta_1$ にバイアスが入ることになり、推奨されていない。限定的なシミュレーションではあるが、投与群指示変数以外の共変量に外れ値が存在する場合に、結果変数 $Y$ を並べ替えた線形回帰をすると Type I error rate がインフレーションするとの報告もある (Kennedy and Cade 1996)。ただし、その後 Manly により提示されたシミュレーション結果 (Manly 1997) は Kennedy and Cade の結果と一貫しておらず、現時点統一したコンセンサスが存在しない。

結果変数 $Y$ を交換可能性が成立するように並び替える必要があるが、その実現は困難でありコンセプト 2 は妥当ではないと考えられた。

### 4.2.3.2 コンセプト 4：フルモデルによる残差の並べ替えの問題点

フルモデルによる残差を並べ替えるアプローチは ter Braak により提案されている (Ter Braak 1992)。本方法は、投与群指示変数の回帰係数についても考慮している点が帰無仮説のもとでの回帰モデルと異なっている。フルモデルによる残差を考えることで興味ある回帰係数 $\beta_1$ の分散を減少させ検出力が向上することが期待される (Ter Braak 1992)。手順の途中で並べ替えた結果変数 $Y^{(b)}$ をフルモデルで解析し、 $t$ 検定統計量を取得する。このときそのままでは対立仮説下の検定統計量を取得してしまうことになる。興味ある検定はあくまでも、 $H_0: \beta_1 = \beta_1^{null}$ なので検定統計量を計算し直す手順を加える必要がある。

帰無仮説下の検定統計量を計算し直すプロセスに妥当な方法を適用することが現時点困難であり、コンセプト 4 は妥当ではないと考えられた。

## 4.2.4 方法 1：投与群指示変数の並べ替え

解析対象となるデータにランダム化の手順が入っていれば、帰無仮説のもとで結果変数と投与群指示変数は独立とみなすことができる。この考え方の起源は Fisher まで遡る (Fisher 1935)。Fisher のランダム化の考え方から、投与群指示変数を並び替える検定が妥当性となる。すなわ

ち、結果変数と投与群指示変数との関係を排除した帰無仮説下の検定統計量が導出可能となる。投与群指示変数を並び替えてデータセットを再構成し、 $\beta_1$ の検定に関する検定統計量を並び替えたデータセットごとに計算する。等確率で治療が割り付けられるという原則を利用したランダム化を繰り返して適用し、帰無仮説下の検定統計量が従う分布を並べ替えにより推定できる。投与群指示変数を並び替える操作は、帰無仮説（治療効果なし、 $H_0: \beta_1 = \beta_1^{null}$ ）下での検定統計量 $T(\beta_1^{null})$ を取得していることと等価な操作である。得られた検定統計量が経験分布に従うことを利用して、実際に観察されたデータから計算した検定統計量の検定を行う。ランダム化という試験デザインが並べ替えを正当化しており、追加の仮定は不要である。実際には、投与群指示変数を並べ替えて全ての組み合わせの統計量を計算するには、組み合わせ数が膨大となり現実的に計算できないケースが多い。そのため、代わりに並べ替えを十分な回数繰り返し、一定の精度を保持した経験分布を推定する。この手順を Monte Carlo 法とも呼ぶ (Dwass 1957; Marozzi 2004; Pesarin and Salmaso 2010)。したがって、いわゆる exact な方法ではなく exact な方法に近似した手法との位置づけとなる。

#### アルゴリズム 1（投与群指示変数の並べ替えによる方法）

**(1)（観測されたデータに対する基本統計量 $T(\beta_1^{null})$ の計算）**

観測されたデータに対して、4.2.1 の手順にしたがって、基本統計量 $T(\beta_1^{null})$ を計算する。

**(2)（投与群指示変数の並べ替え）**

$N$ 人の比較群を示す指示変数を $\{N_0, N_1\}$ の組み合わせに並び替えを行う。ここでは、その数を $B$ 通りとする。すべての組み合わせを取り尽くすことが可能である場合には、 $B = {}_N C_{N_0}$ とし、そうでない場合には、この中から十分大きな数の組み合わせをランダムに選択する（本研究では、 $B = 3000$ ）。

並べ替え後の $B$ 組のデータセットに対して、(1)と同じ手順で、基本統計量 $T(\beta_1^{null})$ を計算する( $b = 1, 2, \dots, B$ )。

**(3)（ $p$ 値の計算）**

(2)で得られた統計量 $T^{(1)}(\beta_1^{null}), T^{(2)}(\beta_1^{null}), \dots, T^{(B)}(\beta_1^{null})$ の分布を参照分布に用いて、以下の式により $p$ 値を計算 (Davison and Hinkley 1997)

$$p = \frac{1}{B+1} \left\{ 1 + \sum_{b=1}^B I[|T^{(b)}(\beta_1^{null})| > |T(\beta_1^{null})|] \right\}$$

ここで、 $I(x)$ は指示関数であり、 $x$ が真のとき 1、 $x$ が偽のとき 0 を返す関数である。

ここで得られる並べ替え $p$ 値は、 $\beta_1$ 以外の局外パラメータを含むモデルに基づくものであることに注意が必要である。回帰モデルなどの局外パラメータを含むモデルにおける検定では、一般的に並べ替え $p$ 値は正確な $p$ 値とはならない。その厳密な正当性は、大標本理論によって説明することが可能である。しかしながら、このモデルに基づく近似的な $p$ 値を用いた検定も、経験的に小標本下で良好な性能を有することが知られている (Anderson and Robinson 2001)。

信頼区間については、 $t$ 検定統計量の経験分布から $\alpha/2\%$ 及び $(1 - \alpha/2)\%$ 点を推定し（それぞれ、 $\hat{q}_{t,(\alpha/2)}$ 及び $\hat{q}_{t,(1-\alpha/2)}$ ）、以下の不等式を満たす $\beta_1^{null}$ の集合を求めることにより、 $100(1 -$

$\alpha$ )%信頼区間を構成することもできる。

$$\hat{q}_{t,(\alpha/2)} \leq T(\beta_1^{null}) \leq \hat{q}_{t,(1-\alpha/2)}$$

#### 4.2.5 方法 2：重み付き残差の並べ替え

Freedman 及び Lane は線形回帰モデルの残差に基づく並べ替え検定を提案した (Freedman and Lane 1983)。線形混合モデルの式 ((1), 8 ページ) に従いデータが発生していると仮定すると、残差  $\varepsilon_i$  が結果変数  $Y_i$  と  $X_i\beta + Z_i b_i$  が張る平面 (線形射影) の差と定義できる。この枠組みでは、残差はいわゆる独立かつ同一な分布を持つランダムな誤差とみなすことはできない。Kennedy (1995) は Freedman 及び Lane の提案法 (Freedman and Lane 1983) と類似した別法を提案した。Kennedy の方法 (Kennedy 1995) では第 1 種の過誤確率の上昇が観測された。Anderson 及び Robinson は、データを並べ替える際、適切に共変量を取り扱わないことで第 1 種の過誤確率の上昇が発生していると指摘した (Anderson and Robinson 2001)。

本研究では、MMRM を適用しており変量効果がモデリングに含まれている。変量効果のような局外パラメータがモデルに含まれる場合の単純な残差は独立かつ同一の分布ではない。MMRM のような線形混合効果モデルにおける残差の並べ替え検定を考える際、Lee and Braun (2012) は残差統計量に重みを付与することにより多変量正規分布に従うことを示した。本研究では、Lee and Braun (2012) の方法を用いて、投与群指示変数をモデル式から除き帰無仮説のもとで推定した残差を並べ替える方法を考えた。

式 (2) (8 ページ) における帰無仮説のもとでの固定効果パラメータ・周辺共分散行列の制約付きパラメータ推定値を  $\{\tilde{\beta}, \tilde{V}_i\}$  とする ( $i = 1, 2, \dots, N$ )。また、 $\tilde{U}_i$  を  $\tilde{V}_i$  をコレスキー分解した上三角行列とする。このとき、この  $\tilde{U}_i$  を用いた、以下の重み付き残差統計量は多変量標準正規分布に従う。

$$(\tilde{U}_i^T)^{-1}(Y_i - X_i\tilde{\beta}) \sim MVN(0, E_{n_i})$$

$E_{n_i}$  は  $n_i$  次の単位行列である。帰無仮説のもとで推定した残差の並べ替えは、モデルによる並べ替え検定とも呼ばれ、帰無仮説下 ( $\beta_1 = \beta_1^{null}$ ) における残差を推定し、並べ替えに用いる (Freedman and Lane 1983)。興味ある投与群指示変数以外の共変量変数はモデルに残っており、これら共変量と結果変数  $Y$  の関係は保存されている点が特徴となる。検定の構成手順は以下の通りである。

#### アルゴリズム 2 (重み付き残差の並べ替えによる方法)

##### (1) (観測されたデータに対する基本統計量 $T(\beta_1^{null})$ の計算)

観測されたデータに対して、4.2.1 の手順にしたがって、基本統計量  $T(\beta_1^{null})$  を計算する。

##### (2) (重み付き残差の並べ替え)

観測されたデータに対して、帰無仮説のもとでの制約付きモデルのもとでの結果変数の予測値  $\tilde{Y}_i = X_i\tilde{\beta}$ 、及び重み付き残差統計量  $e_i = (\tilde{U}_i^T)^{-1}(Y_i - X_i\tilde{\beta})$  を計算する。重み付き残差  $e_1, e_2, \dots, e_N$  をランダムに並び替え、 $B$  通りの並び替え重み付き残差統

計量 $\mathbf{e}_1^{(b)}, \mathbf{e}_2^{(b)}, \dots, \mathbf{e}_N^{(b)}$  ( $b = 1, 2, \dots, B$ )を得る。

並び替えを行った重み付き残差統計量を、結果変数の予測値 $\tilde{\mathbf{Y}}_i$ に加えて、並び替え残差を付与した予測値 $\mathbf{Z}_i^{(b)} = \tilde{\mathbf{Y}}_i + \mathbf{e}_i^{(b)}$ を得る。こうして得られる $\mathbf{Z}_1^{(b)}, \mathbf{Z}_2^{(b)}, \dots, \mathbf{Z}_N^{(b)}$ を結果変数として、(1)と同じ回帰関数モデルと共分散構造を仮定した MMRM に対して、基本統計量 $T^{(b)}(\beta_1^{null})$ を計算する( $b = 1, 2, \dots, B$ )。

### (3) (p値の計算)

(2)で得られた統計量 $T^{(1)}(\beta_1^{null}), T^{(2)}(\beta_1^{null}), \dots, T^{(B)}(\beta_1^{null})$ の分布を参照分布に用いて、以下の式によりp値を計算 (Davison and Hinkley 1997)

$$p = \frac{1}{B+1} \left\{ 1 + \sum_{b=1}^B I[|T^{(b)}(\beta_1^{null})| > |T(\beta_1^{null})|] \right\}$$

ここで、 $I(x)$ は指示関数であり、 $x$ が真のとき 1、 $x$ が偽のとき 0 を返す関数である。

信頼区間については、 $t$ 検定統計量の経験分布から $\alpha/2\%$ 及び $(1-\alpha/2)\%$ 点を推定し（それぞれ、 $\hat{q}_{t,(\alpha/2)}$ 及び $\hat{q}_{t,(1-\alpha/2)}$ ）、以下の不等式を満たす $\beta_1^{null}$ の集合を求めることにより、 $100(1-\alpha)\%$ 信頼区間を構成することもできる。

$$\hat{q}_{t,(\alpha/2)} \leq T(\beta_1^{null}) \leq \hat{q}_{t,(1-\alpha/2)}$$

## 4.3 シミュレーション実験

### 4.3.1 シナリオ

尤度比検定における Bartlett 補正法とブートストラップ検定に基づく近似推測手法と同様であるため、記載を割愛する (3.5.1 を参照)。

### 4.3.2 相関構造

尤度比検定における Bartlett 補正法とブートストラップ検定に基づく近似推測手法と同様であるため、記載を割愛する (3.5.2 を参照)。

### 4.3.3 欠測メカニズム

尤度比検定における Bartlett 補正法とブートストラップ検定に基づく近似推測手法と同様であるため、記載を割愛する (3.5.3 を参照)。

### 4.3.4 評価方法

本研究で提案する 2 種類の方法を含めて合計 5 種類の方法を評価した。比較対象とした方法の略語を説明する。PermTrt 及び PermWR は、それぞれ投与群指示変数を並べ替える方法（アルゴリズム 1）及び帰無仮説のもとで推定した残差に重みを付与した残差を並べ替える方法（アルゴリズム 2）である。

並べ替えたデータセットに対して MMRM により治療効果の回帰係数に関する $t$ 検定統計量を求める際、自由度調整法を Between-within 法 (Schluchter and Elashoff 1990) とした REML 推定を適用している。MMRM<sub>BW-REML</sub> は提案法の直接的な比較対象である。また、推定方法を最尤法とした MMRM<sub>BW-ML</sub> についても検討した。

本研究では、小標本下での漸近近似を改善した方法として提案されている Kenward-Roger 法



(Kenward and Roger 1997) を適用した MMRM (MMRM<sub>KR</sub>) を比較対象として設定した。以下の表 4.3-1 に略語と方法の対応表を示す。

表 4.3-1 略語と方法の対応表

略語	方法
PermTrt	投与群指示変数を並べ替えたデータセットを MMRM <sub>BW-REML</sub> により解析する方法 (アルゴリズム 1 に対応)
PermWR	帰無仮説のもとで推定した重み付き残差を並べ替えたデータセットを MMRM <sub>BW-REML</sub> により解析する方法 (アルゴリズム 2 に対応)
MMRM <sub>KR</sub>	固定効果の標準誤差及び自由度調整法に Kenward and Roger 法を適用した MMRM
MMRM <sub>BW-REML</sub>	固定効果の推定に REML 及び自由度調整に Between-within 法を適用した MMRM
MMRM <sub>BW-ML</sub>	固定効果の推定に ML 及び自由度調整に Between-within 法を適用した MMRM

MMRM では、投与群及び時点を因子 (クラス変数) として扱い、投与群と時点の交互作用項を含めたモデルにより解析を実施した。分散共分散構造は、無構造 (Unstructured) を指定した。5 種類の比較方法ごとに最終測定時点における群間差に関する検定を実施し  $p$  値を計算した。MMRM<sub>KR</sub>, MMRM<sub>BW-REML</sub> 及び MMRM<sub>BW-ML</sub> については、固定効果の標準誤差が 50 を超えた場合に推定が不適切であると判断し評価から除外する処理を行った。反復回数を分母としたときに  $p < 0.05$  を満たす  $p$  値の割合を、Type I error rate として計算した。5 種類の方法について収束割合を算出した。本研究におけるシミュレーションプログラムは SAS Ver 9.4 (SAS Institute, Cary, USA) で作成しており、収束判定の基準は SAS GLIMMIX プロシーチャーのデフォルト設定を適用した。

観察されたデータに対して治療効果の回帰係数に関する  $t$  検定統計量を求めようとするときに、脱落割合によっては推定ができない場合がある。このような場合は、評価から除外した。また、並べ替えたデータセットを MMRM で解析した結果、帰無仮説下の  $t$  検定統計量が算出できない場合もある。こうした場合を除外して合計 3000 回のリサンプリングがされるように調整した。尤度比検定における Bartlett 補正項を推定する方法と同じ条件下で評価可能なようにリサンプリング回数を 3000 回とした。

MMRM に並べ替え検定を適用した方法については、シナリオ 1, 2, 5 及び 6 についてシミュレーション実験の結果を掲載する。その他のシナリオについては、時間の制約上検討できていない。また反復回数は 2000 回とした。

#### 4.3.5 従来法に対する提案法の基礎的挙動

従来法に対する提案法の基礎的挙動を確認するために、非欠測下の帰無仮説を仮定し従来法と位置付けている MMRM<sub>BW-REML</sub> と提案法 (PermTrt 及び PermWR) について QQ プロットを作成した。

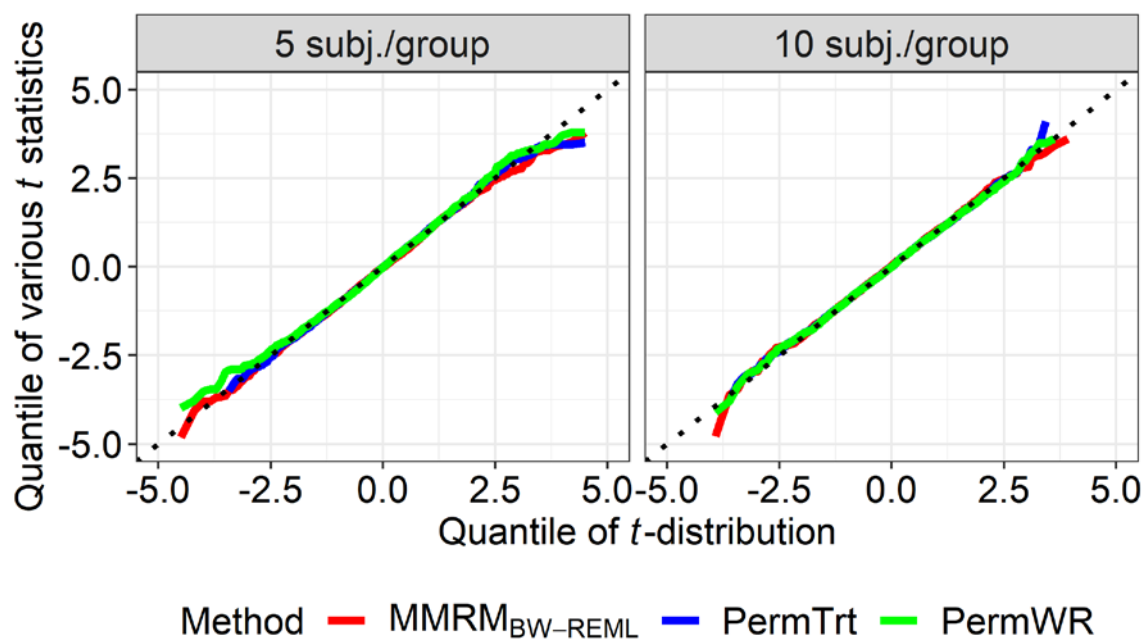


図 4.3-1 シナリオ 1 の非欠測下における QQ プロット

横軸： $t$ 分布の分位点（1 群 5 例及び 10 例の場合に、 $t$ 分布の自由度はそれぞれ 8 及び 18）、縦軸：MMRM<sub>BW-REML</sub>，固定効果の推定に REML 法及び自由度調整を Between-within 法とした MMRM；PermTrt，投与群指示変数を並べ替える方法；PermWR，重み付き残差統計量を並べ替える方法

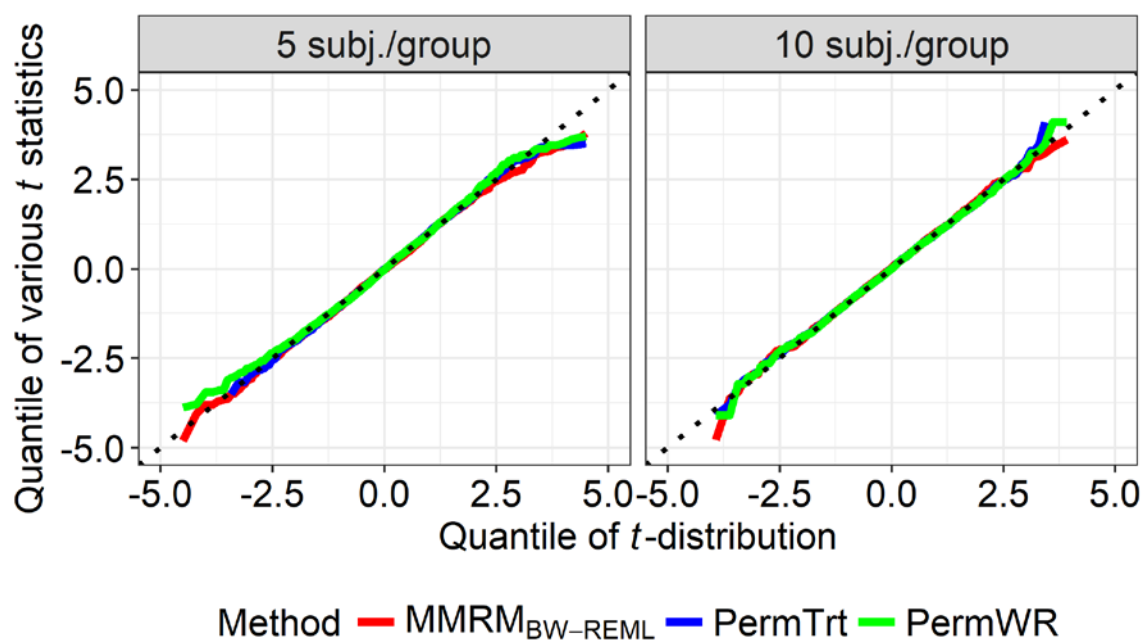


図 4.3-2 シナリオ 2 の非欠測下における QQ プロット

横軸： $t$ 分布の分位点（1 群 5 例及び 10 例の場合に、 $t$ 分布の自由度はそれぞれ 8 及び 18）、縦軸：MMRM<sub>BW-REML</sub>，固定効果の推定に REML 法及び自由度調整を Between-within 法とした MMRM；PermTrt，投与群指示変数を並べ替える方法；PermWR，重み付き残差統計量を並べ替える方法

表 4.3-2 帰無仮説下における $t$ 分布の上側 2.5%点

シナリオ	1 群例数	MMRM <sub>BW-REML</sub>	PermTrt	PermWR
1	5	2.240	2.471	2.482
1	10	2.140	2.037	2.037
2	5	2.240	2.471	2.418
2	10	2.140	2.037	2.043

1 群 5 例及び 10 例の場合の $t$ 分布の自由度は、それぞれ 8 及び 18。自由度 8 及び 18 における $t$ 分布の理論上の上側 2.5%点は、それぞれ 2.306 及び 2.101。

表 4.3-3 欠測がない状況での Type I error rate (%)

シナリオ	1 群例数	MMRM <sub>BW-REML</sub>	PermTrt	PermWR
1	5	4.85	5.25	4.95
1	10	5.05	4.95	5.10
2	5	4.85	5.25	4.80
2	10	5.05	4.95	5.00

欠測がない状況で帰無仮説を仮定したシナリオ 1 で推定する $t$ 検定統計量 $T(\beta_1^{null})$ は、理論的には 1 群 5 例及び 10 例の場合にそれぞれ自由度 8 及び 18 の $t$ 分布に従う。図 4.3-1 に従来法 MMRM<sub>BW-REML</sub> と提案法 PermTrt 及び PermWR の QQ プロットを示した。同様にシナリオ 2 における QQ プロットを図 4.3-2 に示した。

視覚的に確認すると、いずれの方法であってもほぼ理論上の $t$ 分布に従ったプロットとなっていた。

図 4.3-1 及び図 4.3-2 に示した QQ プロットで $t$ 分布の上側 2.5%点を表 4.3-2 に示した。 $t$ 分布の理論上の上側 2.5%点と各方法で得られた 2.5%点を比較すると、ほぼ理論値に一致した値であった。

表 4.3-3 に各方法での Type I error rate を示した。シミュレーション実験の反復回数に対するモンテカルロエラーを考慮すると、Type I error rate は 5%に維持されていると判断できた。

#### 4.3.6 $N = 10$ (1 群 5 例)

$N = 10$ の検討で、定義した欠測メカニズムにより欠測を発生させたデータを生成し、比較方法を適用した。すべての比較方法は制限付最尤法もしくは最尤法である。したがって、小標本かつ欠測が発生する $N = 10$ の状況ではパラメータの収束がほとんど得られなかった。そこで、 $N = 10$ のセクションでは、欠測がない完全データのための検討結果を示す。

##### 4.3.6.1 Type I error rate

$N = 10$  (1 群 5 例) における Type I error rate の結果を、表 4.3-4 に示す。シナリオ 1 ではすべての時点で真の平均構造は群間に差がない。シナリオ 2 は最終時点の平均構造にのみ群間に差がある。したがって、Type I error rate は 5%となるはずである。最尤法を用いている MMRM<sub>BW-ML</sub> の Type I error rate は 7.5%であり、5%から大きくインフレーションする結果であった。

並び替え検定と MMRM を組み合わせた提案法の中では、シナリオ 1 で重み付き残差の並べ替え検定である PermWR は 4.95%であり最も 5%に近い Type I error rate であった。また、同様にシナリオ 1 で投与群指示変数の並べ替え検定である PermTrt は 5.25%であった。

表 4.3-4 Type I error rate (%)のシミュレーション実験結果<sup>†</sup> (シナリオ 1 及び 2, 欠測なし, 1 群 5 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	PermTrt	5.25	---	---	5.25	---	---
	PermWR	4.95	---	---	4.95	---	---
	MMRM <sub>KR</sub>	4.87	---	---	4.87	---	---
	MMRM <sub>BW-REML</sub>	4.85	---	---	4.85	---	---
	MMRM <sub>BW-ML</sub>	7.50	---	---	7.50	---	---
2	PermTrt	5.25	---	---	5.25	---	---
	PermWR	4.80	---	---	4.80	---	---
	MMRM <sub>KR</sub>	4.93	---	---	4.93	---	---
	MMRM <sub>BW-REML</sub>	4.85	---	---	4.85	---	---
	MMRM <sub>BW-ML</sub>	7.50	---	---	7.50	---	---

<sup>†</sup>PermTrt, MMRM<sub>BW-REML</sub> を用いた投与群指示変数による並べ替え検定；PermWR, 帰無仮説のもとで推定した重み付き残差を並べ替え MMRM<sub>BW-REML</sub> で検定；MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定；MMRM<sub>BW-REML</sub>, Between-within 法による REML 推定を用いた Wald 型検定；MMRM<sub>BW-ML</sub>, Between-within 法による最尤推定を用いた Wald 型検定

#### 4.3.6.2 検出力

$N = 10$  (1 群 5 例) における検出力の結果を, 表 4.3-5 に示す。シナリオ 5 及び 6 において最も高い検出力を示した方法は MMRM<sub>BW-ML</sub> であり, それぞれ 57.93% 及び 57.90% であった。MMRM<sub>BW-ML</sub> 以外の方法でも約 50% 以上の検出力を示した。

表 4.3-5 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 5 及び 6, 欠測なし, 1 群 5 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
5	PermTrt	50.25	---	---	50.25	---	---
	PermWR	49.00	---	---	49.00	---	---
	MMRM <sub>KR</sub>	49.37	---	---	49.37	---	---
	MMRM <sub>BW-REML</sub>	49.65	---	---	49.65	---	---
	MMRM <sub>BW-ML</sub>	57.93	---	---	57.93	---	---
6	PermTrt	50.35	---	---	50.35	---	---
	PermWR	48.80	---	---	48.80	---	---
	MMRM <sub>KR</sub>	49.55	---	---	49.55	---	---
	MMRM <sub>BW-REML</sub>	49.65	---	---	49.65	---	---
	MMRM <sub>BW-ML</sub>	57.90	---	---	57.90	---	---

<sup>†</sup>PermTrt, MMRM<sub>BW-REML</sub> を用いた投与群指示変数による並べ替え検定；PermWR, 帰無仮説のもとで推定した重み付き残差を並べ替え MMRM<sub>BW-REML</sub> で検定；MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定；MMRM<sub>BW-REML</sub>, Between-within 法による REML 推定を用いた Wald 型検定；MMRM<sub>BW-ML</sub>, Between-within 法による最尤推定を用いた Wald 型検定

#### 4.3.6.3 収束割合

$N = 10$  (1 群 5 例) における収束割合の結果を, 表 4.3-6 に示す。 $N = 10$  (1 群 5 例) は欠測がない完全データでの検討であるため, パラメータの収束値が得られる。したがって, 収束割合はいずれの方法においても 100% であった。

表 4.3-6 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 1, 2, 5 及び 6, 欠測なし, 1 群 5 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	PermTrt	100.0	---	---	100.0	---	---
	PermWR	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW-REML</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW-ML</sub>	100.0	---	---	100.0	---	---
2	PermTrt	100.0	---	---	100.0	---	---
	PermWR	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW-REML</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW-ML</sub>	100.0	---	---	100.0	---	---
5	PermTrt	100.0	---	---	100.0	---	---
	PermWR	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW-REML</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW-ML</sub>	100.0	---	---	100.0	---	---
6	PermTrt	100.0	---	---	100.0	---	---
	PermWR	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW-REML</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW-ML</sub>	100.0	---	---	100.0	---	---

<sup>†</sup>PermTrt, MMRM<sub>BW-REML</sub> を用いた投与群指示変数による並べ替え検定 ; PermWR, 帰無仮説のもとで推定した重み付き残差を並べ替え MMRM<sub>BW-REML</sub> で検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW-REML</sub>, Between-within 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW-ML</sub>, Between-within 法による最尤推定を用いた Wald 型検定

#### 4.3.7 $N = 20$ (1 群 10 例)

##### 4.3.7.1 Type I error rate

$N = 20$  (1 群 10 例) における Type I error rate の結果を表 4.3-7 及び図 4.3-3 に示す。シナリオ 1 及び 2 の最終測定時点における群間差は 0 であるから, Type I error rate は 5% に維持されるはずである。MMRM<sub>BW-REML</sub> 及び MMRM<sub>BW-ML</sub> では, 欠測割合が増加すると MCAR 及び MAR の両方で Type I error rate が増加した。MMRM<sub>BW-REML</sub> ではシナリオ 1 の欠測割合 40% のとき Type I error rate は, MCAR 及び MAR でそれぞれ 9.77% 及び 8.62% であった。

表 4.3-7 Type I error rate (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 10 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	PermTrt	4.95	4.45	4.89	4.95	4.50	5.17
	PermWR	5.10	4.20	4.08	5.10	3.85	4.15
	MMRM <sub>KR</sub>	5.05	4.75	5.75	5.05	3.90	4.14
	MMRM <sub>BW-REML</sub>	5.05	5.95	9.77	5.05	5.51	8.62
	MMRM <sub>BW-ML</sub>	6.10	7.80	12.11	6.10	7.00	10.64
2	PermTrt	4.95	4.30	5.56	4.95	4.56	5.42
	PermWR	5.00	3.95	4.30	5.00	4.01	4.24
	MMRM <sub>KR</sub>	5.05	4.40	6.11	5.05	4.11	4.29
	MMRM <sub>BW-REML</sub>	5.05	5.75	9.70	5.05	5.61	9.52

MMRM <sub>BW-ML</sub>	6.10	7.70	11.72	6.10	6.91	11.08
-----------------------	------	------	-------	------	------	-------

<sup>†</sup>PermTrt, MMRM<sub>BW-REML</sub>を用いた投与群指示変数による並べ替え検定；PermWR, 帰無仮説のもとで推定した重み付き残差を並べ替え MMRM<sub>BW-REML</sub> で検定；MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定；MMRM<sub>BW-REML</sub>, Between-within 法による REML 推定を用いた Wald 型検定；MMRM<sub>BW-ML</sub>, Between-within 法による最尤推定を用いた Wald 型検定

シミュレーションで得られた Type I error rate は、モンテカルロ誤差を考慮すると青点線の内側にあるとき Type I error rate が 5%に維持されていると考えられた。PermTrt は欠測メカニズムや欠測割合に関わらず、5%を維持していた。重み付き残差の並べ替えを利用した PermWR では、シナリオ 1 の MAR 下で欠測割合が 20%のとき、またシナリオ 2 の MCAR 下で欠測割合が 20%のとき、Type I error rate 5%を維持すると定義した基準を下回った。

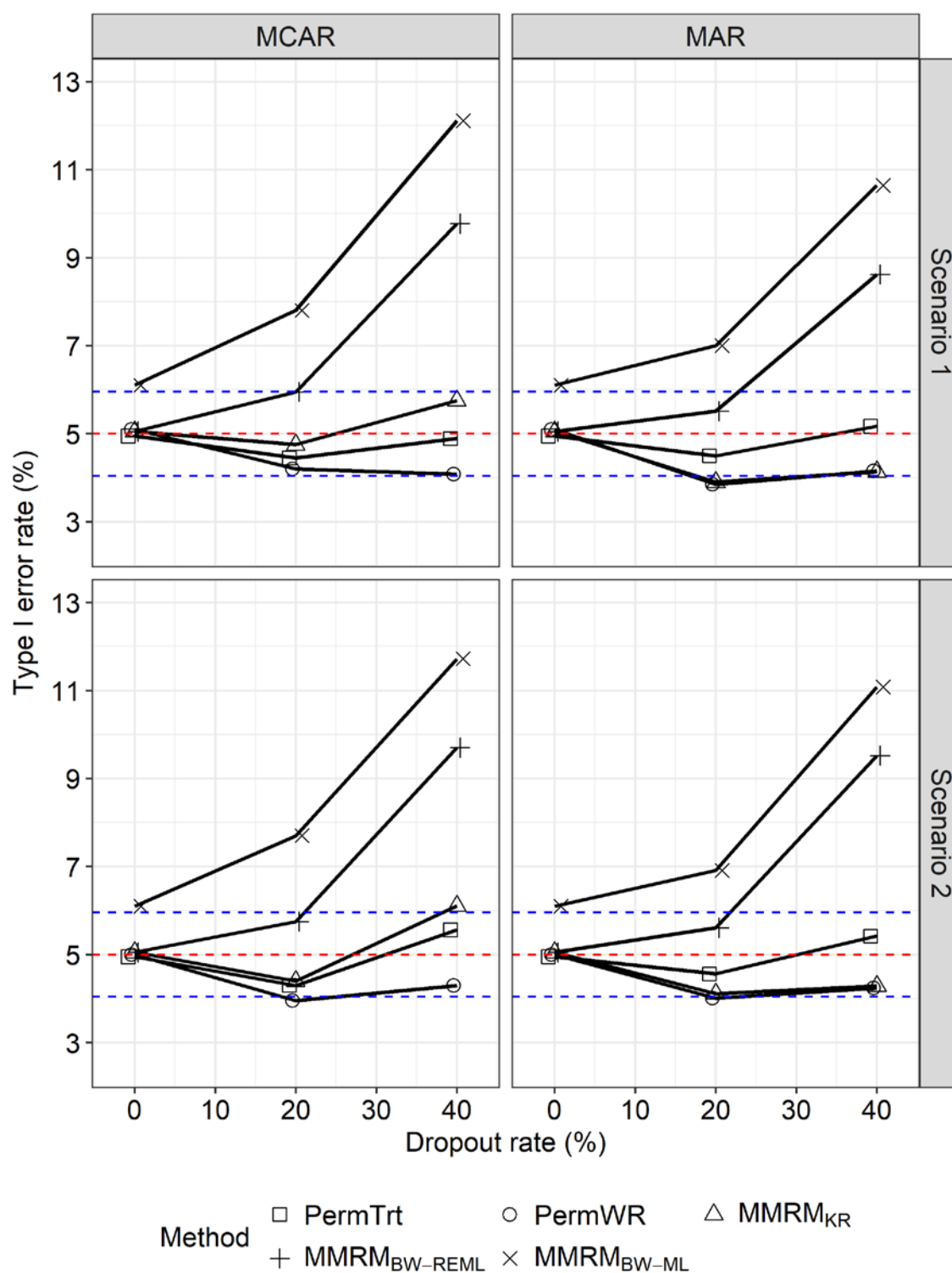


図 4.3-3 Type I error rate (%)のシミュレーション実験結果<sup>†</sup> (シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 10 例)

赤点線, 5%; 青点線, 95%信頼限界 (真の Type I error rate を 5% と仮定した場合の 2000 回反復によるモンテカルロ誤差); <sup>†</sup>PermTrt, MMRM<sub>BW-REML</sub> を用いた投与群指示変数による並べ替え検定; PermWR, 帰無仮説のもとで推定した重み付き残差を並べ替え MMRM<sub>BW-REML</sub> で検定; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定; MMRM<sub>BW-REML</sub>, Between-within 法による REML 推定を用いた Wald 型検定; MMRM<sub>BW-ML</sub>, Between-within 法による最尤推定を用いた Wald 型検定

#### 4.3.7.2 検出力

$N = 20$ （1 群 10 例）における検出力の結果を表 4.3-8 及び図 4.3-4 に示す。シナリオ 5 の最終測定時点の群間差は-9.6 であり、欠測がない下で  $\text{MMRM}_{\text{KR}}$ 、 $\text{MMRM}_{\text{BW-REML}}$  及び  $\text{MMRM}_{\text{BW-ML}}$  では検出力はそれぞれ 86.25%、86.25% 及び 88.45% であった。

図 4.3-4 において、 $\text{MMRM}_{\text{BW-ML}}$  の検出力が比較した方法の中で最も高く、欠測割合に応じて 88.5% から 70.0% の範囲に値をとった。しかしながら、 $\text{MMRM}_{\text{BW-REML}}$  及び  $\text{MMRM}_{\text{BW-ML}}$  は、シナリオ 1 及び 2 の下で、Type I error rate が 5% に制御されていなかったため、それを反映した結果であるものと考えられる。

シナリオ 5 において PermTrt の検出力は、欠測割合が 40% の MCAR 及び MAR 下でそれぞれ 59.54% 及び 59.39% であった。また、PermWR の検出力は、欠測割合が 40% の MCAR 及び MAR 下でそれぞれ 56.12% 及び 53.80% であった。

PermTrt のみが、欠測を伴う小標本の経時測定データに対して名義の有意水準を維持することができた。検出力は比較した方法の中で高くはないものの、Type I error rate を 5% に維持することを優先して考えれば、最も性能が良いと言える。

PermWR もシナリオ 1 の欠測割合 20% の MAR 下及びシナリオ 2 の欠測割合 20% の MCAR 下で名義の有意水準を維持することはできなかったが、PermTrt とほぼ同等な性能を有する可能性がある。

表 4.3-8 検出力（%）のシミュレーション実験結果<sup>†</sup>（シナリオ 5 及び 6，MCAR 及び MAR 下，1 群 10 例）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
5	PermTrt	86.20	75.35	59.54	86.20	76.83	59.39
	PermWR	86.50	73.60	56.12	86.50	74.57	53.80
	$\text{MMRM}_{\text{KR}}$	86.25	75.65	60.29	86.25	75.28	53.13
	$\text{MMRM}_{\text{BW-REML}}$	86.25	78.65	68.64	86.25	79.13	66.26
	$\text{MMRM}_{\text{BW-ML}}$	88.45	82.15	73.21	88.45	82.55	70.11
6	PermTrt	86.20	75.30	59.89	86.20	76.55	59.22
	PermWR	86.35	73.85	56.31	86.35	74.65	53.46
	$\text{MMRM}_{\text{KR}}$	86.25	75.65	60.44	86.25	75.15	52.25
	$\text{MMRM}_{\text{BW-REML}}$	86.25	78.65	68.84	86.25	78.75	65.91
	$\text{MMRM}_{\text{BW-ML}}$	88.45	82.15	73.31	88.45	82.55	69.98

<sup>†</sup>PermTrt,  $\text{MMRM}_{\text{BW-REML}}$  を用いた投与群指示変数による並べ替え検定；PermWR, 帰無仮説のもとで推定した重み付き残差を並べ替え  $\text{MMRM}_{\text{BW-REML}}$  で検定； $\text{MMRM}_{\text{KR}}$ , Kenward and Roger 法による REML 推定を用いた Wald 型検定； $\text{MMRM}_{\text{BW-REML}}$ , Between-within 法による REML 推定を用いた Wald 型検定； $\text{MMRM}_{\text{BW-ML}}$ , Between-within 法による最尤推定を用いた Wald 型検定



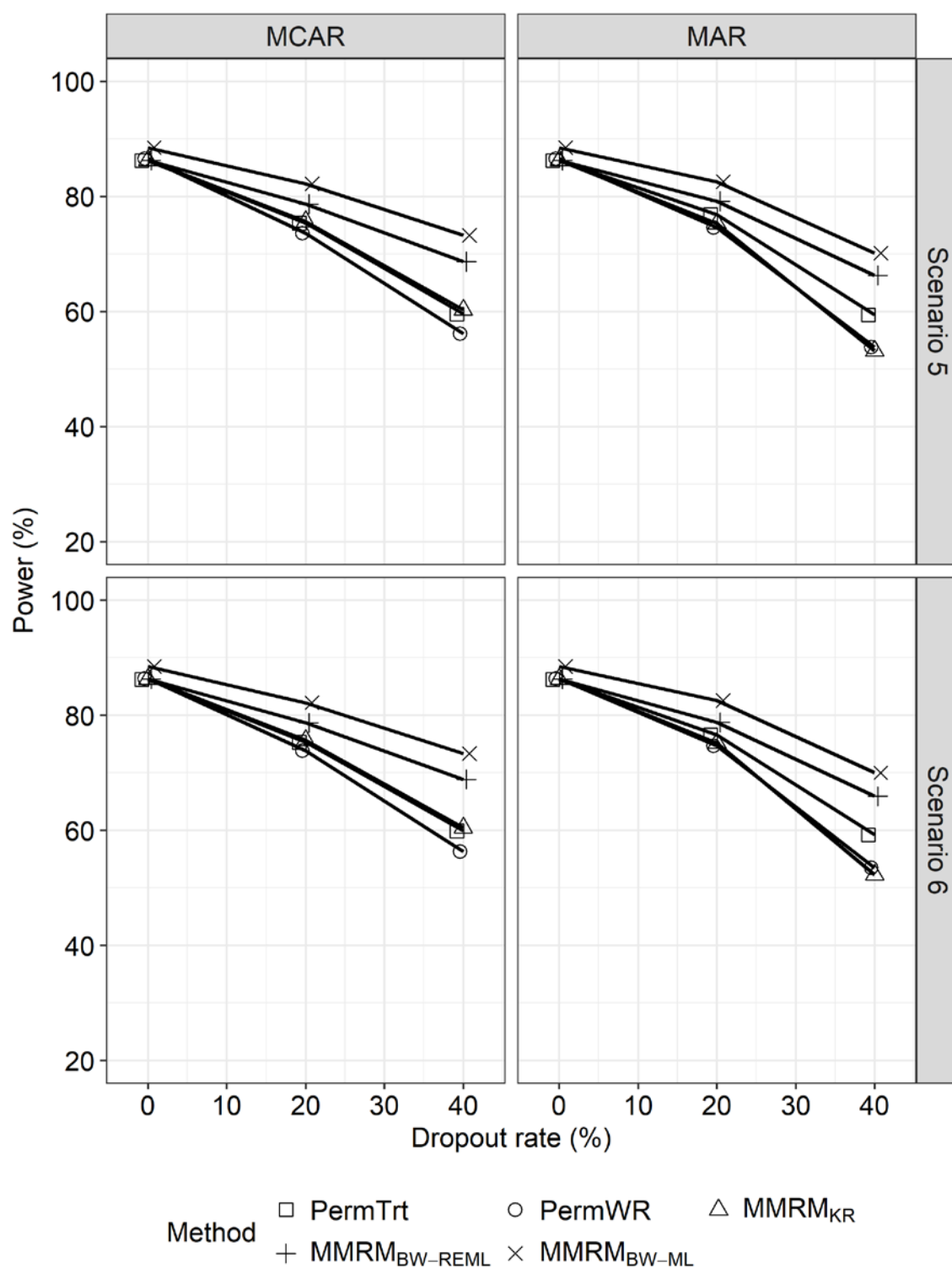


図 4.3-4 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 10 例)

<sup>†</sup>PermTrt, MMRM<sub>BW-REML</sub> を用いた投与群指示変数による並べ替え検定； PermWR, 帰無仮説のもとで推定した重み付き残差を並べ替え MMRM<sub>BW-REML</sub> で検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW-REML</sub>, Between-within 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW-ML</sub>, Between-within 法による最尤推定を用いた Wald 型検定

#### 4.3.7.3 収束割合

$N = 20$  (1 群 10 例) における収束割合の結果を表 4.3-9, 図 4.3-5 及び図 4.3-6 に示した。比較したすべての方法で欠測割合が 40% となった場合でも約 90% 前後の収束割合が得られた。また, 方法間での違いもほとんど認められなかった。

表 4.3-9 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 1, 2, 5 及び 6, MCAR 及び MAR 下, 1 群 10 例)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	PermTrt	100.0	100.0	92.1	100.0	99.9	88.1
	PermWR	100.0	100.0	92.0	100.0	99.9	87.9
	MMRM <sub>KR</sub>	100.0	100.0	92.1	100.0	99.9	88.2
	MMRM <sub>BW-REML</sub>	100.0	100.0	92.1	100.0	99.9	88.2
	MMRM <sub>BW-ML</sub>	100.0	100.0	91.7	100.0	100.0	87.9
2	PermTrt	100.0	100.0	95.4	100.0	99.9	84.9
	PermWR	100.0	100.0	95.4	100.0	99.9	84.9
	MMRM <sub>KR</sub>	100.0	100.0	95.4	100.0	99.9	85.1
	MMRM <sub>BW-REML</sub>	100.0	100.0	95.4	100.0	99.9	85.1
	MMRM <sub>BW-ML</sub>	100.0	100.0	95.1	100.0	99.9	84.8
5	PermTrt	100.0	100.0	92.0	100.0	99.9	94.0
	PermWR	100.0	100.0	92.0	100.0	99.9	93.5
	MMRM <sub>KR</sub>	100.0	100.0	92.2	100.0	99.9	94.0
	MMRM <sub>BW-REML</sub>	100.0	100.0	92.2	100.0	99.9	94.0
	MMRM <sub>BW-ML</sub>	100.0	100.0	91.7	100.0	100.0	93.9
6	PermTrt	100.0	100.0	92.0	100.0	100.0	94.1
	PermWR	100.0	100.0	91.6	100.0	100.0	93.9
	MMRM <sub>KR</sub>	100.0	100.0	92.1	100.0	100.0	94.2
	MMRM <sub>BW-REML</sub>	100.0	100.0	92.1	100.0	100.0	94.2
	MMRM <sub>BW-ML</sub>	100.0	100.0	91.8	100.0	100.0	94.0

<sup>†</sup>PermTrt, MMRM<sub>BW-REML</sub> を用いた投与群指示変数による並べ替え検定 ; PermWR, 帰無仮説のもとで推定した重み付き残差を並べ替え MMRM<sub>BW-REML</sub> で検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW-REML</sub>, Between-within 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW-ML</sub>, Between-within 法による最尤推定を用いた Wald 型検定

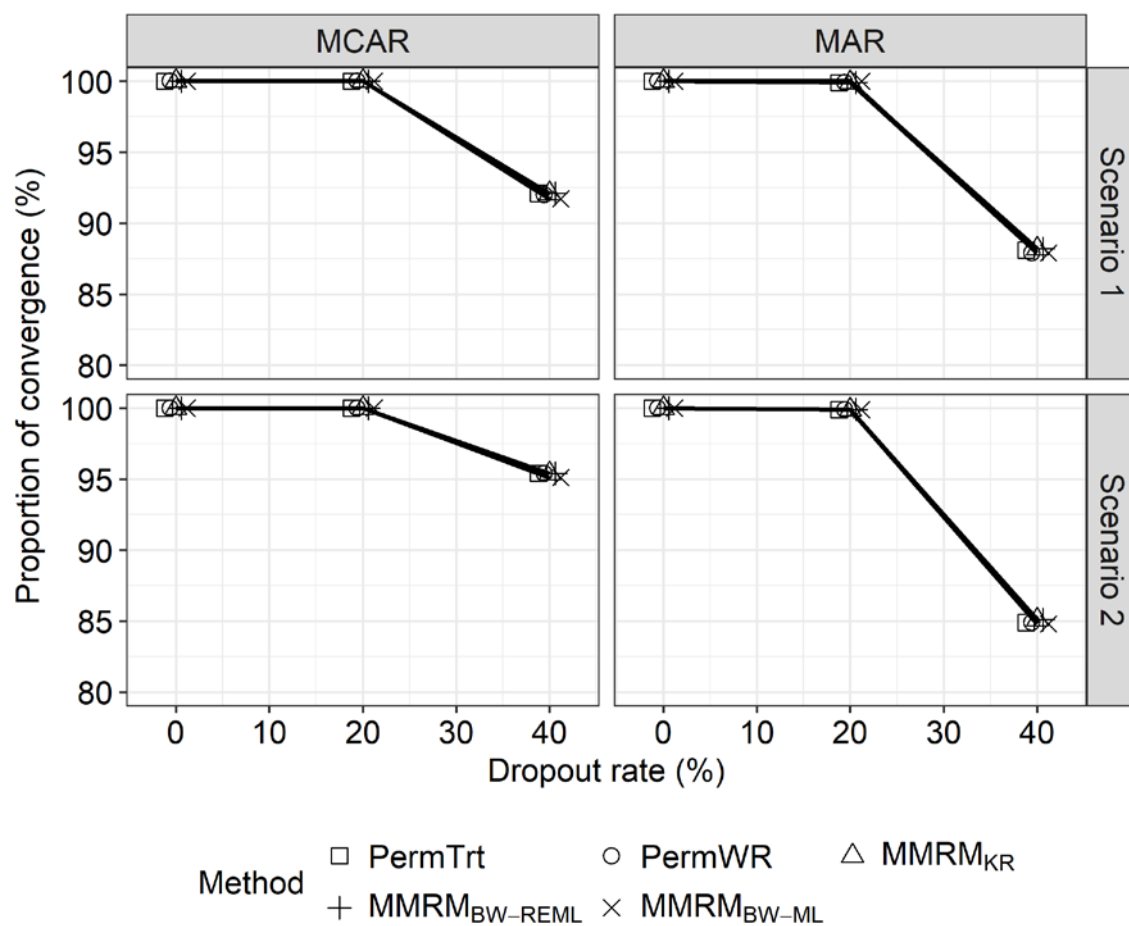


図 4.3-5 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 10 例)

<sup>†</sup>PermTrt, MMRM<sub>BW-REML</sub> を用いた投与群指示変数による並べ替え検定； PermWR, 帰無仮説のもとで推定した重み付き残差を並べ替え MMRM<sub>BW-REML</sub> で検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW-REML</sub>, Between-within 法による REML 推定を用いた Wald 型検； MMRM<sub>BW-ML</sub>, Between-within 法による最尤推定を用いた Wald 型検定

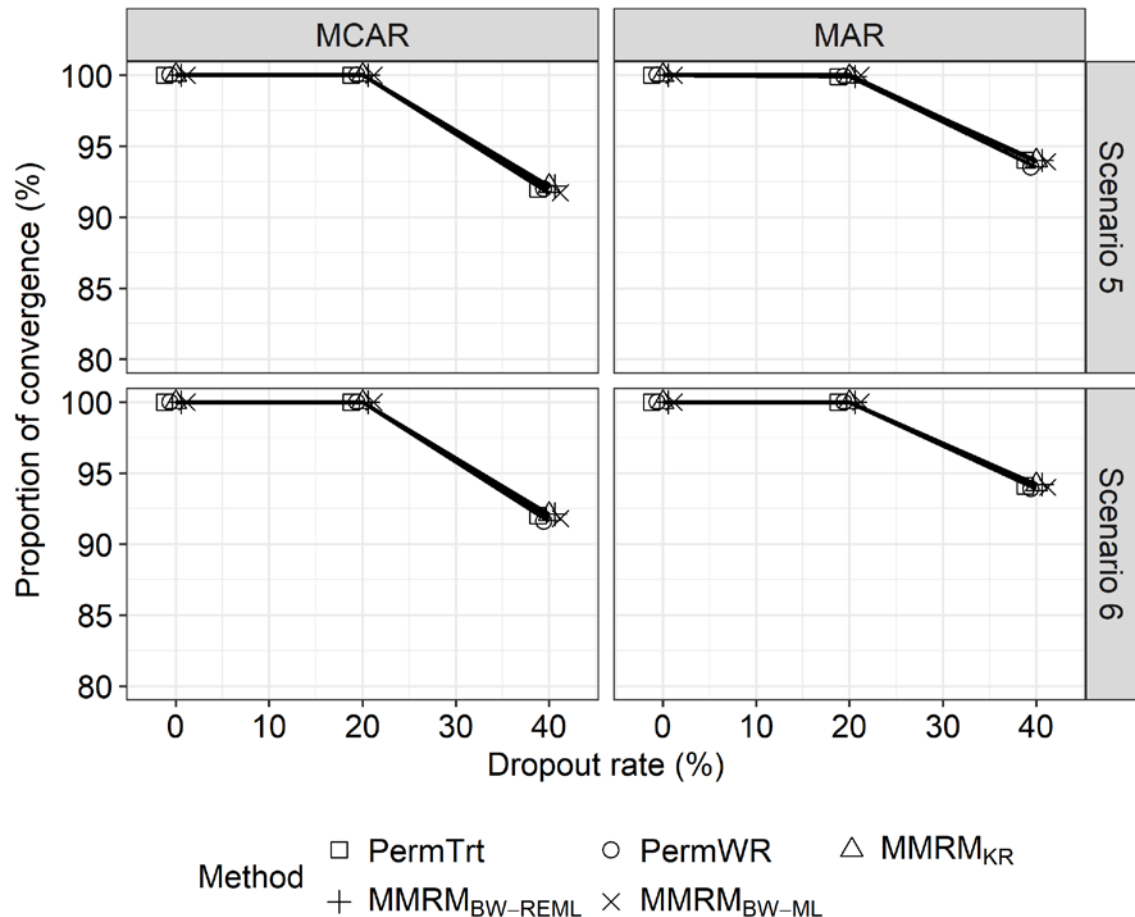


図 4.3-6 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 10 例)

<sup>†</sup>PermTrt, MMRM<sub>BW-REML</sub> を用いた投与群指示変数による並べ替え検定； PermWR, 帰無仮説のもとで推定した重み付き残差を並べ替え MMRM<sub>BW-REML</sub> で検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW-REML</sub>, Between-within 法による REML 推定を用いた Wald 型検； MMRM<sub>BW-ML</sub>, Between-within 法による最尤推定を用いた Wald 型検定

#### 4.3.8 $N = 40$ (1 群 20 例)

時間の制約があったため、本研究では検討していない。

### 4.4 実データへの適用

尤度比検定における Bartlett 補正項の近似推測手法で実データとして用いた産後うつ臨床試験データを事例データとして適用する。事例データの背景及び基本的な解析結果は、重複を避けるため記載を割愛する (3.7.1 を参照)。MMRM に並べ替え検定を適用した提案法とその比較方法に関する解析結果について以下に述べる。

提案法及び比較方法の解析では、被験者を変量効果とした混合効果モデルを用いた。

$$y_{ij} = \beta_0 + \beta_1 G_i + \beta_2 t_{ij} + \beta_3 G_i t_{ij} + b_{0i} + \varepsilon_{ij}$$

被験者は、 $i = 1, 2, \dots, 61$  であり、各被験者ごとに最終測定時点が異なっているため、 $j =$

1, 2, ...,  $n_i$ となる。また、被験者の変量効果を $b_{0i} \sim N(0, \sigma_b^2)$ とし、誤差を $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ とした。最終測定時点 $j = 8$ における $\beta_{3j}$ が主要評価項目である EPDS スコアの治療効果に相当しており、用いた方法ごとに $p$ 値を算出した。さらに、参考としてランダム化以降の測定時点についても、同様に解析結果を提示した（表 4.4-1）。

部分集団において、PermTrt 及び PermWR の $p$ 値は比較対象となる MMRM<sub>BW-REML</sub> よりも保守的な値となる傾向が確認できた。この傾向は、1 群 10 例のシミュレーション結果とも一貫している。また、最終測定時点（8 か月目）の部分集団解析結果で、MMRM<sub>BW-ML</sub> は比較対象とした方法の中で最も小さい $p$ 値であったが、この結果もシミュレーション結果と一貫していた。

表 4.4-1 産後うつ臨床試験における群間差及びその 95%信頼区間（測定時点ごと）

Item	全体集団		部分集団	
	Estimate [95%CI]	$p$ -value	Estimate [95%CI]	$p$ -value
<3 month>				
PermTrt	3.13 [0.36, 6.02]	0.0327	3.53 [-0.90, 7.96]	0.1193
PermWR	3.13 [0.42, 7.31]	0.0347	3.53 [-1.00, 7.90]	0.1153
MMRM <sub>KR</sub>	3.13 [0.33, 5.93]	0.0293	3.53 [-0.90, 7.97]	0.1138
MMRM <sub>BW-REML</sub>	3.13 [0.33, 5.93]	0.0293	3.53 [-0.90, 7.97]	0.1138
MMRM <sub>BW-ML</sub>	3.13 [0.37, 5.88]	0.0268	3.53 [-0.75, 7.82]	0.1022
<4 month>				
PermTrt	3.91 [0.29, 7.42]	0.0350	3.80 [-1.40, 9.26]	0.1546
PermWR	3.91 [0.42, 7.31]	0.0277	3.80 [-1.72, 9.18]	0.1523
MMRM <sub>KR</sub>	3.91 [0.43, 7.38]	0.0283	3.80 [-1.48, 9.08]	0.1509
MMRM <sub>BW-REML</sub>	3.91 [0.45, 7.36]	0.0275	3.80 [-1.43, 9.03]	0.1474
MMRM <sub>BW-ML</sub>	3.91 [0.51, 7.30]	0.0248	3.80 [-1.23, 8.83]	0.1328
<5 month>				
PermTrt	5.09 [2.16, 8.17]	0.0017	4.60 [-0.27, 9.49]	0.0680
PermWR	5.09 [1.97, 8.17]	0.0013	4.60 [-0.41, 9.61]	0.0666
MMRM <sub>KR</sub>	5.09 [2.04, 8.15]	0.0015	4.60 [-0.18, 9.38]	0.0584
MMRM <sub>BW-REML</sub>	5.09 [2.07, 8.12]	0.0013	4.60 [-0.09, 9.30]	0.0545
MMRM <sub>BW-ML</sub>	5.09 [2.13, 8.06]	0.0011	4.60 [0.10, 9.11]	0.0456
<6 month>				
PermTrt	3.28 [0.24, 6.31]	0.0370	3.27 [-1.32, 7.83]	0.1413
PermWR	3.28 [0.07, 6.47]	0.0440	3.27 [-1.67, 7.76]	0.1573
MMRM <sub>KR</sub>	3.28 [0.13, 6.42]	0.0414	3.27 [-1.17, 7.71]	0.1420
MMRM <sub>BW-REML</sub>	3.28 [0.18, 6.37]	0.0382	3.27 [-1.07, 7.61]	0.1337
MMRM <sub>BW-ML</sub>	3.28 [0.25, 6.31]	0.0344	3.27 [-0.89, 7.43]	0.1182
<7 month>				
PermTrt	3.99 [0.77, 7.20]	0.0167	3.97 [-1.30, 9.34]	0.1333
PermWR	3.99 [0.78, 7.19]	0.0157	3.97 [-1.54, 9.57]	0.1433
MMRM <sub>KR</sub>	3.99 [0.79, 7.20]	0.0158	3.97 [-1.25, 9.18]	0.1296
MMRM <sub>BW-REML</sub>	3.99 [0.84, 7.15]	0.0141	3.97 [-1.09, 9.02]	0.1191
MMRM <sub>BW-ML</sub>	3.99 [0.90, 7.09]	0.0123	3.97 [-0.86, 8.80]	0.1037
<8 month>				
PermTrt	4.34 [1.48, 7.08]	0.0037	3.93 [-1.26, 9.16]	0.1436
PermWR	4.34 [1.52, 7.15]	0.0033	3.93 [-1.37, 9.22]	0.1383
MMRM <sub>KR</sub>	4.34 [1.45, 7.23]	0.0040	3.93 [-1.23, 9.09]	0.1288
MMRM <sub>BW-REML</sub>	4.34 [1.51, 7.18]	0.0033	3.93 [-1.03, 8.90]	0.1160
MMRM <sub>BW-ML</sub>	4.34 [1.57, 7.12]	0.0027	3.93 [-0.81, 8.67]	0.1004

部分集団の定義：EPDS スコアのベースライン値>21 である被験者集団

†PermTrt, MMRM<sub>BW-REML</sub> を用いた投与群指示変数による並べ替え検定；PermWR, 帰無仮説のもとで推定した重み付き残差を並べ替え MMRM<sub>BW-REML</sub> で検定；MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定；MMRM<sub>BW-REML</sub>, Between-within 法による REML 推定を用いた Wald 型検定；MMRM<sub>BW-ML</sub>, Between-within 法による最尤推定を用いた Wald 型検定

図 4.4-1 に最終測定時点である 8 か月目において、並べ替え検定を MMRM に適用した手法によりリサンプリングした  $t$  検定統計量の経験分布に関するヒストグラムを示した。リサンプリング回数は 3000 回であり、上側が全体集団、下側が部分集団のヒストグラムとなっている。それぞれのヒストグラムの垂直な赤点線は観測データの  $t$  検定統計量の値を示しており、全体集団及び部分集団でそれぞれ 3.06 及び 1.62 であった。計算された  $p$  値は、図の垂直な赤点線よりも右側に位置するリサンプリング回数の割合となっている。

$p$  値の大きさについてさらに検討するために、垂直な赤点線よりも右側部分の状況を詳細に検討することとした。理論的には、経験分布の平均はゼロである。また、本事例で  $t$  検定統計量が従う  $t$  分布の自由度は、全体集団及び部分集団でそれぞれ 59 及び 28 である。そこで、部分集団のデータを対象にリサンプリングした  $t$  検定統計量から推定したカーネル密度関数と理論上の  $t$  分布を比較した。図 4.4-2 に比較した結果を示した。赤実線が推定したカーネル密度関数であり、青点線が理論上の  $t$  分布の密度関数である。

図 4.4-3 は、図 4.4-2 の横軸 0 から 2 の部分のみを拡大して描いた図である。赤実線が推定したカーネル密度関数であり、青点線が理論上の  $t$  分布の密度関数である。分布の裾の部分が PermTrt では、わずかに赤実線が青点線よりも大きく（推定したカーネル密度関数の右裾が理論上の  $t$  分布の右裾よりも重い）なっている。PermWR では、ほぼ赤実線と青点線は同一である。この違いが、PermWR よりも PermTrt の  $p$  値が大きい（保守的である）原因であると考えられた。

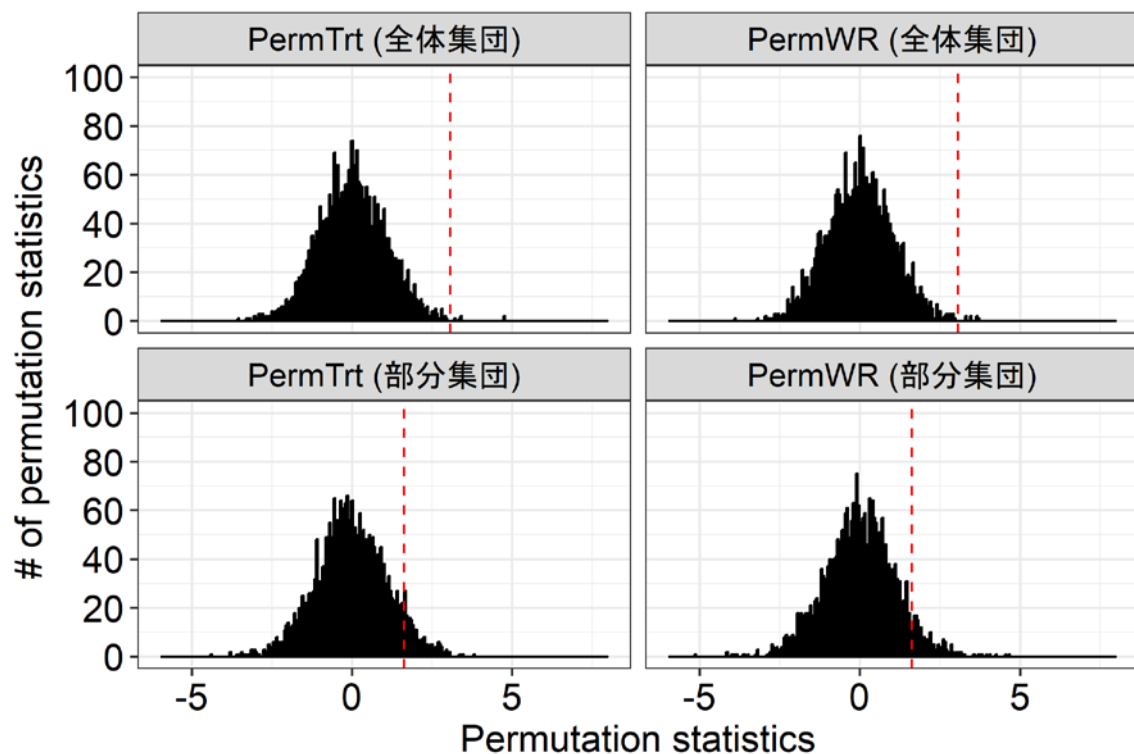


図 4.4-1 産後うつ臨床試験における並べ替え検定による推測手法における  $t$  検定統計量の経験分布<sup>†</sup> (8 ヶ月時点)

上側ボックスの赤点線, 3.06 (全体集団を対象とした  $t$  検定統計量) ; 下側ボックスの赤点線, 1.62 (部分集団を対象とした  $t$  検定統計量) <sup>†</sup>PermTrt, MMRM<sub>BW-REML</sub> を用いた投与群指示変数による並べ替え検定 ; PermWR, 帰無仮説のもとで推定した重み付き残差を並べ替え MMRM<sub>BW-REML</sub> で検定

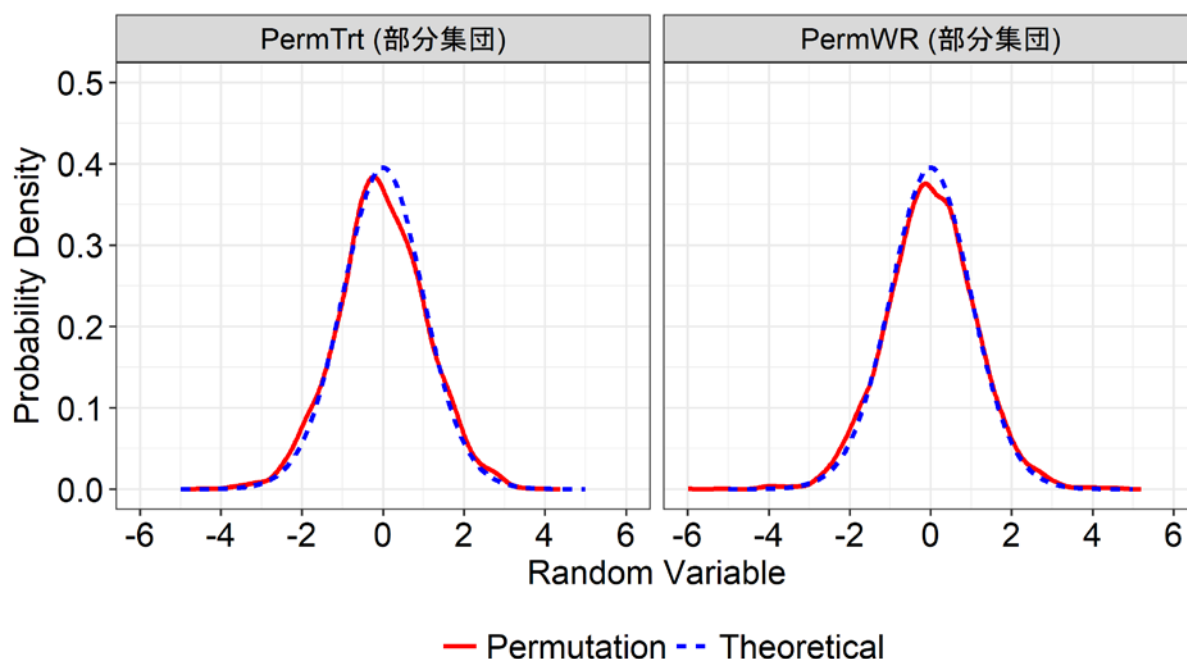


図 4.4-2 産後うつ臨床試験における並べ替え検定による推測手法における $t$ 検定統計量の経験分布<sup>†</sup> (8 ヶ月時点) [部分集団]

赤色実線，リサンプリングされた $t$ 検定統計量により推定したカーネル密度関数；青色点線，自由度 28 の $t$ 分布の確率密度関数；<sup>†</sup>PermTrt,  $\text{MMRM}_{\text{BW-REML}}$  を用いた投与群指示変数による並べ替え検定；PermWR, 帰無仮説のもとで推定した重み付き残差を並べ替え  $\text{MMRM}_{\text{BW-REML}}$  で検定

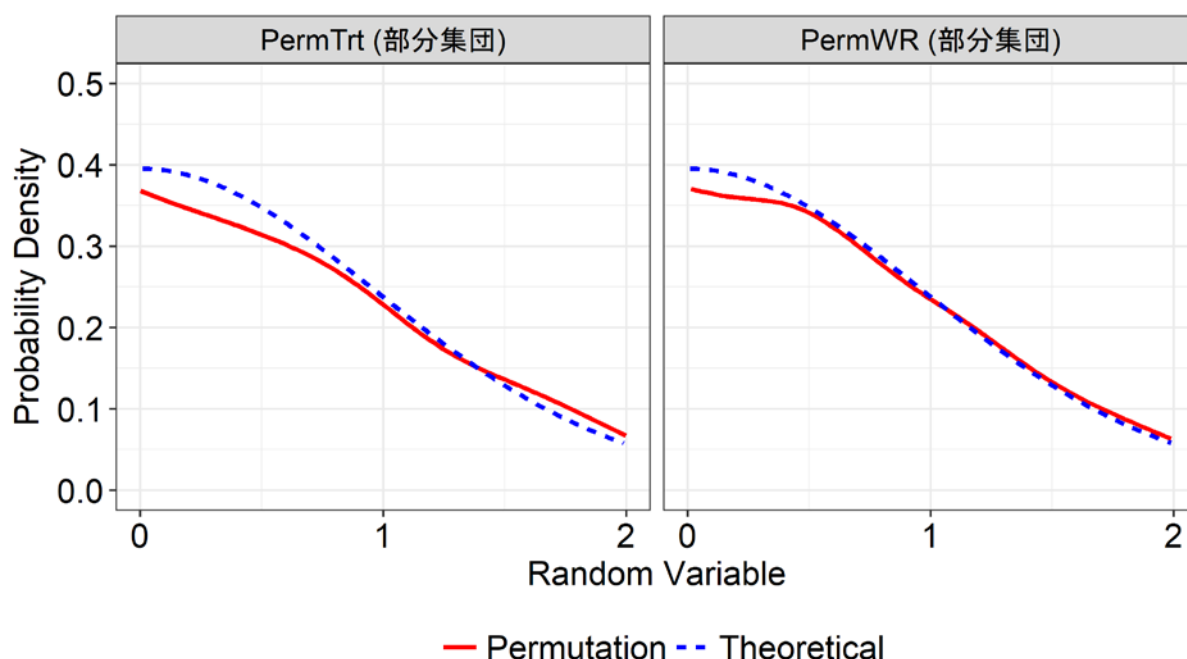


図 4.4-3 産後うつ臨床試験における並べ替え検定による推測手法における $t$ 検定統計量の経験分布<sup>†</sup> (確率変数 0~2 部分のみ, 8 ヶ月時点) [部分集団]

赤色実線，リサンプリングされた $t$ 検定統計量により推定したカーネル密度関数；青色点線，自由度 28 の $t$ 分布の確率密度関数；<sup>†</sup> PermTrt,  $\text{MMRM}_{\text{BW-REML}}$  を用いた投与群指示変数による並べ替え検定；PermWR, 帰無仮説のもとで推定した重み付き残差を並べ替え  $\text{MMRM}_{\text{BW-REML}}$  で検定

## 4.5 考察

### 4.5.1 並べ替え検定とパラメトリックブートストラップの比較

Good (2005)は、ノンパラメトリックブートストラップよりも並べ替え検定の方がより検出力が高いことを示した。パラメトリックブートストラップでは、まず帰無仮説のもとで線形モデルを適用させて推定したパラメータにより新しいデータを発生させる。発生したデータは、経験分布を構成する検定統計量を計算するために使われる。パラメトリックブートストラップでは、仮定したモデルにしたがってデータを発生させるため、モデルを誤特定した場合にパラメトリックブートストラップの妥当性が疑われることとなる。したがって、ブートストラップ法で発生させたデータによる分布は正しくないことになる。並べ替え検定は漸近的に名義の検定サイズを満たし、パラメトリック検定と同等の検出力を持つと考えられている (Hoeffding 1952)。

並べ替え検定は、残差が漸近的に交換可能ではない時のみ検定に影響を受ける。そのため、パラメトリックブートストラップよりもモデルの誤特定に対して頑健である。

### 4.5.2 計算の効率性

本研究の設定や実務上では、すべての並べ替えを行うことは組み合わせ数が膨大となり現実的ではないことが多い。そこで、本研究では Monte Carlo 法により並べ替えを実施した。しかし、本アプローチには小さなことだが欠点がある。並べ替えの Monte Carlo 法によるランダムサンプリングは実際には2項確率変数となっている点である (Good 2005)。これにより、検定の検出力が本来の並べ替え検定よりもわずかに低下する (Dwass 1957; Marozzi 2004; Pesarin and Salmaso 2010)。ただし、これは通常大きな問題とはならない。

並べ替えの分布のばらつきを最小限にするために、並べ替えに重みを付けたインポートランスサンプリングにより検定の検出力を改善することが可能である (Mehta et al. 1988)。

並べ替え検定をモンテカルロ法で実施する場合の適切な回数について、Marozzi (2004)にまとめられている。単純な1変数による検討だが、5%有意水準の検定を実施する場合に500~1000回程度、1%有意水準の検定を実施する場合に2000-5000回程度の Monte Carlo 法によるサンプリング回数が目安として提案されている。

慣習的な線形モデルの回帰係数に対する並べ替え検定は Anderson and Legendre (1999)が頻繁に引用されるが、本論文での Monte Carlo 法によるサンプリング回数は1000回であった。

### 4.5.3 調整する共変量が多い場合の性能

本研究のシミュレーションによる検討では、複数の共変量が存在する場合の性能について評価できていない。実務上も複数の共変量を調整する状況は容易に想像できるため、複数の共変量が存在した場合でも性能が良い方法はニーズが高いと考えられる。今後の残された課題となっている。

### 4.5.4 結論

シミュレーションによる検討及び事例データによる解析結果を踏まえると、並べ替え検定の考え方を適用した MMRM によるアプローチでは、投与群指示変数を並べ替える方法及び帰無仮説のもとでの回帰モデルにより推定した重み付き残差を並べ替える方法の有用性が示された。



欠測を伴う小標本の経時測定データに対して，投与群指示変数を並べ替える方法ではシミュレーション実験により Type I error rate が適切に維持されることが示された。

## 5. 結論

希少疾患を対象とした臨床試験及び疫学研究では、欠測を伴う小標本の経時測定データを解析対象とせざるを得ない状況が容易に想定される。一般的に普及している経時測定データの標準的な解析方法は、大標本近似に基づく手法であり、小標本ではその統計的推測の妥当性が保証されない。また、経時測定データにほぼ確実に発生する欠測により、統計的推測の妥当性はさらに保証されないこととなる。近年、新医薬品の開発対象や観察研究の研究対象も **common disease** から **rare disease** に移行しており、実務で遭遇する問題に対して実践的な解決方法が求められている。

本研究では、こうした問題を動機とし大きく 2 種類のアプローチによる実践的な解析方法を提案した。1 つ目は尤度比検定統計量の **Bartlett** 補正項とブートストラップ検定における経験分布を **MMRM** により近似的に推測する手法であり、2 つ目は並べ替え検定を **MMRM** に適用した手法である。

尤度比検定統計量の **Bartlett** 補正項を近似的に推測する手法は、得られた解析対象データから統計量の従う経験分布をパラメトリックブートストラップ法により推定し **Bartlett** 補正項を推定する。これまでに提案された高次漸近展開を伴う複雑な手法では容易に対応が難しい時点間相関のモデリングについても **MMRM** を利用することで柔軟に対応できる簡便かつ実践的な手法である。小標本で問題となる **Type I error rate** のインフレーションも適切に制御可能であることをシミュレーション実験及び事例解析により本研究で明らかにした。また、推定した尤度比検定統計量の経験分布をそのまま用いた手法も同等の性能を有していることを確認している。

並べ替え検定の手法自体は、これまでも臨床試験のデータ解析で用いられてきている。一方、並べ替え検定を **MMRM** に適用して詳細に検討・評価した研究は、本研究が初めてとなる。これまでに並べ替え検定を経時測定データに応用した研究報告は、経済学領域で認められるものの医学領域での事例は少ない。特に臨床試験の経時データ解析に標準的に利用されている **MMRM** との組み合わせで論じられている先行研究は皆無である。研究の動機は十分に存在するものの、ほとんど研究されていなかった領域であった。本研究ではこれまでに提案された並べ替え検定の考え方を 4 種類に整理し、そのうち 2 種類の並べ替えと **MMRM** を組み合わせた手法について性能を評価した。シミュレーション実験の結果、投与群指示変数を並べ替え **MMRM** と組み合わせた方法が最も適切に **Type I error rate** を制御できることを明らかにした。また、帰無仮説を仮定した回帰モデルにより推定した残差に重みを付与して並べ替え **MMRM** と組み合わせる方法も実践的には問題ないレベルで **Type I error rate** を制御できていた。

本研究で実施したシミュレーション実験の結果から、尤度比検定統計量の **Bartlett** 補正項を近似的に推測する方法及びブートストラップ検定、並べ替え検定を **MMRM** と組み合わせた投与群指示変数を並べ替える方法及び帰無仮説を仮定した回帰モデルにより推定した残差に重みを付与して並べ替える方法は、ほぼ同等なレベルで **Type I error rate** を制御できる可能性がある。対立仮説下での検出力についてもほぼ同等な性能である可能性が示されている。

一方、いずれの方法もブートストラップ法もしくは **Monte Carlo** 法による並べ替えが必要で、計算コストが大きい点はデメリットである。リサンプリング回数を減らすことにより計算コストを軽減することもできるが、検定精度をある一定以上に維持するためには 1000 回程度以上のリサンプリングは必要と考えられる。最近ではこうした計算負荷にも耐えうる高性能 PC が

入手可能な状況が整いつつあり、実践的には問題は徐々に軽減されてきている。

本研究は、医薬品開発や臨床研究及び疫学研究の経時測定データ解析の標準的な解析方法として普及している **MMRM** の実用上の問題に解決策を提示した応用研究である。実務家にとっても簡便で適用が容易な方法であると考えている。欠測を伴う小標本の経時測定データに対する妥当な解析手法は今後ますますニーズが増加すると予想される。本研究において提案したような統計的推測の妥当性が確保された手法が普及することで、新医薬品の承認に関する科学的評価や臨床研究及び疫学研究における結果・結論の妥当性がより確かなものとなることが期待される。

## 6. 謝辞

本研究に関して指導ならびに議論をして下さった総合研究大学院大学複合科学研究科統計科学専攻 野間久史 准教授に心より感謝申し上げます。

本研究の論文原稿に関して、共著者として大変貴重なレビューコメントを頂きました筑波大学医学医療系臨床試験・臨床疫学研究室 五所正彦 教授ならびに丸尾和司 准教授に深く御礼申し上げます。

勤務先であるヤンセンファーマ株式会社クオンティタティブサイエンス統括部 塚本友子 統括部長ならびに生物統計部 宮里盛幸 部長には、学位取得の機会を与えて頂き深く感謝申し上げます。ならびに所属部門のみなさまに深く感謝申し上げます。最後に支えてくださった家族に感謝を表します。

## 7. 参考文献

1. Anderson MJ, Legendre P. An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of statistical computation and simulation*. 1999; 62(3): 271-303.
2. Anderson MJ, Robinson J. Permutation tests for linear models. *Australian & New Zealand Journal of Statistics*. 2001; 43(1): 75-88.
3. Arnau J, Bono R, Vallejo G. Analyzing small samples of repeated measures data with the mixed-model adjusted F test. *Communications in Statistics-Simulation and Computation*. 2009; 38(5): 1083-1103.
4. Ashbeck EL and Bell ML. Single time point comparisons in longitudinal randomized controlled trials: power and bias in the presence of missing data. *BMC medical research methodology*. 2016; 16(1): 43.
5. Barndorff-Nielsen O. On a formula for the distribution of the maximum likelihood estimator. *Biometrika*. 1983; 70(2): 343-365.
6. Barndorff-Nielsen, OE. Inference on full or partial parameters based on the standardized signed log-likelihood ratio. *Biometrika*. 1986; 73(2): 307-322.
7. Barndorff-Nielsen OE and Hall P. On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika*. 1988; 75(2): 374-378.
8. Bartlett MS. Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London Series A, Mathematical and Physical Sciences* 1937; 160(901): 268-282.
9. Barnes S. A., Mallinckrodt C. H., Lindborg S. R., and Carter M. K. The impact of missing data and how it is handled on the rate of false-positive results in drug development. *Pharmaceutical Statistics*. 2008; 7: 215-225.
10. Beck, A. T., Steer, A. & Brown, G. K. Beck Depression Inventory Manual, 2<sup>nd</sup> ed. The Psychological Corporation: San Antonio, TX. 1996.
11. Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in medicine*. 1997; 16(20): 2349-2380.
12. Commenges D. Transformations which preserve exchangeability and application to permutation tests. *Journal of Nonparametric Statistics*. 2003; 15(2): 171-185.
13. Cordeiro GM, Cribari-Neto F. An Introduction to Bartlett Correction and Bias Reduction. New York: Springer; 2014.
14. Cox DR, Hinkley DV. Theoretical Statistics. London: Chapman & Hall; 1974.
15. Cox DR, Reid N. Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1987; 49(1): 1-39.
16. Davison AC, Hinkley DV. Bootstrap Methods and Their Application (Vol. 1). Cambridge: Cambridge University Press; 1997.
17. DeRubeis, R. J., Gelfland, L. A., Tang, T. Z. & Simons, A. D. Medications versus cognitive behavior therapy for severely depressed outpatients: mega-analysis of four randomized comparisons. *American Journal of Psychiatry*. 1999; 156(7): 1007-1013.
18. DeSouza C. M., Legedza A. T., and Sankoh A. J. An overview of practical approaches for handling missing data in clinical trials. *Journal of biopharmaceutical statistics*. 2009; 19(6): 1055-1073.

19. Diggle P, Heagerty P, Liang K-Y, Zeger S. Analysis of Longitudinal Data. 2002. Oxford University Press, Oxford.
20. Dwass M. Modified randomization tests for nonparametric hypotheses. The Annals of Mathematical Statistics. 1957; 28(1): 181-187.
21. Efron B. Bootstrap Methods: Another Look at the Jackknife. New York: Springer; 1992.
22. EMA. Guideline on missing data in confirmatory clinical trials. European Medicines Agency 2010. [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2010/09/WC500096793.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf). Accessed 15 Sep 2018.
23. Everitt BS, & Der G. A Handbook of Statistical Analyses using SAS (Second Edition). Boca Raton, FL: Chapman & Hall/CRC. 2001.
24. Everitt B, Hothorn T. An Introduction to Applied Multivariate Analysis with R. Springer. 2011. New York.
25. Fisher RA. The design of experiments. 1935. Oliver & Boyd, Edinburgh.
26. Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis. 2012. Vol. 998. John Wiley & Sons, New York.
27. Freedman D, Lane D. A nonstochastic interpretation of reported significance levels. Journal of Business & Economic Statistics. 1983; 1(4): 292-298.
28. Gillings, D. and Koch G. The application of the principle of intention-to-treat to the analysis of clinical trials. Drug Information Journal. 1991; 25(3): 411-424.
29. Good PI. Permutation, Parametric, and Bootstrap Tests of Hypotheses, 2005. Springer-Verlag: New York,
30. Gosh M, Hirakawa A, Noma H, Maruo K, Sato Y. Comparison of bias-corrected covariance estimators for MMRM analysis in longitudinal data with dropouts. Statistical methods in medical research. 2017; 26(5): 2389-2406.
31. Gregoire AJP, Kumar R, Everitt B, Henderson AF, Studd JWW. Transdermal oestrogen for treatment of severe postnatal depression. The Lancet. 1996; 347(9006): 930-933.
32. Guolo A, Brazzale AR, Salvan A. Improved inference on a scalar fixed effect of interest in nonlinear mixed-effects models. Computational statistics & data analysis. 2006; 51(3): 1602-1613.
33. Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association. 1977; 72(358): 320-338.
34. Hoeffding W. The large-sample power of tests based on permutations of observations. The Annals of Mathematical Statistics. 1952; 169-192.
35. Hogan JW, Roy J, Korkontzelou C. Handling drop-out in longitudinal studies. Statistics in medicine. 2004; 23(9): 1455-1497.
36. Kackar RN, and Harville DA. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. Journal of the American Statistical Association. 1984; 79(388): 853-862.
37. Kauermann G and Carroll RJ. A note on the efficiency of sandwich covariance matrix estimation. Journal of the American Statistical Association. 2001; 96(456): 1387-1396.
38. Kennedy FE. Randomization tests in econometrics. Journal of Business & Economic Statistics. 1995; 13(1): 85-94.

39. Kennedy PE and Cade BS. Randomization tests for multiple regression. *Communications in Statistics-Simulation and Computation*. 1996; 25(4): 923-936.
40. Kenward MG and Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*. 1997; 53(3): 983-997.
41. Kenward MG and Roger JH. An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics & Data Analysis*. 2009; 53(7): 2583-2595.
42. Laird NM and Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982; 38(4): 963-974.
43. Laird NM. Missing data in longitudinal studies. *Statistics in medicine*. 1988; 7(1-2): 305-315.
44. Lane P. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. *Pharmaceutical statistics*. 2008; 7(2): 93-106.
45. Lawley DN. A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*. 1956; 43(3/4): 295-303.
46. Lee OE, and Braun TM. Permutation tests for random effects in linear mixed models. *Biometrics*. 2012; 68(2): 486-493.
47. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, Frangakis C, Hogan JW, Molenberghs G, Murphy SA, Neaton JD, Rotnitzky A, Scharfstein D, Shih WJ, Siegel JP, Stern H. The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*. 2012a; 367(14): 1355-1360.
48. Little RJ, Cohen ML, Dickersin K, Emerson SS, Farrar JT, Neaton JD, Shih W, Siegel JP, Stern H. The design and conduct of clinical trials to limit missing data. *Statistics in medicine*. 2012b; 31(28): 3433-3443.
49. Liu G., and Gould A. L. Comparison of alternative strategies for analysis of longitudinal trials with dropouts. *Journal of Biopharmaceutical Statistics*. 2002; 12(2): 207-226.
50. Liu M., Wei L., and Zhang J. Review of guidelines and literature for handling missing data in longitudinal clinical trials with a case study. *Pharmaceutical statistics*. 2006; 5(1): 7-18.
51. Lu K and Mehrotra DV. Specification of covariance structure in longitudinal data analysis for randomized clinical trials. *Statistics in medicine*. 2010; 29(4): 474-488.
52. Lyons B, Peters D. Applying skovgaard's modified directed likelihood statistic to mixed linear models. *Journal of statistical computation and simulation*. 2000; 65(1-4): 225-242.
53. Mallinckrodt C. H., Clark W. S., and David S. R. Accounting for dropout bias using mixed-effects models. *Journal of biopharmaceutical statistics*. 2001a; 11(1-2): 9-21.
54. Mallinckrodt C. H., Clark W. S., and David S. R. Type I error rates from mixed effects model repeated measures versus fixed effects ANOVA with missing values imputed via last observation carried forward. *Drug Information Journal*. 2001b; 35(4): 1215-1225.
55. Mallinckrodt CH, Sanger TM, Dube S, DeBrot DJ, Molenberghs G, Carroll RJ, Potter WZ, Tollefson GD. Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. *Biological psychiatry*. 2003; 53(8): 754-760.
56. Mallinckrodt C. H., Kaiser C. J., Watkin J. G., Detke M. J., Molenberghs G., and Carroll R. J. Type I

- error rates from likelihood-based repeated measures analyses of incomplete longitudinal data. *Pharmaceutical statistics*. 2004a; 3(3): 171-186.
57. Mallinckrodt C. H., Kaiser C. J., Watkin J. G., Molenberghs G., and Carroll R. J. The effect of correlation structure on treatment contrasts estimated from incomplete clinical trial data with likelihood-based repeated measures compared with last observation carried forward ANOVA. *Clinical trials*. 2004b; 1(6): 477-489.
  58. Mallinckrodt CH, Watkin JG, Molenberghs G, Carroll RJ. Choice of the primary analysis in longitudinal clinical trials. *Pharmaceutical statistics*. 2004c; 3(3): 161-169.
  59. Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Information Journal*. 2008; 42(4): 303-319.
  60. Mancl LA and DeRouen TA. A covariance estimator for GEE with improved small-sample properties. *Biometrics*. 2001; 57(1): 126-134.
  61. Manly BF. *Randomization, bootstrap and Monte Carlo methods in biology*. 1997. Chapman & Hall, London.
  62. Marozzi M. Some remarks about the number of permutations one should consider to perform a permutation test. *Statistica*. 2004; 64(1): 193-201.
  63. Matsuyama Y. Analysis of Missing Data in Longitudinal Studies: A Review. *Japanese Journal of Biometrics*. 2004; 25: 89-116.
  64. McCabe-Beane JE, Segre LS, Perkhounkova Y, Stuart S, O'Hara MW. The identification of severity ranges for the Edinburgh Postnatal Depression Scale. *Journal of Reproductive and Infant Psychology*. 2016; 34(3): 293-303.
  65. McCulloch C, Searle S, and Neuhaus J. *Generalized, Linear, and Mixed Models*, 2nd edition. Wiley. 2008.
  66. Mehta CR, Patel NR, Senchaudhuri P. Importance sampling for estimating exact probabilities in permutational inference. *Journal of the American Statistical Association*. 1988; 83(404): 999-1005.
  67. Melo TFN, Ferrari SLP, Cribari-Neto F. Improved testing inference in mixed linear models. *Computational Statistics & Data Analysis*. 2009; 53(7): 2573-2582.
  68. Mielke PW, and Berry KJ. *Permutation Methods: a Distance Function Approach*. New York: Springer-Verlag. 2007.
  69. National Research Council. *The prevention and treatment of missing data in clinical trials*. National Academies Press: Washington, DC, 2010.
  70. Noma H, Nagashima K, Maruo K, Goshio M, Furukawa TA. Bartlett-type corrections and bootstrap adjustments of likelihood-based inference methods for network meta-analysis. *Statistics in medicine*. 2018; 37(7): 1178-1190.
  71. O'Neill RT, Temple R. The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. *Clinical Pharmacology & Therapeutics*. 2012; 91(3): 550-554.
  72. Ottoboni K, Lewis F, Salmaso L. An Empirical Comparison of Parametric and Permutation Tests for Regression Analysis of Randomized Experiments. *Statistics in Biopharmaceutical Research*. 2018;



(just-accepted): 1-29.

73. Patterson HD, Thompson R. Recovery of inter-block information when block sizes are unequal. *Biometrika*. 1971; 58(3): 545-554.
74. Peace KE, Chen D-G, Menon S. *Biopharmaceutical Applied Statistics Symposium: Volume 2 Biostatistical Analysis of Clinical Trials*. 2018. Springer.
75. Pesarin F, and Salmaso L. *Permutation Tests for Complex Data: Theory, Applications and Software*. John Wiley and Sons. 2010. West Sussex, England UK.
76. Pinheiro, J. C. and Bates, D. M. *Mixed-Effects Models in S and S-PLUS*. Springer. 2000. New York.
77. Proudfoot J, Goldberg D, Mann A, Everitt B, Marks I, Gray JA. Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. *Psychological medicine*. 2003; 33(2): 217-227.
78. Rabinowitz J, Levine SZ, Barkai O, Davidov O. Dropout rates in randomized clinical trials of antipsychotics: a meta-analysis comparing first-and second-generation drugs and an examination of the role of trial design features. *Schizophrenia bulletin*. 2008; 35(4): 775-788.
79. Ratitch B, O’Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceutical statistics*. 2013; 12(6): 337-347.
80. Rocke DM. Bootstrap Bartlett adjustment in seemingly unrelated regression. *Journal of the American Statistical Association*. 1989; 84(406): 598-601.
81. Rubin DB. Inference and missing data. *Biometrika*. 1976; 63(3): 581-592.
82. SAS Institute Inc. *SAS/STAT 14.3 User’s Guide The GLIMMIX Procedure*. 2017. <https://support.sas.com/documentation/onlinedoc/stat/indexproc.html>. Accessed on 17 Oct 2018.
83. Satterthwaite, F. An approximate distribution of estimates of variance components. *Biometrics bulletin*. 1946; 2(6): 110-114.
84. Schaalje GB, McBride JB, Fellingham GW. Adequacy of approximations to distributions of test statistics in complex mixed linear models. *Journal of Agricultural Biological and Environmental Statistics*. 2002; 7(4): 512-524.
85. Schluchter MD, Elashoff JT. Small-sample adjustments to tests with unbalanced repeated measures assuming several covariance structures. *Journal of Statistical Computation and Simulation*. 1990; 37(1-2): 69-87.
86. Senn SJ, Auclair P. The graphical representation of clinical trials with particular reference to measurements over time. *Statistics in Medicine*. 1990; 9(11): 1287-1302.
87. Severini TA. An approximation to the modified profile likelihood function. *Biometrika*. 1998; 85(2): 403-411.
88. Siddiqui O., Hung H. J., and O’Neill R. MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets. *Journal of biopharmaceutical statistics*. 2009; 19(2): 227-246.
89. Skovgaard IM. An explicit large-deviation approximation to one-parameter tests. *Bernoulli*. 1996; 2(2): 145-165.
90. Stein MC, da Silva MF, Duczmal LH. Alternatives to the usual likelihood ratio test in mixed linear models. *Computational Statistics & Data Analysis*. 2014; 69: 184-197.

91. Ter Braak CJ. Permutation versus bootstrap significance tests in multiple regression and ANOVA. In: Bootstrapping and related techniques. 1992. Springer, pp 79-85.
92. Tibshirani RJ, Efron B. An introduction to the Bootstrap. Monographs on Statistics and Applied Probability. 1993; 57: 1-436.
93. Verbeke G. Linear Mixed Models in Practice. Springer, 1997.
94. Verbeke G. and Molenberghs G. Linear Mixed Models for Longitudinal Data. New York: Springer; 2000.
95. Verbeke G. and Molenberghs G. Linear Mixed Models for Longitudinal Data. New York: Springer, 2009.
96. Wehrens R, Putter H, Buydens LMC. The bootstrap: a tutorial. Chemometrics and intelligent laboratory systems. 2000; 54(1): 35-52.
97. Zeng Q, Davidian M. Bootstrap-adjusted calibration confidence intervals for immunoassay. Journal of the American Statistical Association. 1997; 92(437): 278-290.
98. Zucker DM, Lieberman O, Manor O. Improved small sample inference in the mixed linear model: Bartlett correction and adjusted likelihood. Journal of the royal statistical society. Series B. 2000; 62(4): 827-838.
99. 小西貞則, 越智義道, 大森裕浩. 計算統計学の方法. 朝倉書店. 2008.
100. 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. 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—NAS レポート, EMA ガイドライン, estimand と 解 析 方 法 の 概 説 — . 2014.  
[http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/missing\\_data.html](http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/missing_data.html). Accessed 10 Oct 2018.
101. 松山裕・山口拓洋. 医学統計のための線型混合モデル (編訳) . サイエンティスト社. 2001.

## 8. 付録

### 8.1 シミュレーションデータの要約

実施したシミュレーションにおける最終測定時点のサンプルサイズについて、反復回数 1000 回の平均値を表に示した。

表 8.1-1 最終測定時点のサンプルサイズの平均値（1 群 10 例）

シナリオ	投与群	MCAR		MAR	
		20%	40%	20%	40%
シナリオ 1	プラセボ群	7.81	5.91	7.81	5.68
	実薬群	7.83	5.81	7.75	5.66
シナリオ 2	プラセボ群	7.66	5.64	7.51	5.28
	実薬群	8.12	6.40	7.86	5.78
シナリオ 3	プラセボ群	7.41	5.31	7.38	5.21
	実薬群	8.29	6.40	8.15	6.18
シナリオ 4	プラセボ群	7.41	5.31	7.32	5.05
	実薬群	8.29	6.40	8.21	6.18
シナリオ 5	プラセボ群	7.41	5.31	7.12	4.86
	実薬群	8.29	6.40	8.95	7.27
シナリオ 6	プラセボ群	7.41	5.31	7.20	4.88
	実薬群	8.29	6.40	8.94	7.25

反復回数 1000 回の平均値を表示

表 8.1-2 最終測定時点のサンプルサイズの平均値（1 群 20 例）

シナリオ	投与群	MCAR		MAR	
		20%	40%	20%	40%
シナリオ 1	プラセボ群	15.71	11.55	15.52	11.38
	実薬群	15.63	11.77	15.65	11.46
シナリオ 2	プラセボ群	15.40	10.98	14.90	10.56
	実薬群	16.38	12.98	15.88	11.70
シナリオ 3	プラセボ群	14.82	10.34	14.75	10.38
	実薬群	16.80	12.91	16.42	12.44
シナリオ 4	プラセボ群	14.82	10.34	14.57	10.12
	実薬群	16.80	12.91	16.52	12.45
シナリオ 5	プラセボ群	14.82	10.34	14.22	9.71
	実薬群	16.80	12.91	17.98	14.61
シナリオ 6	プラセボ群	14.82	10.34	14.31	9.77
	実薬群	16.80	12.91	18.02	14.61

反復回数 1000 回の平均値を表示

## 8.2 ブートストラップ法による Bartlett 補正項の推定値

ブートストラップ法により推定した Bartlett 補正項の分布を明らかにするために、推定した Bartlett 補正項についてヒストグラムを作成した。付録にヒストグラムを掲載する。

図の番号及び図の種類，対象シナリオの対応表

図の種類	図の番号	対象シナリオ	1 群当たりの例数
ヒストグラム	図 8.2-1	1	10
	図 8.2-2	2	10
	図 8.2-3	3	10
	図 8.2-4	4	10
	図 8.2-5	5	10
	図 8.2-6	6	10
ヒストグラム	図 8.2-7	1	20
	図 8.2-8	2	20
	図 8.2-9	3	20
	図 8.2-10	4	20
	図 8.2-11	5	20
	図 8.2-12	6	20

### 8.2.1 Bartlett 補正項推定値のヒストグラム : $N = 20$ (1 群 10 例)

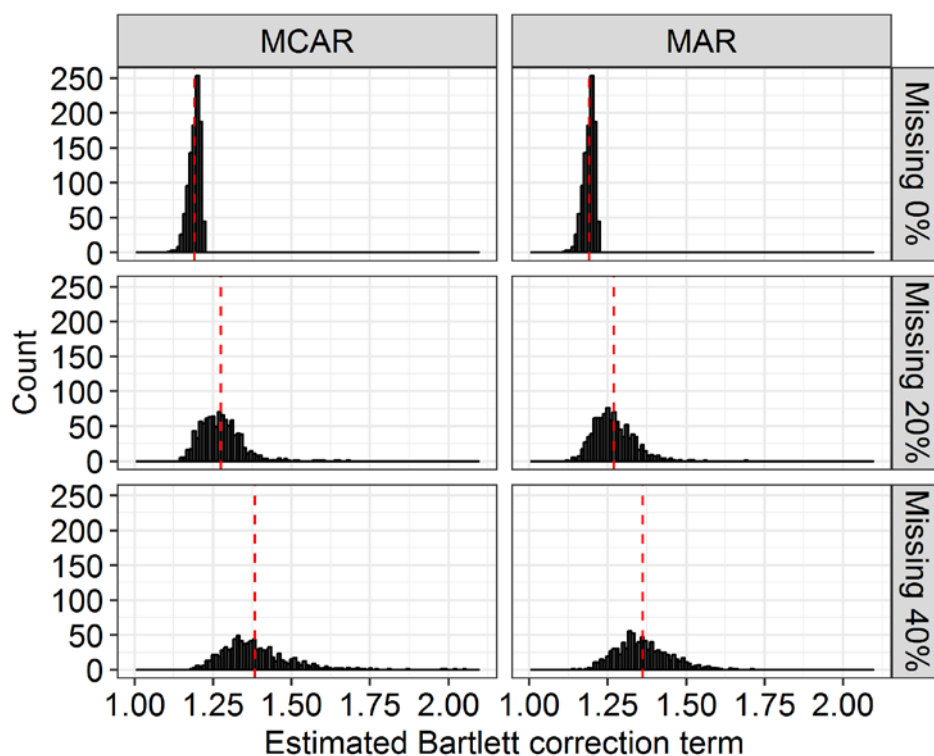


図 8.2-1 Bartlett 補正項推定値のヒストグラムのシミュレーション実験結果 (シナリオ 1, MCAR 及び MAR 下, 1 群 10 例)

赤点線, Bartlett 補正項推定値の平均

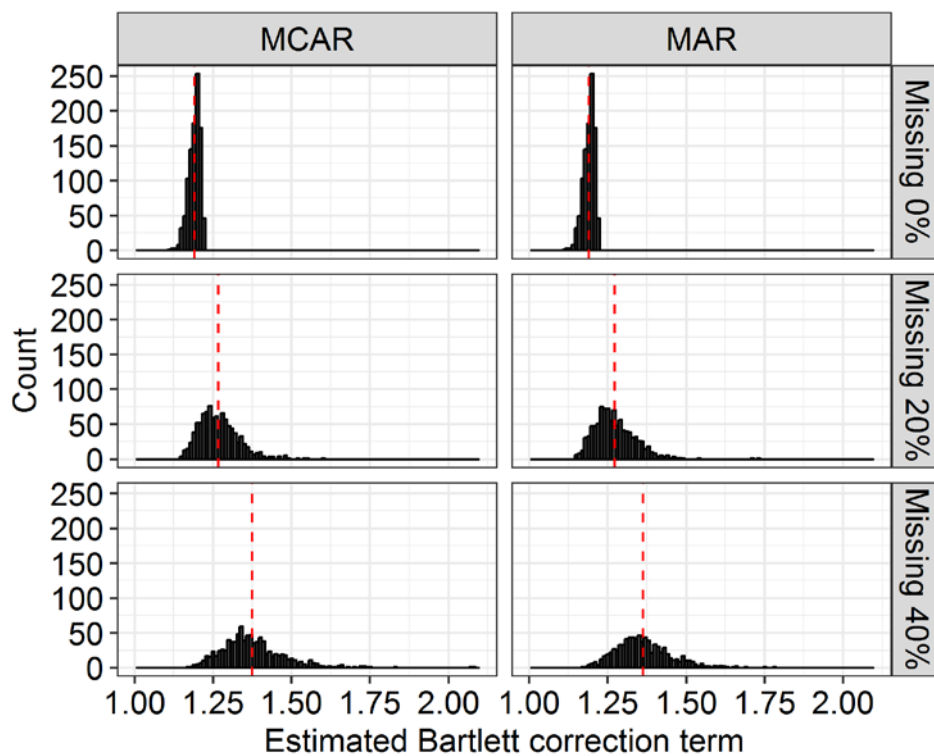


図 8.2-2 Bartlett 補正項推定値のヒストグラムのシミュレーション実験結果 (シナリオ 2, MCAR 及び MAR 下, 1 群 10 例)

赤点線, Bartlett 補正項推定値の平均

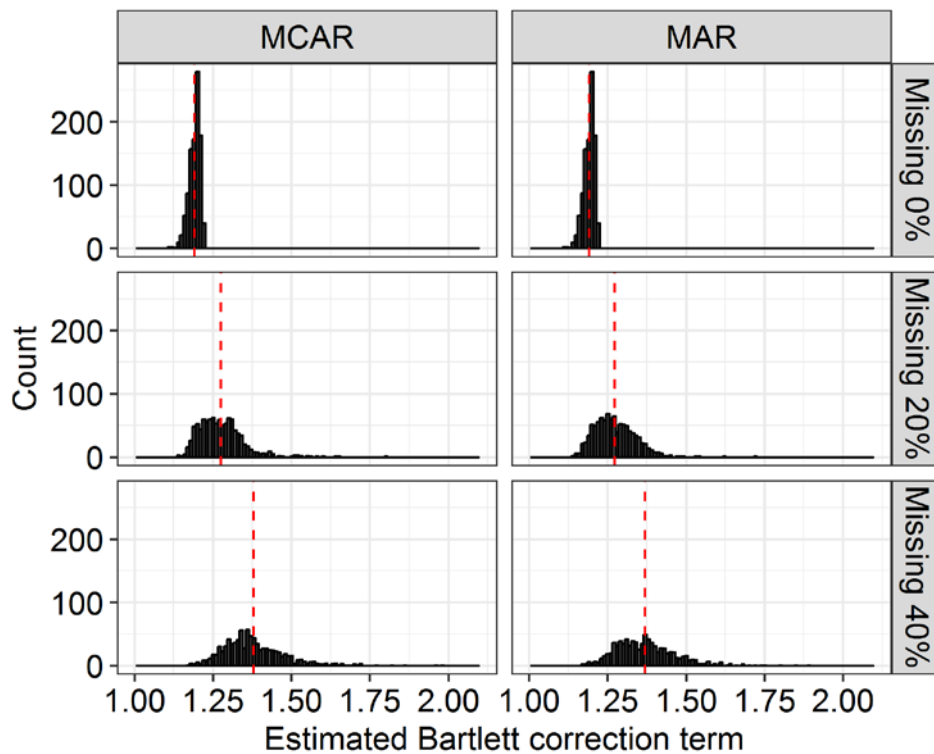


図 8.2-3 Bartlett 補正項推定値のヒストグラムのシミュレーション実験結果（シナリオ 3, MCAR 及び MAR 下, 1 群 10 例）

赤点線, Bartlett 補正項推定値の平均

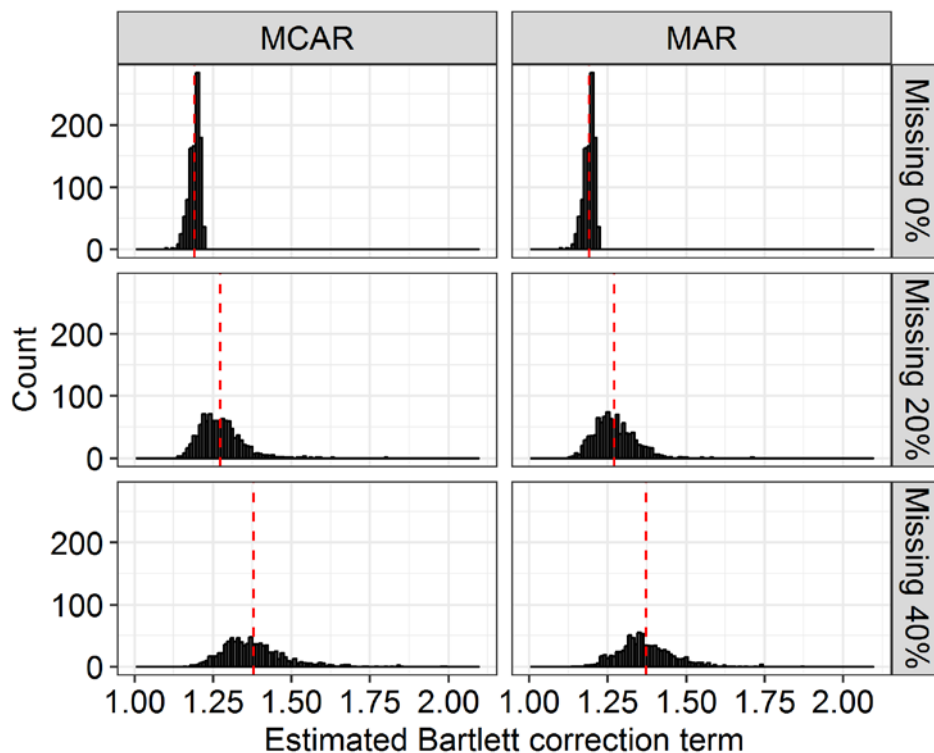


図 8.2-4 Bartlett 補正項推定値のヒストグラムのシミュレーション実験結果（シナリオ 4, MCAR 及び MAR 下, 1 群 10 例）

赤点線, Bartlett 補正項推定値の平均

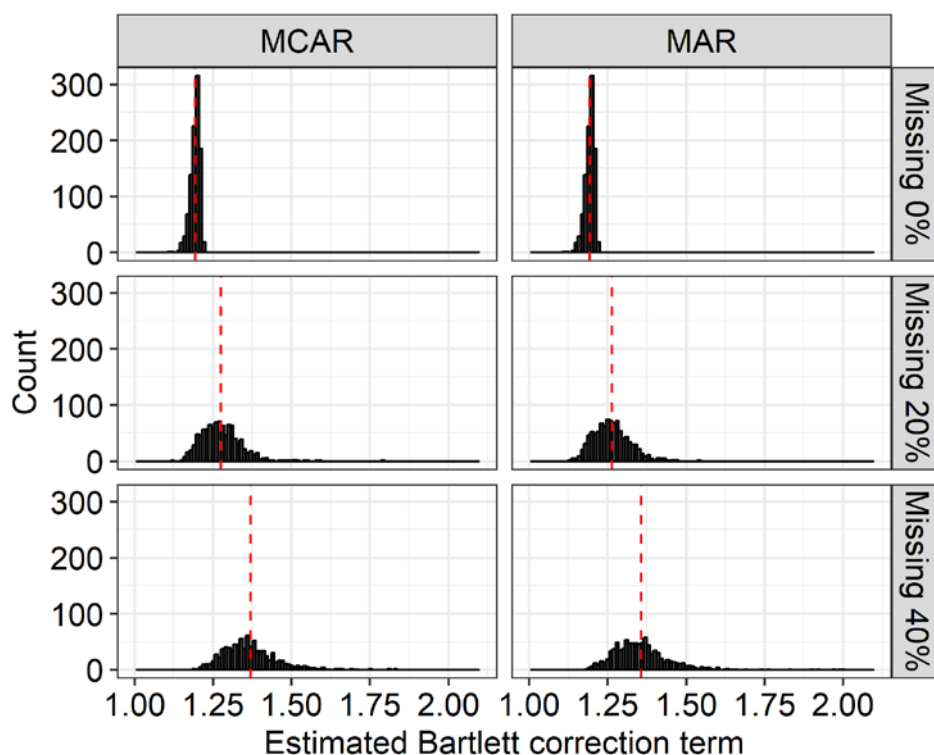


図 8.2-5 Bartlett 補正項推定値のヒストグラムのシミュレーション実験結果（シナリオ 5, MCAR 及び MAR 下, 1 群 10 例）

赤点線, Bartlett 補正項推定値の平均

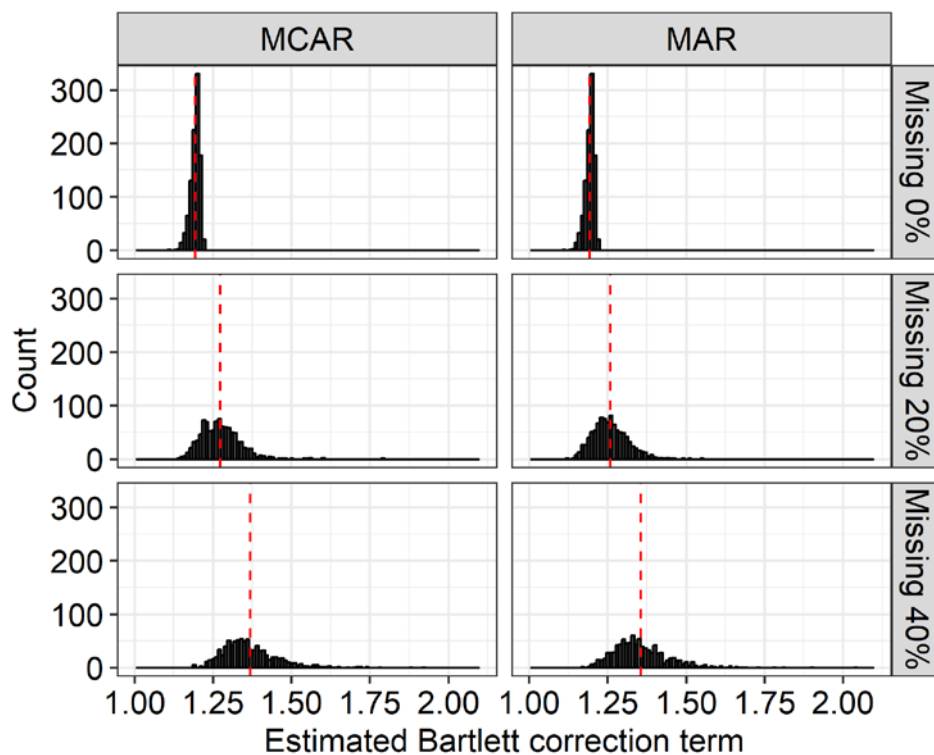


図 8.2-6 Bartlett 補正項推定値のヒストグラムのシミュレーション実験結果（シナリオ 6, MCAR 及び MAR 下, 1 群 10 例）

赤点線, Bartlett 補正項推定値の平均

## 8.2.2 Bartlett 補正項推定値のヒストグラム : $N = 40$ (1 群 20 例)

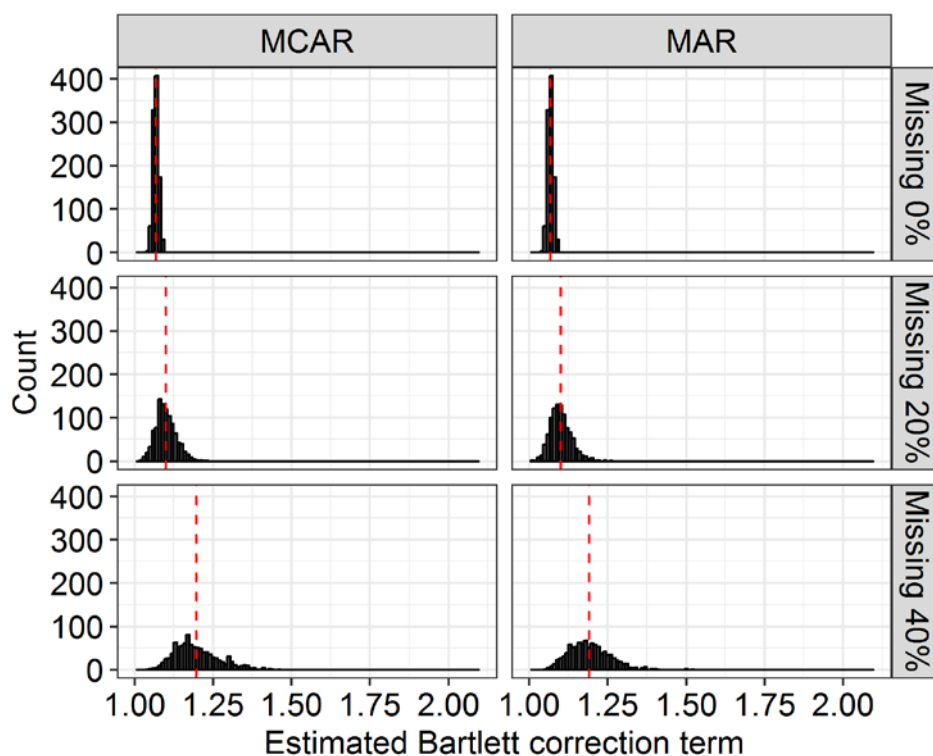


図 8.2-7 Bartlett 補正項推定値のヒストグラムのシミュレーション実験結果 (シナリオ 1, MCAR 及び MAR 下, 1 群 20 例)

赤点線, Bartlett 補正項推定値の平均

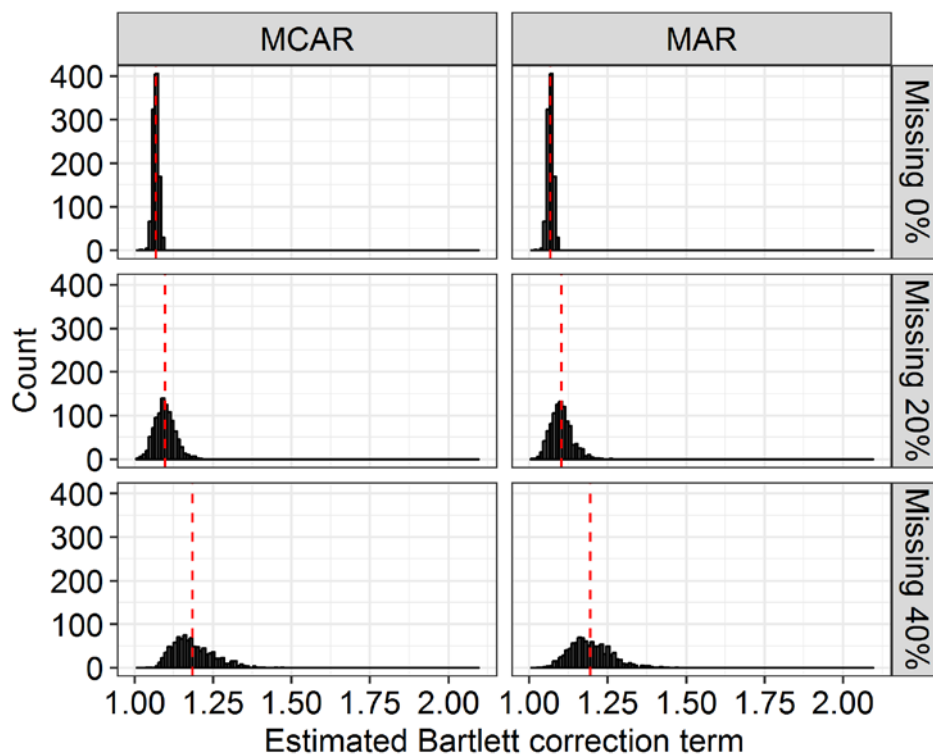


図 8.2-8 Bartlett 補正項推定値のヒストグラムのシミュレーション実験結果 (シナリオ 2, MCAR 及び MAR 下, 1 群 20 例)

赤点線, Bartlett 補正項推定値の平均



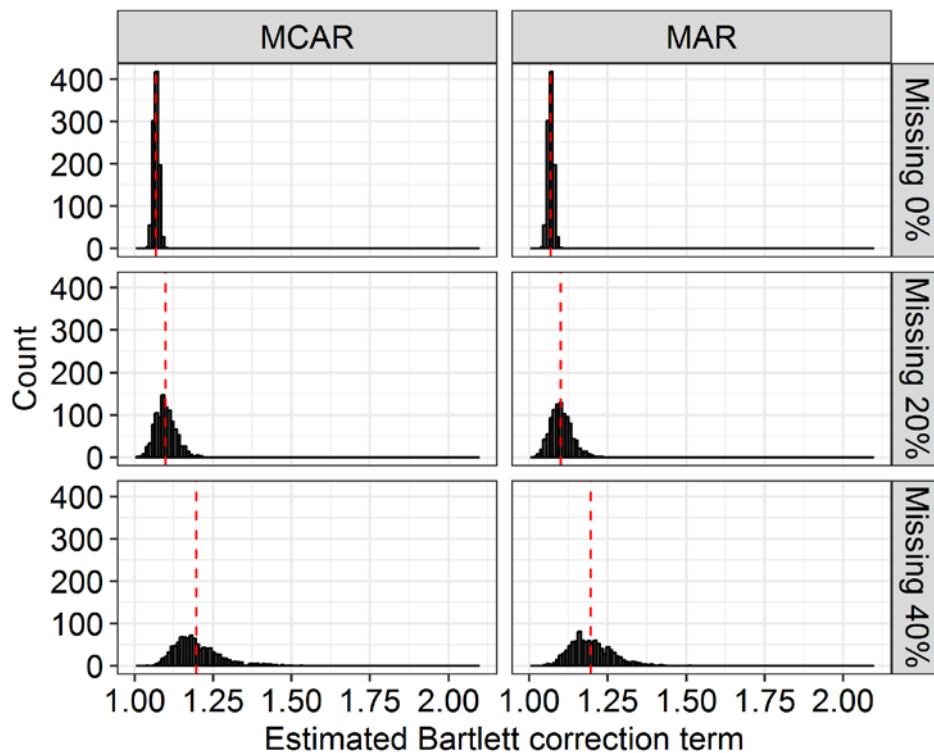


図 8.2-9 Bartlett 補正項推定値のヒストグラムのシミュレーション実験結果（シナリオ 3, MCAR 及び MAR 下, 1 群 20 例）

赤点線, Bartlett 補正項推定値の平均

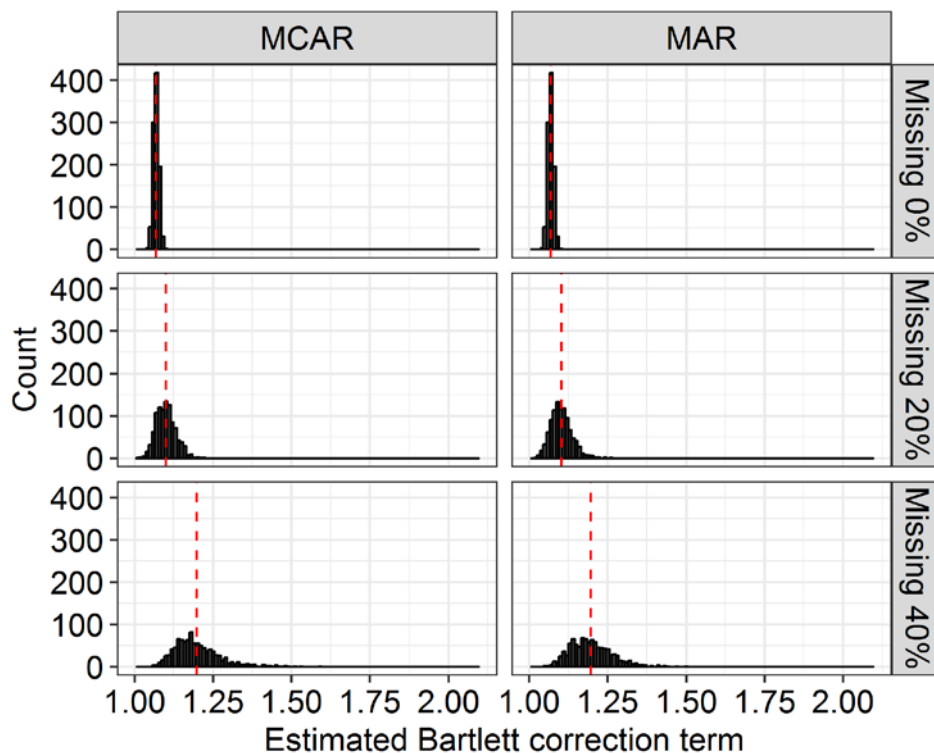


図 8.2-10 Bartlett 補正項推定値のヒストグラムのシミュレーション実験結果（シナリオ 4, MCAR 及び MAR 下, 1 群 20 例）

赤点線, Bartlett 補正項推定値の平均

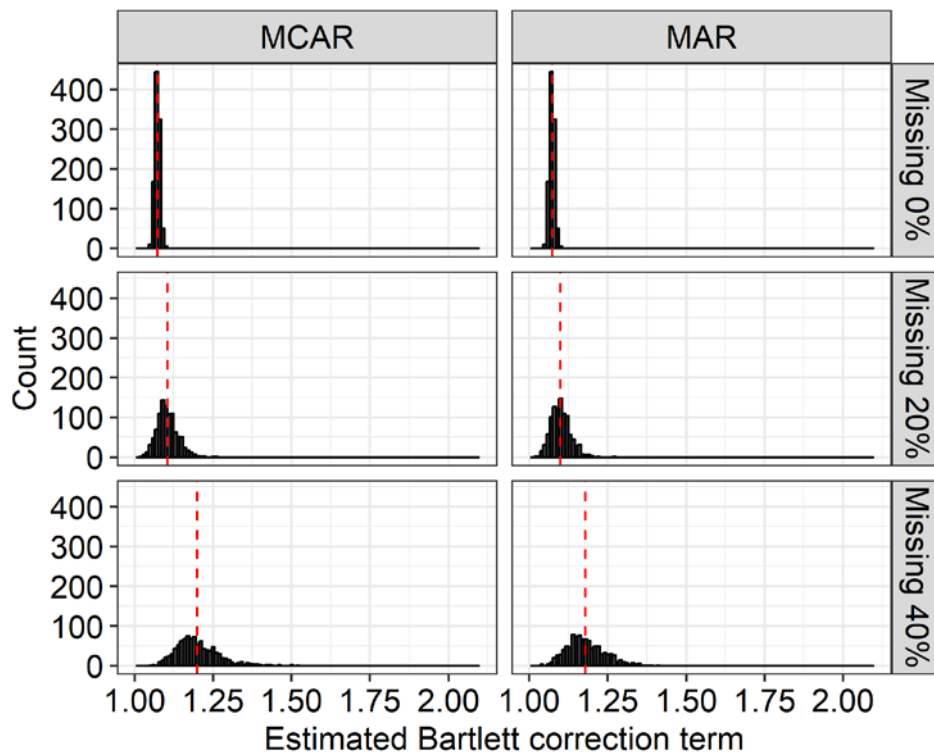


図 8.2-11 Bartlett 補正項推定値のヒストグラムのシミュレーション実験結果（シナリオ 5, MCAR 及び MAR 下, 1 群 20 例）

赤点線, Bartlett 補正項推定値の平均

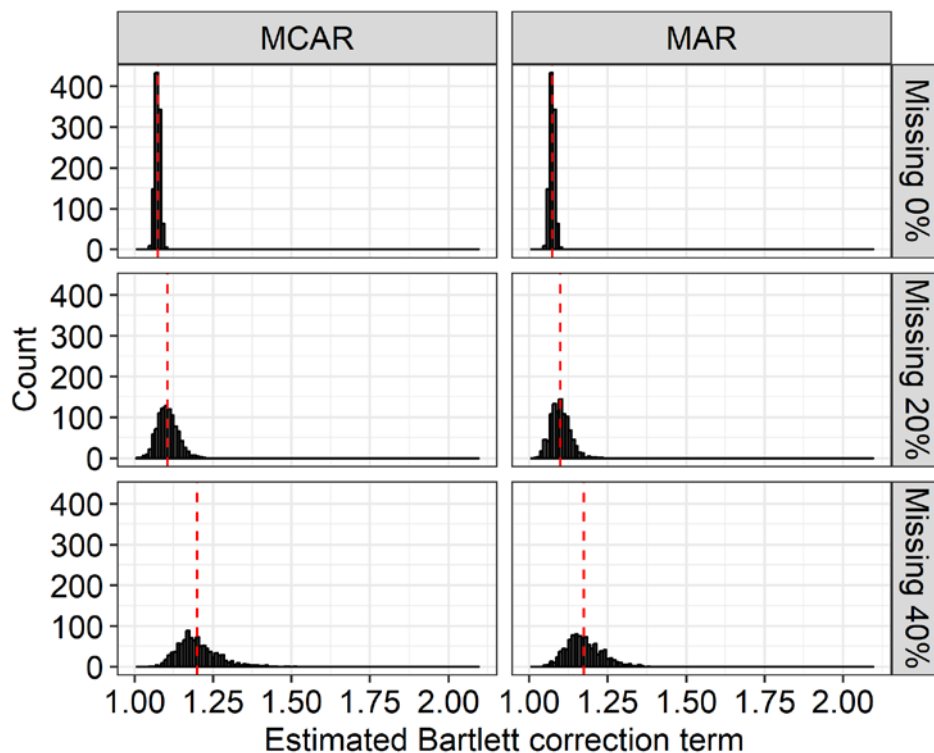


図 8.2-12 Bartlett 補正項推定値のヒストグラムのシミュレーション実験結果（シナリオ 6, MCAR 及び MAR 下, 1 群 20 例）

赤点線, Bartlett 補正項推定値の平均

### 8.3 Kenward and Roger 法の性能に関する検討

反復回数 1000 回とした場合の図（収束割合，（ $t$ 検定の）自由度），及び反復回数 10000 回とした場合の検討結果を掲載する。

#### 8.3.1 Type I error rate 及び検出力

表 8.3-1 Type I error rate（％）及び検出力（％）のシミュレーション実験結果<sup>†</sup>（シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 5 例, 反復回数 10000 回）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	5.02	---	---	5.02	---	---
	MMRM <sub>KR</sub>	5.03	---	---	5.03	---	---
	MMRM <sub>BW</sub>	5.02	---	---	5.02	---	---
2	MMRM <sub>SAT</sub>	5.02	---	---	5.02	---	---
	MMRM <sub>KR</sub>	5.05	---	---	5.05	---	---
	MMRM <sub>BW</sub>	5.02	---	---	5.02	---	---
3	MMRM <sub>SAT</sub>	10.24	---	---	10.24	---	---
	MMRM <sub>KR</sub>	10.30	---	---	10.30	---	---
	MMRM <sub>BW</sub>	10.24	---	---	10.24	---	---
4	MMRM <sub>SAT</sub>	10.24	---	---	10.24	---	---
	MMRM <sub>KR</sub>	10.33	---	---	10.33	---	---
	MMRM <sub>BW</sub>	10.24	---	---	10.24	---	---
5	MMRM <sub>SAT</sub>	52.33	---	---	52.33	---	---
	MMRM <sub>KR</sub>	52.19	---	---	52.19	---	---
	MMRM <sub>BW</sub>	52.33	---	---	52.33	---	---
6	MMRM <sub>SAT</sub>	52.33	---	---	52.33	---	---
	MMRM <sub>KR</sub>	52.24	---	---	52.24	---	---
	MMRM <sub>BW</sub>	52.33	---	---	52.33	---	---

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

表 8.3-2 Type I error rate（％）及び検出力（％）のシミュレーション実験結果<sup>†</sup>（シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	4.93	5.79	8.03	4.93	5.03	6.08
	MMRM <sub>KR</sub>	4.93	4.94	5.90	4.93	4.33	4.08
	MMRM <sub>BW</sub>	4.93	6.12	9.29	4.93	5.95	8.47
2	MMRM <sub>SAT</sub>	4.93	5.73	7.94	4.93	5.14	6.27
	MMRM <sub>KR</sub>	4.93	5.01	6.02	4.93	4.38	4.22
	MMRM <sub>BW</sub>	4.93	6.10	9.14	4.93	6.08	8.71
3	MMRM <sub>SAT</sub>	15.97	15.02	14.78	15.97	13.26	11.89
	MMRM <sub>KR</sub>	15.97	13.62	11.77	15.97	11.49	8.59
	MMRM <sub>BW</sub>	15.97	15.74	16.63	15.97	14.90	15.59
4	MMRM <sub>SAT</sub>	15.97	15.01	14.78	15.97	13.11	11.71
	MMRM <sub>KR</sub>	15.97	13.63	11.77	15.97	11.74	8.60
	MMRM <sub>BW</sub>	15.97	15.74	16.62	15.97	14.77	15.24
5	MMRM <sub>SAT</sub>	86.38	78.07	65.50	86.38	76.47	60.28
	MMRM <sub>KR</sub>	86.38	75.79	59.70	86.38	74.27	53.37
	MMRM <sub>BW</sub>	86.38	79.11	68.81	86.38	78.81	66.18
6	MMRM <sub>SAT</sub>	86.38	78.09	65.46	86.38	76.41	60.30

MMRM <sub>KR</sub>	86.38	75.82	59.63	86.38	74.39	53.28
MMRM <sub>BW</sub>	86.38	79.12	68.82	86.38	78.60	66.16

†MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

表 8.3-3 Type I error rate (%) 及び検出力 (%) のシミュレーション実験結果† (シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 20 例, 反復回数 10000 回)

シナリオ	方法†	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	4.77	5.37	6.35	4.77	4.92	5.36
	MMRM <sub>KR</sub>	4.77	5.05	5.16	4.77	4.48	4.00
	MMRM <sub>BW</sub>	4.77	5.47	6.73	4.77	5.38	6.51
2	MMRM <sub>SAT</sub>	4.77	5.22	6.26	4.77	4.95	5.47
	MMRM <sub>KR</sub>	4.77	4.92	5.34	4.77	4.57	4.06
	MMRM <sub>BW</sub>	4.77	5.34	6.66	4.77	5.39	6.53
3	MMRM <sub>SAT</sub>	28.58	25.46	22.26	28.58	23.73	19.34
	MMRM <sub>KR</sub>	28.58	24.63	19.53	28.58	22.53	16.01
	MMRM <sub>BW</sub>	28.58	25.81	23.08	28.58	24.84	21.46
4	MMRM <sub>SAT</sub>	28.58	25.44	22.26	28.58	23.69	18.80
	MMRM <sub>KR</sub>	28.58	24.63	19.55	28.58	22.41	15.82
	MMRM <sub>BW</sub>	28.58	25.82	23.09	28.58	24.77	20.86
5	MMRM <sub>SAT</sub>	99.32	97.98	93.62	99.32	97.96	93.36
	MMRM <sub>KR</sub>	99.32	97.81	92.25	99.32	97.82	91.57
	MMRM <sub>BW</sub>	99.32	98.05	94.01	99.32	98.11	94.16
6	MMRM <sub>SAT</sub>	99.32	97.98	93.63	99.32	98.07	93.05
	MMRM <sub>KR</sub>	99.32	97.81	92.26	99.32	97.82	91.43
	MMRM <sub>BW</sub>	99.32	98.05	94.02	99.32	98.22	93.83

†MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

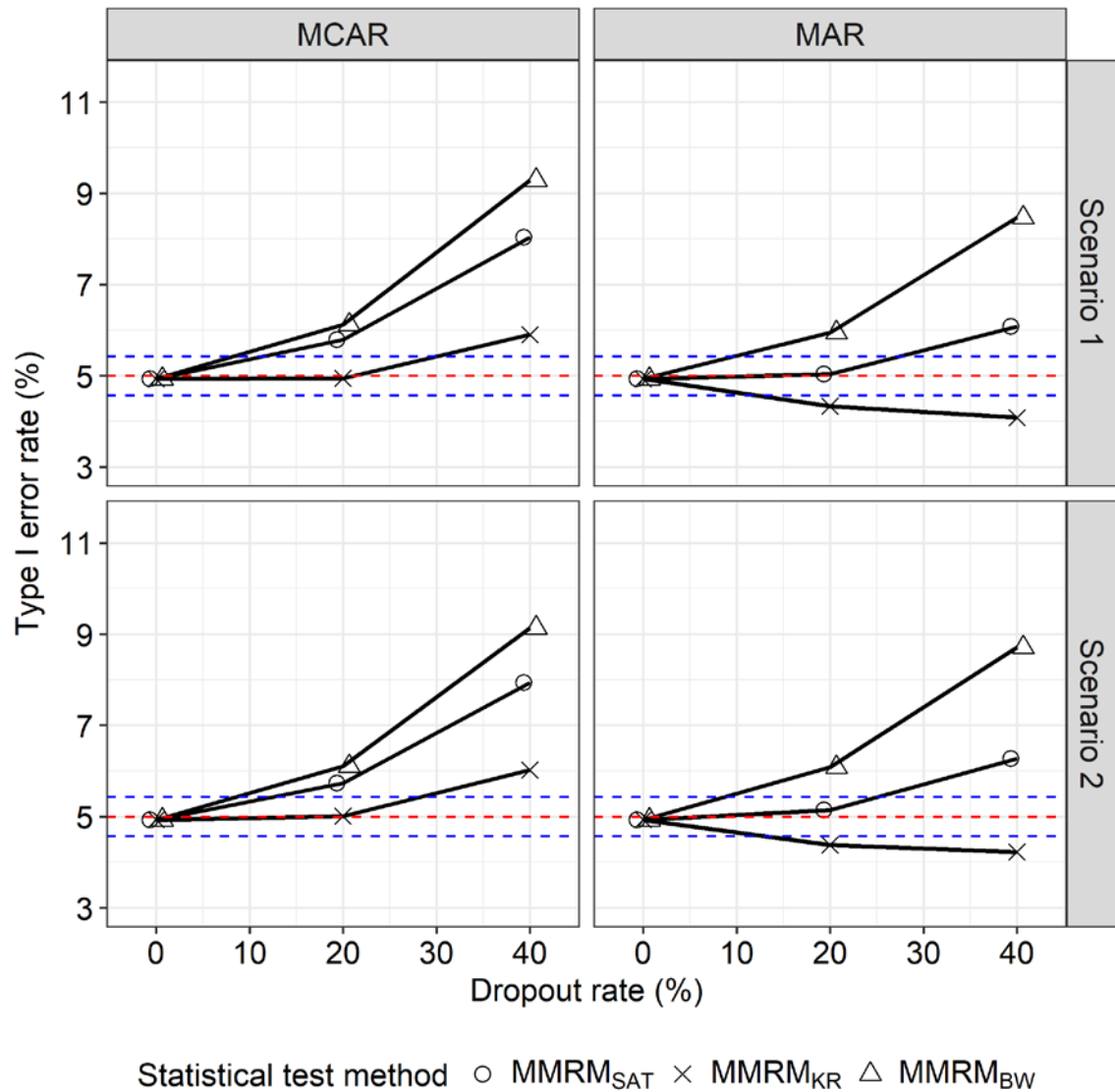


図 8.3-1 Type I error rate (%)のシミュレーション実験結果<sup>†</sup> (シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回)

赤点線, 5% ; 青点線, 95%信頼限界 (真の Type I error rate を 5%と仮定した場合の 1000 回反復によるモンテカルロ誤差) ; <sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

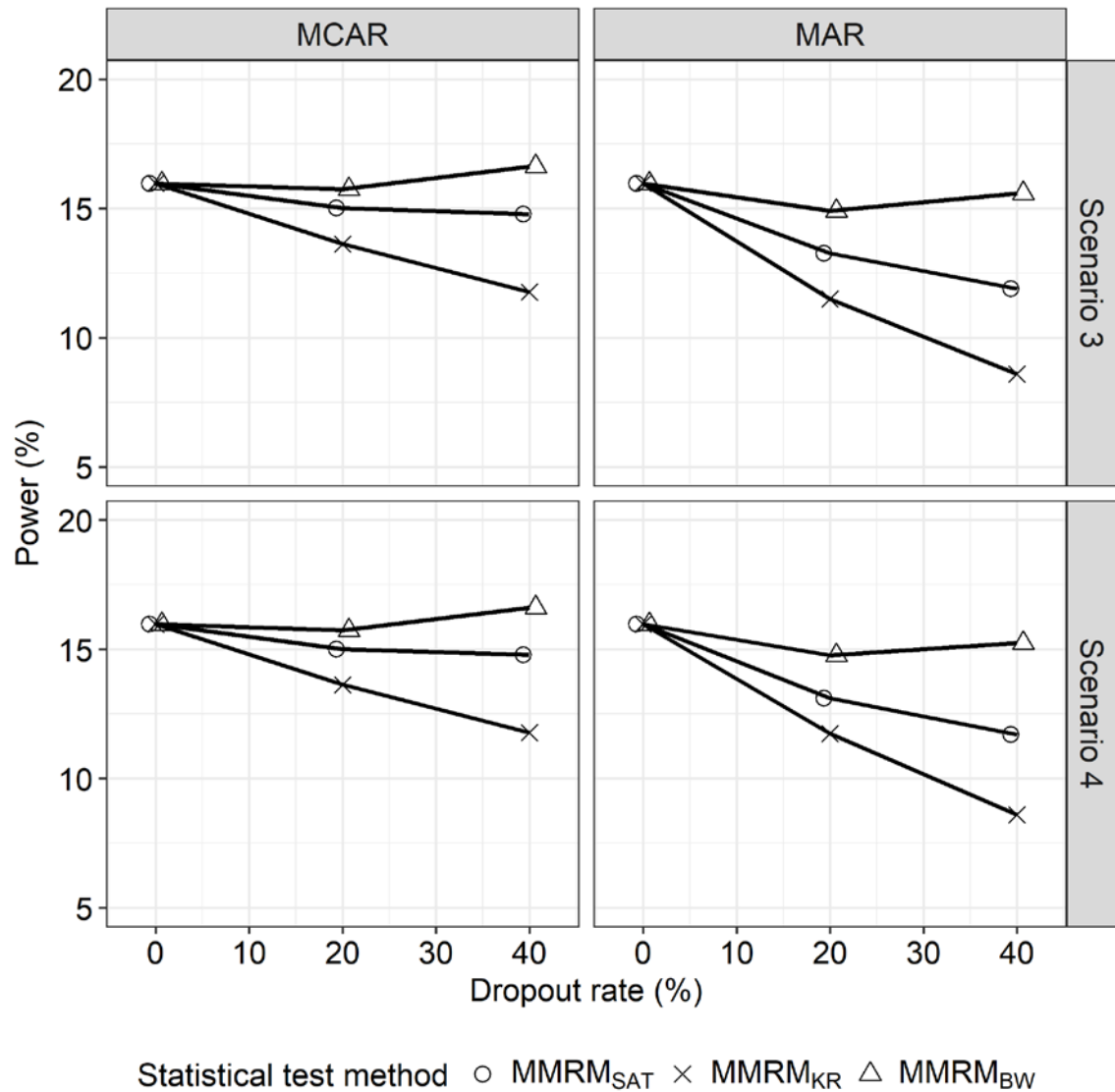


図 8.3-2 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回)

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

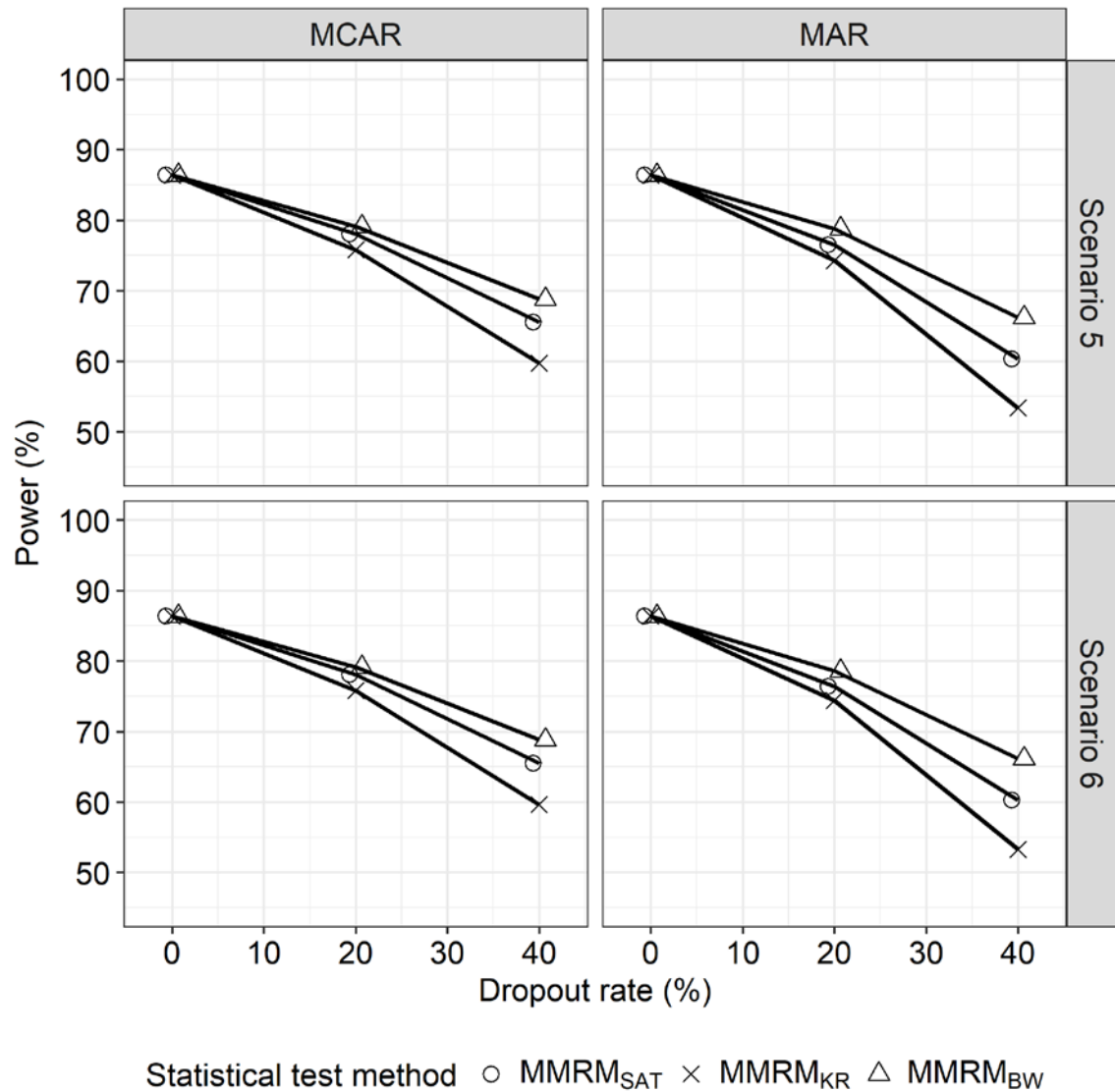


図 8.3-3 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回)

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

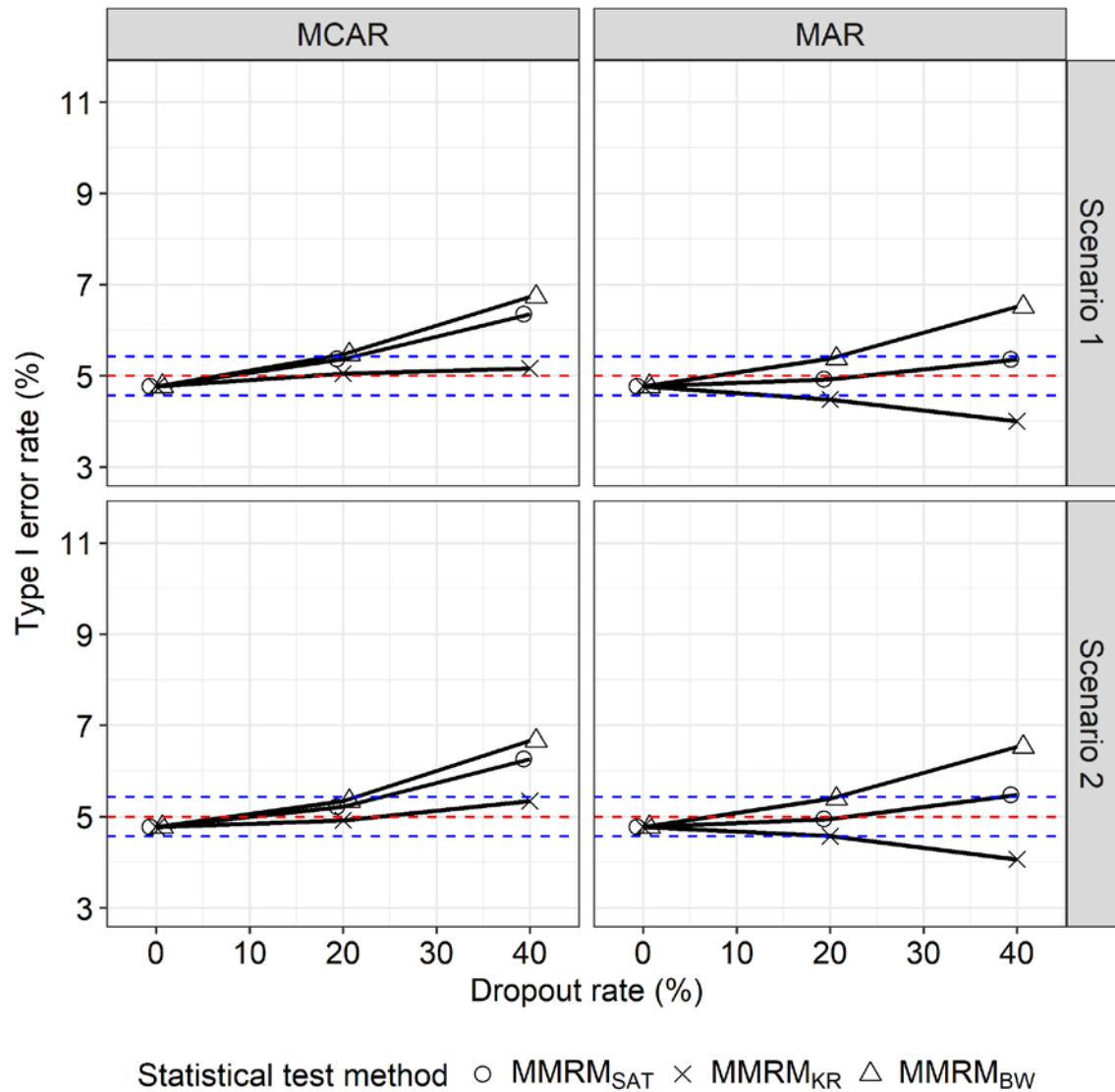


図 8.3-4 Type I error rate (%)のシミュレーション実験結果<sup>†</sup> (シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 20 例, 反復回数 10000 回)

赤点線, 5% ; 青点線, 95%信頼限界 (真の Type I error rate を 5% と仮定した場合の 1000 回反復によるモンテカルロ誤差) ; <sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定



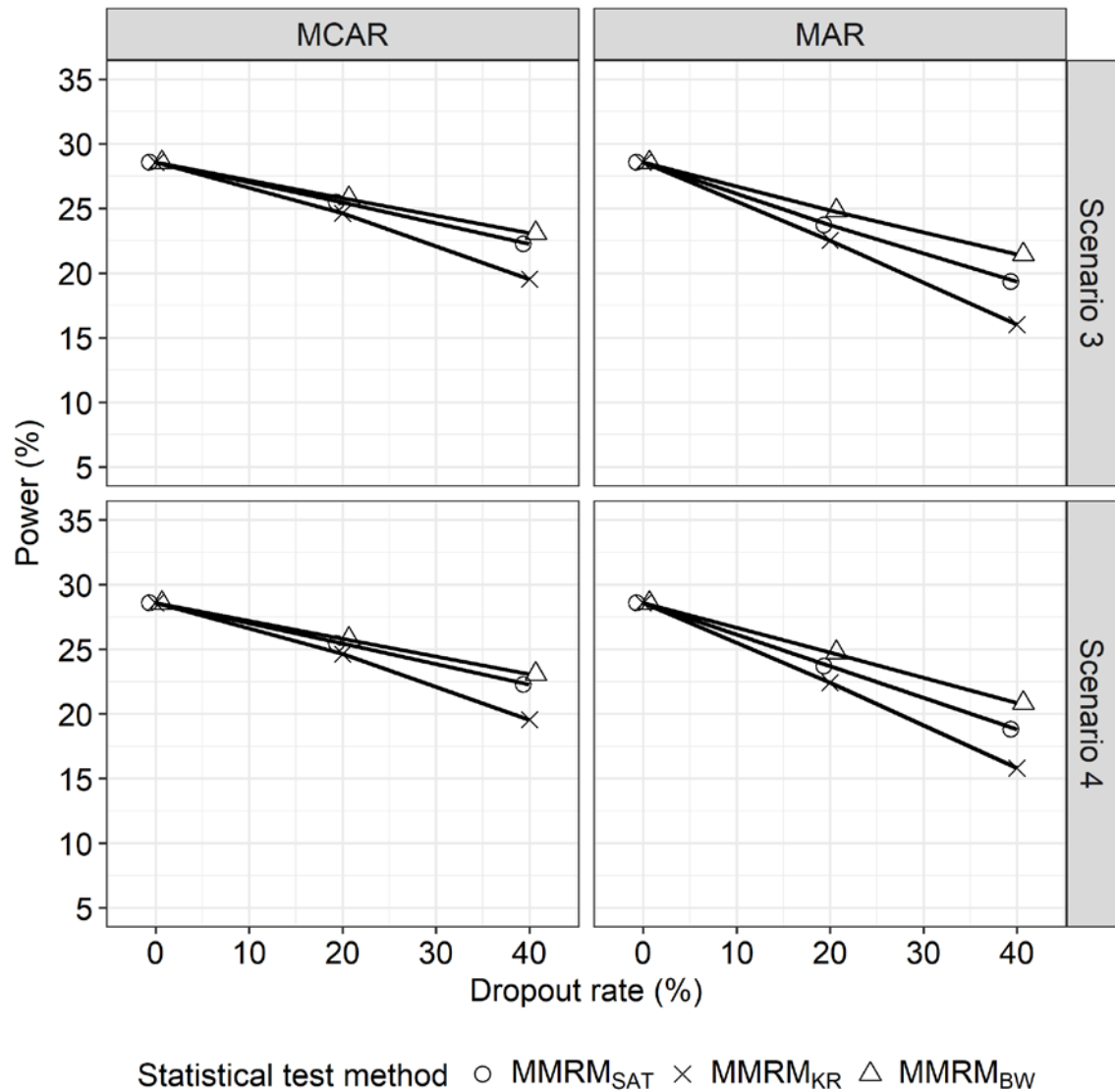


図 8.3-5 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 20 例, 反復回数 10000 回)

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

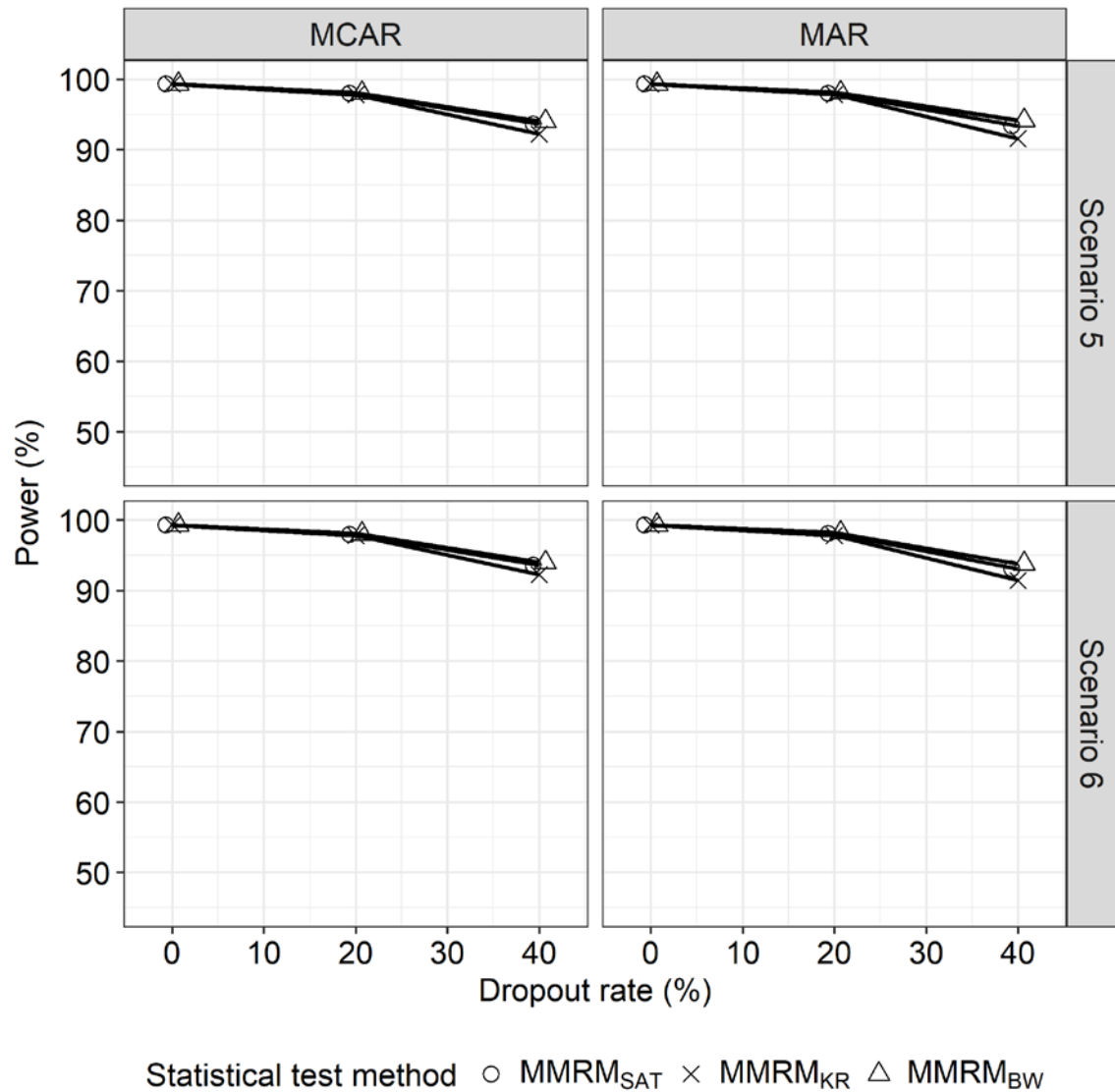


図 8.3-6 検出力 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 20 例, 反復回数 10000 回)

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

### 8.3.2 収束割合

#### 8.3.2.1 反復回数 1000 回の結果

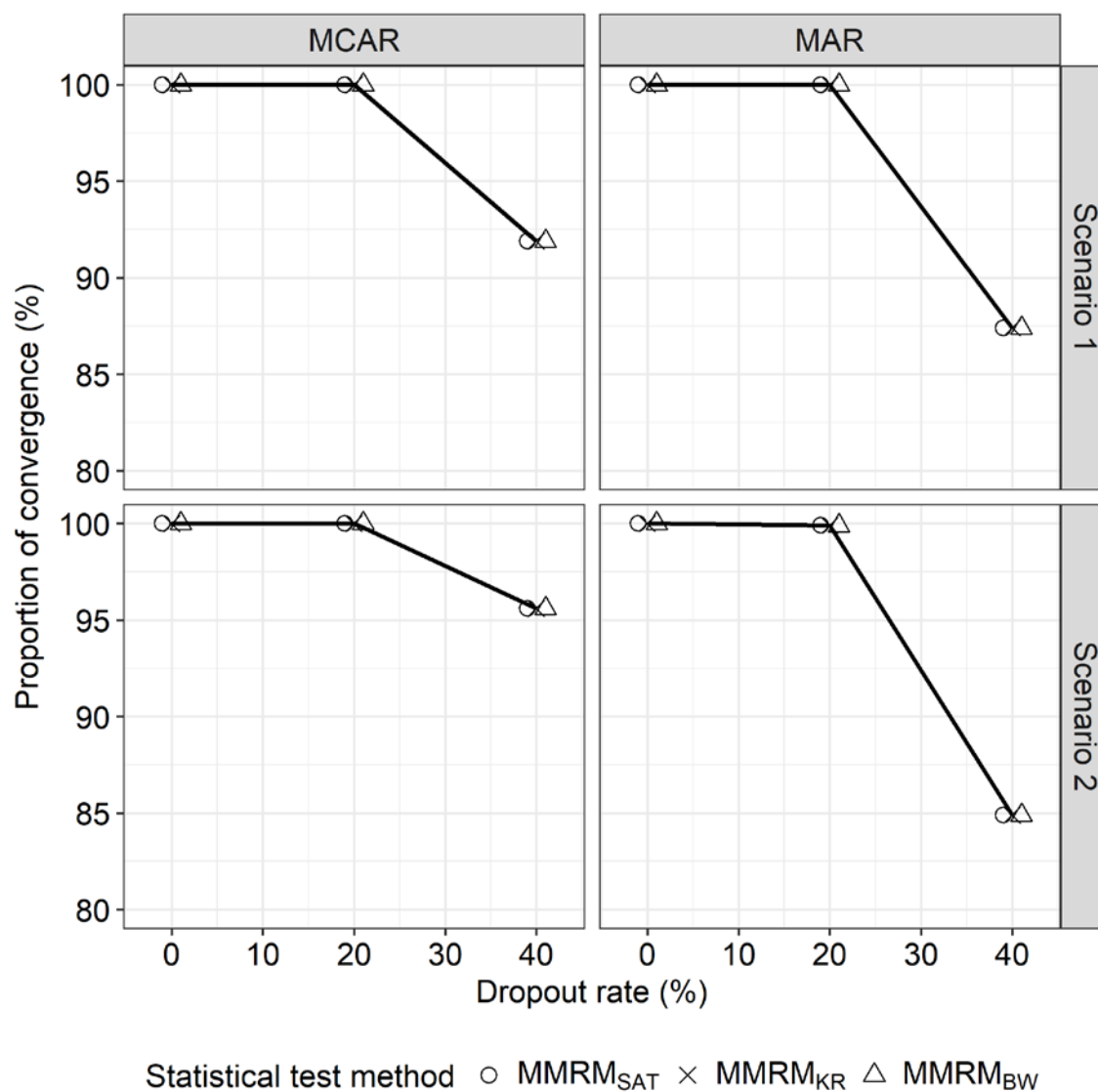


図 8.3-7 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 10 例)

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

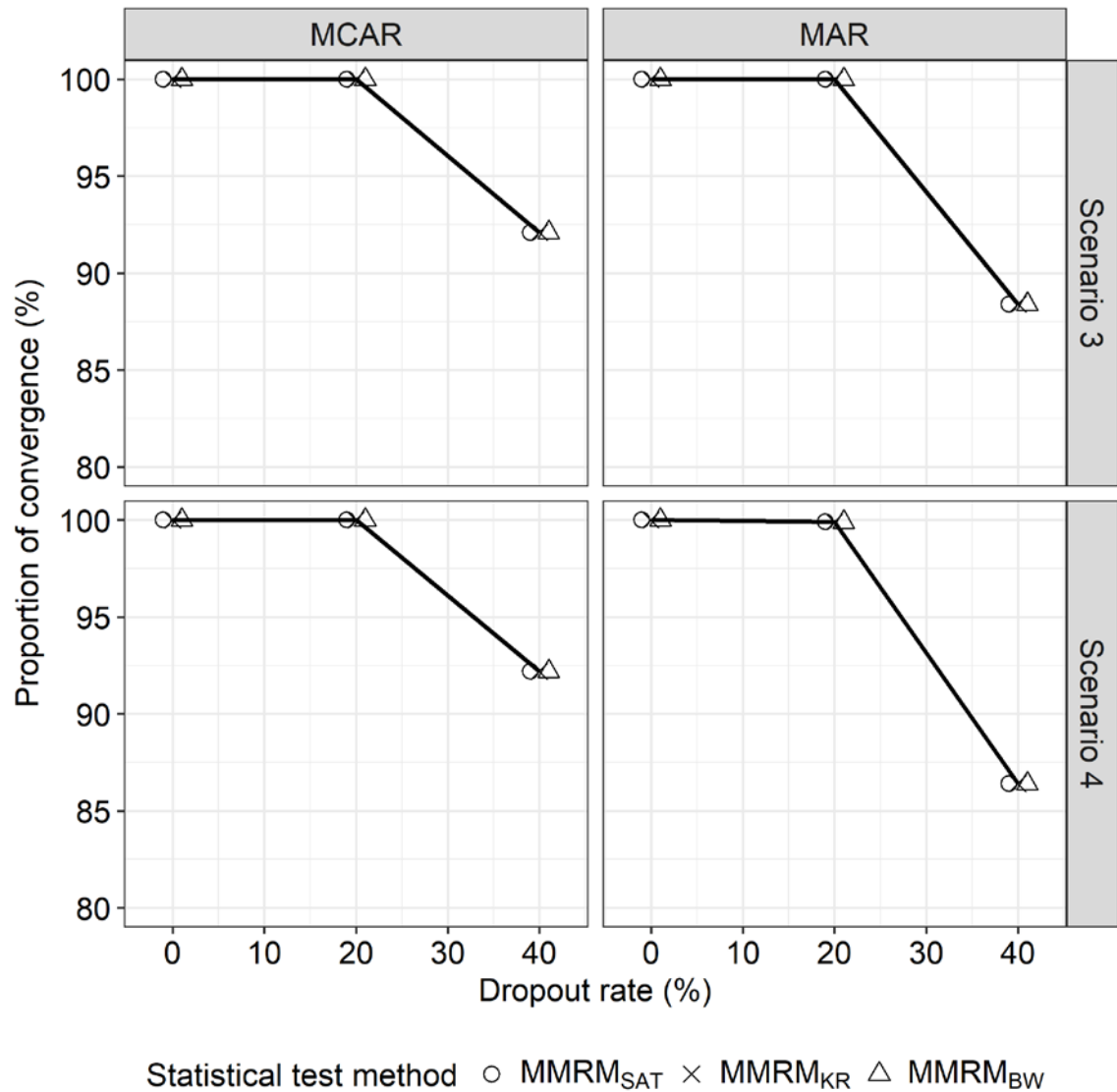


図 8.3-8 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 10 例)

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

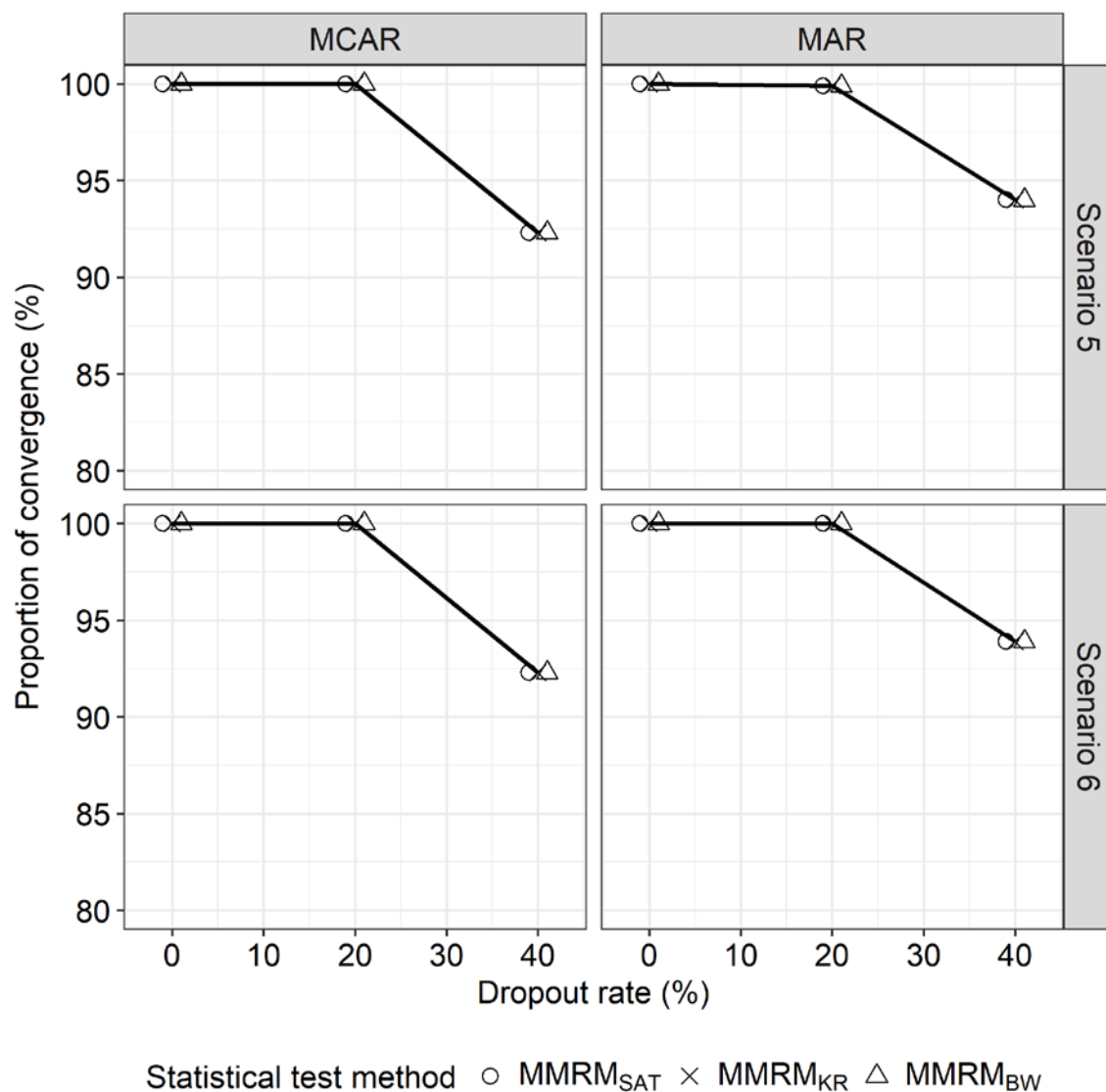


図 8.3-9 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 10 例)

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

### 8.3.2.2 反復回数 10000 回の結果

表 8.3-4 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 5 例, 反復回数 10000 回)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW</sub>	100.0	---	---	100.0	---	---
2	MMRM <sub>SAT</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---

3	MMRM <sub>BW</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>SAT</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW</sub>	100.0	---	---	100.0	---	---
4	MMRM <sub>SAT</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW</sub>	100.0	---	---	100.0	---	---
5	MMRM <sub>SAT</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW</sub>	100.0	---	---	100.0	---	---
6	MMRM <sub>SAT</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>KR</sub>	100.0	---	---	100.0	---	---
	MMRM <sub>BW</sub>	100.0	---	---	100.0	---	---

†MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

表 8.3-5 収束割合（％）のシミュレーション実験結果<sup>†</sup>（シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	100.0	100.0	92.5	100.0	99.9	88.0
	MMRM <sub>KR</sub>	100.0	100.0	92.5	100.0	99.9	88.0
	MMRM <sub>BW</sub>	100.0	100.0	92.5	100.0	99.9	88.0
2	MMRM <sub>SAT</sub>	100.0	100.0	94.2	100.0	99.8	84.8
	MMRM <sub>KR</sub>	100.0	100.0	94.2	100.0	99.8	84.8
	MMRM <sub>BW</sub>	100.0	100.0	94.2	100.0	99.8	84.8
3	MMRM <sub>SAT</sub>	100.0	100.0	92.6	100.0	99.9	87.8
	MMRM <sub>KR</sub>	100.0	100.0	92.6	100.0	99.9	87.8
	MMRM <sub>BW</sub>	100.0	100.0	92.6	100.0	99.9	87.8
4	MMRM <sub>SAT</sub>	100.0	100.0	92.6	100.0	99.9	86.5
	MMRM <sub>KR</sub>	100.0	100.0	92.6	100.0	99.9	86.5
	MMRM <sub>BW</sub>	100.0	100.0	92.6	100.0	99.9	86.5
5	MMRM <sub>SAT</sub>	100.0	100.0	92.6	100.0	100.0	94.2
	MMRM <sub>KR</sub>	100.0	100.0	92.6	100.0	100.0	94.2
	MMRM <sub>BW</sub>	100.0	100.0	92.6	100.0	100.0	94.2
6	MMRM <sub>SAT</sub>	100.0	100.0	92.6	100.0	100.0	94.1
	MMRM <sub>KR</sub>	100.0	100.0	92.6	100.0	100.0	94.1
	MMRM <sub>BW</sub>	100.0	100.0	92.6	100.0	100.0	94.1

†MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

表 8.3-6 収束割合（％）のシミュレーション実験結果<sup>†</sup>（シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 20 例, 反復回数 10000 回）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>BW</sub>	100.0	100.0	100.0	100.0	100.0	100.0
2	MMRM <sub>SAT</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0

3	MMRM <sub>BW</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>SAT</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>BW</sub>	100.0	100.0	100.0	100.0	100.0	100.0
4	MMRM <sub>SAT</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>BW</sub>	100.0	100.0	100.0	100.0	100.0	100.0
5	MMRM <sub>SAT</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>BW</sub>	100.0	100.0	100.0	100.0	100.0	100.0
6	MMRM <sub>SAT</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>KR</sub>	100.0	100.0	100.0	100.0	100.0	100.0
	MMRM <sub>BW</sub>	100.0	100.0	100.0	100.0	100.0	100.0

†MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

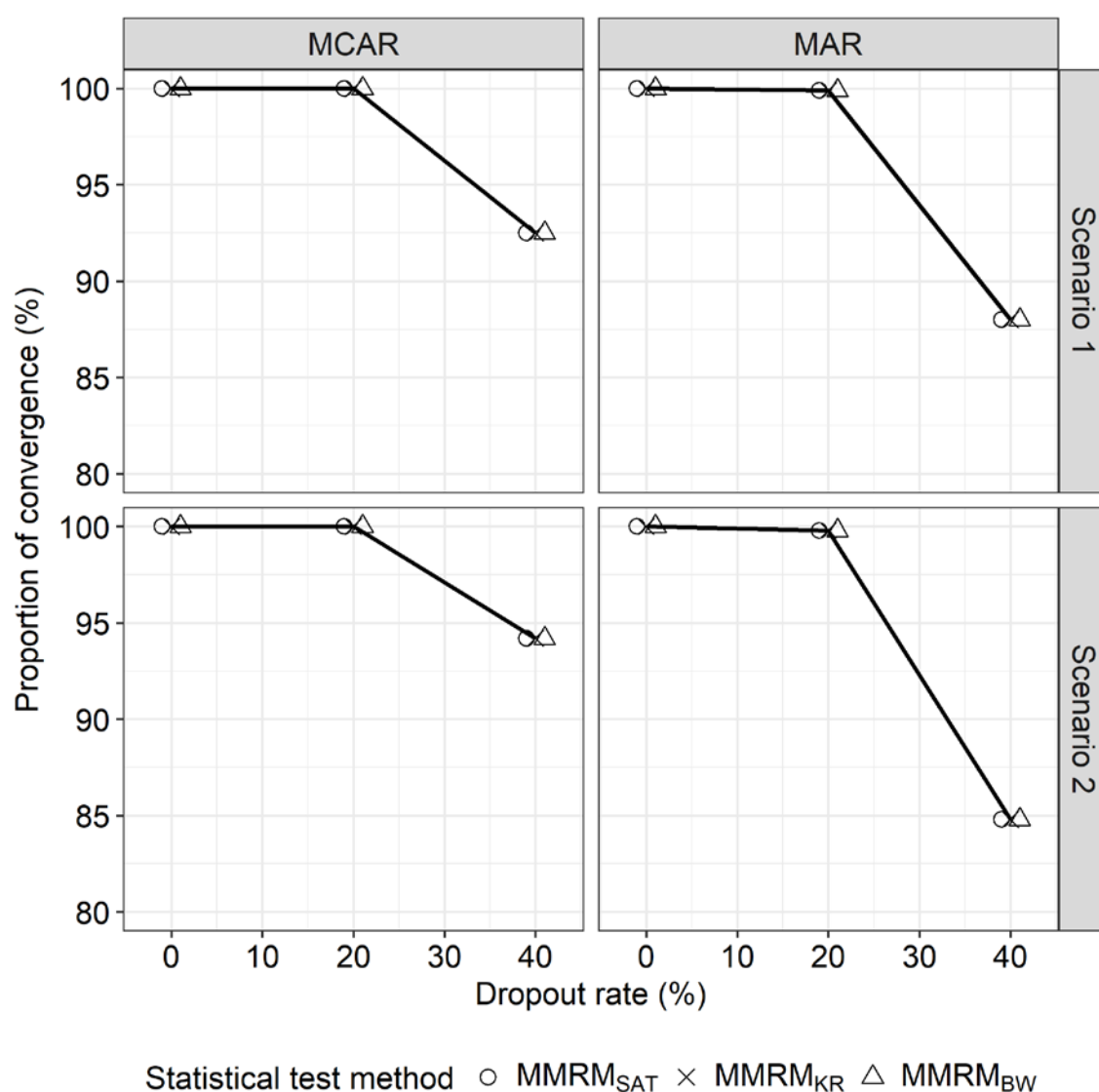


図 8.3-10 収束割合 (%) のシミュレーション実験結果† (シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回)

†MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用い

た Wald 型検定

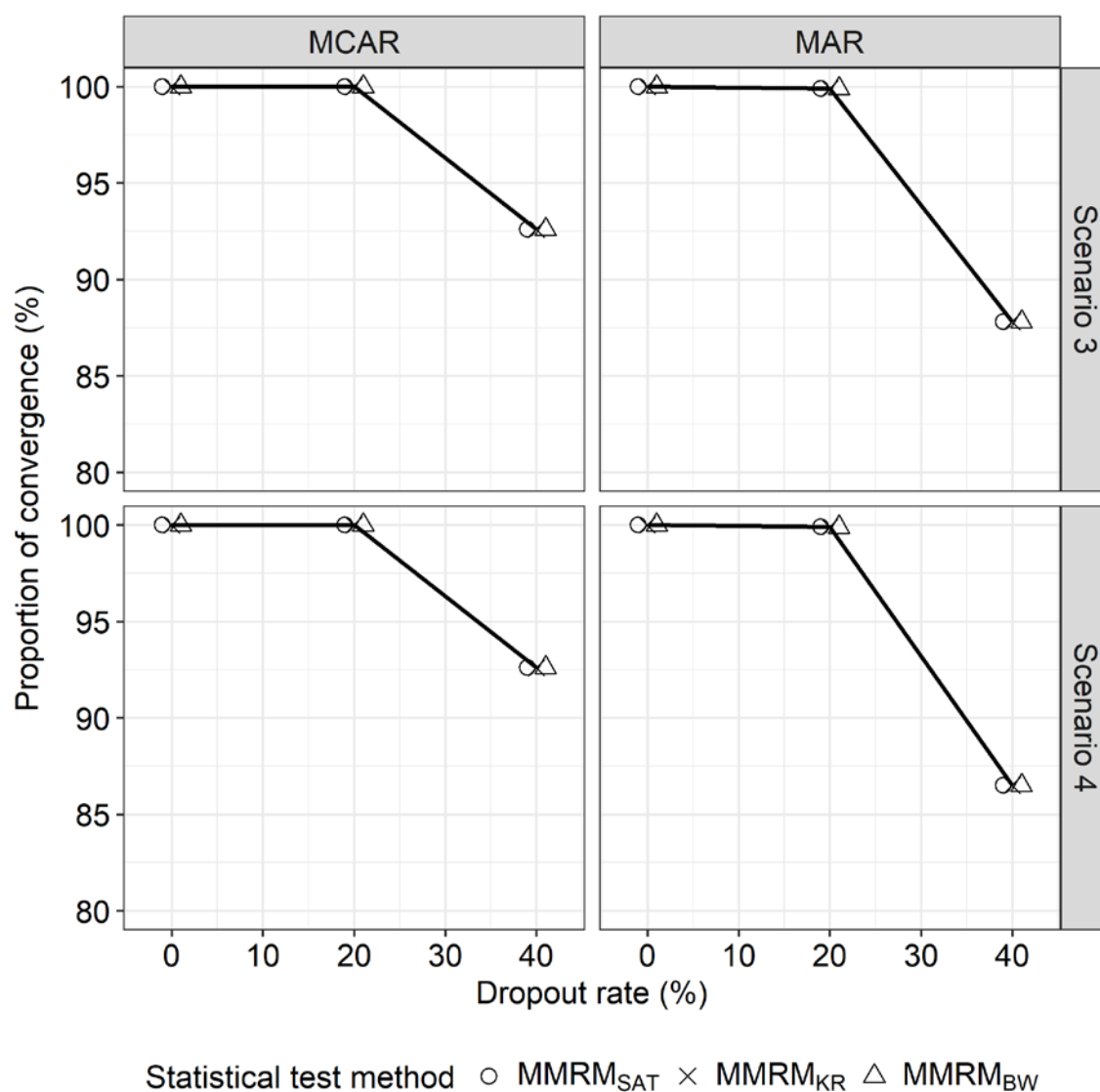
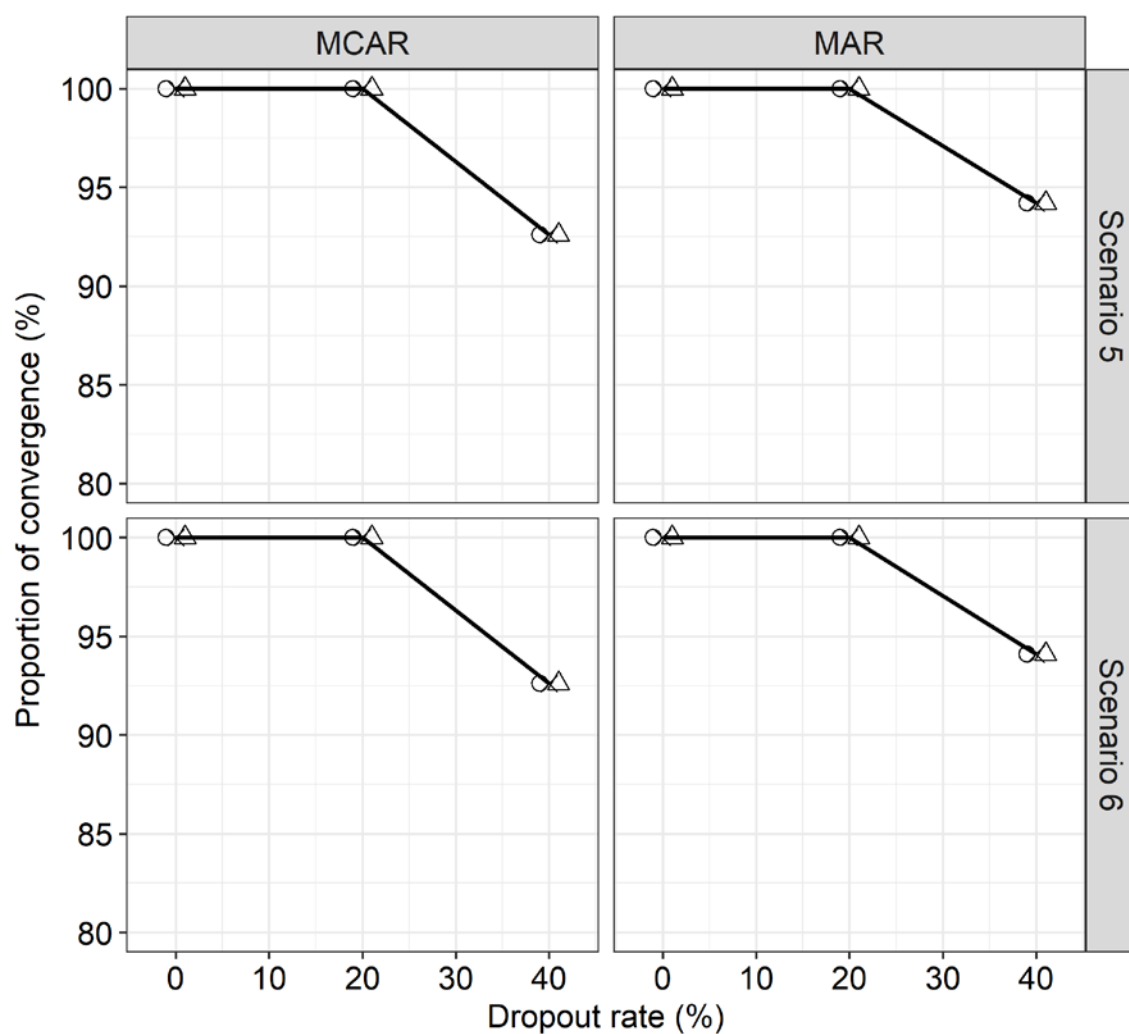


図 8.3-11 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回)

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

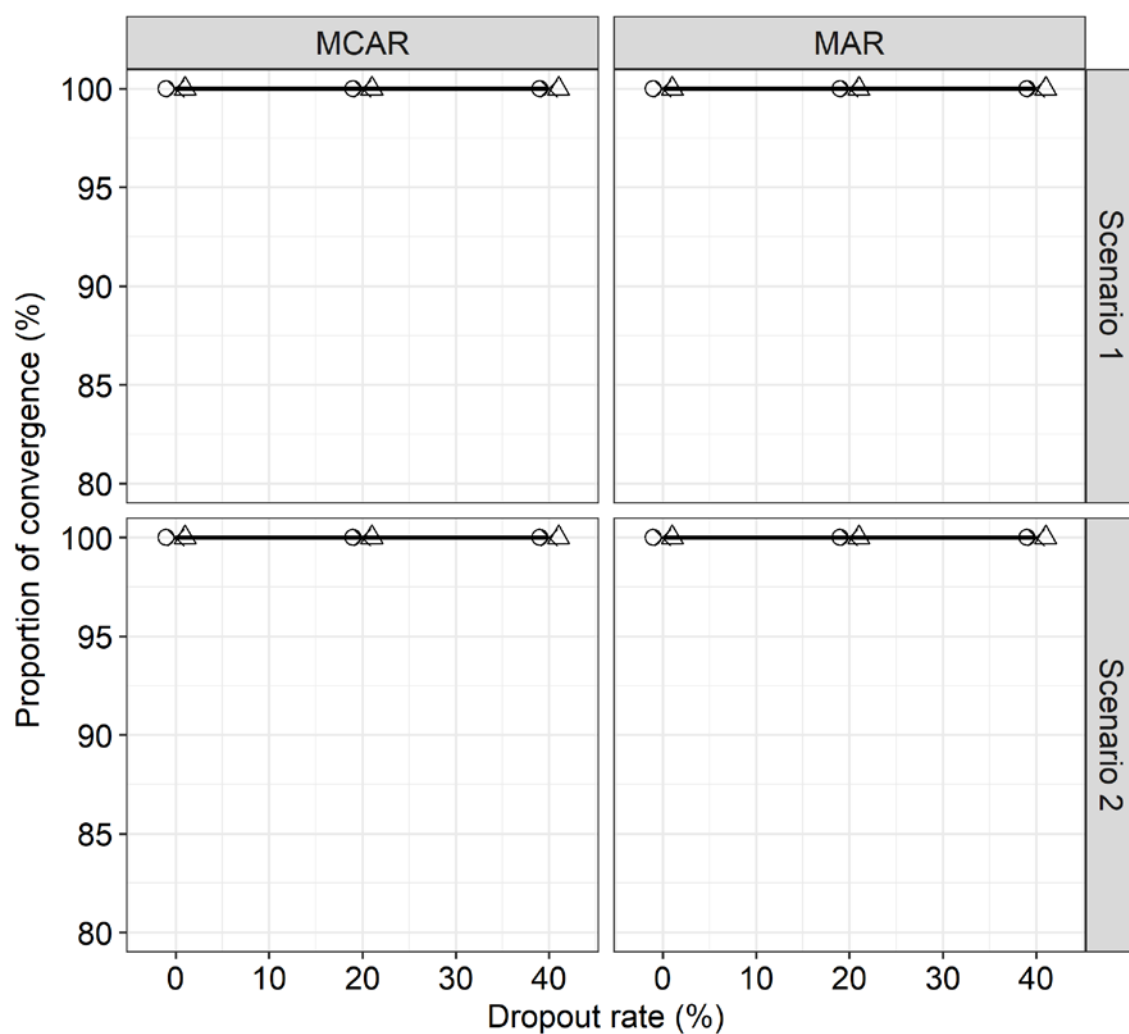




Statistical test method ○ MMRM<sub>SAT</sub> × MMRM<sub>KR</sub> △ MMRM<sub>BW</sub>

図 8.3-12 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回)

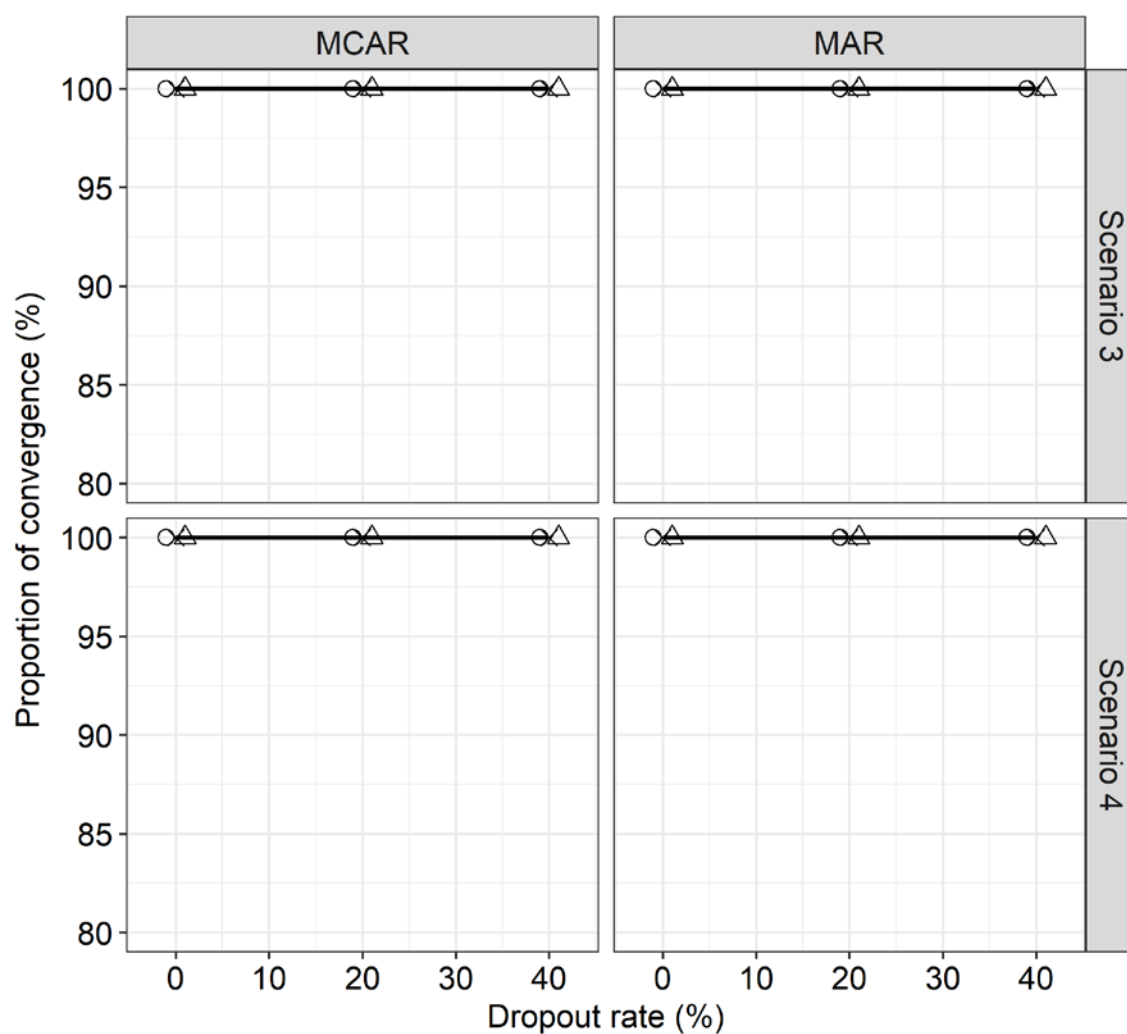
<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定



Statistical test method ○ MMRM<sub>SAT</sub> × MMRM<sub>KR</sub> △ MMRM<sub>BW</sub>

図 8.3-13 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 20 例, 反復回数 10000 回)

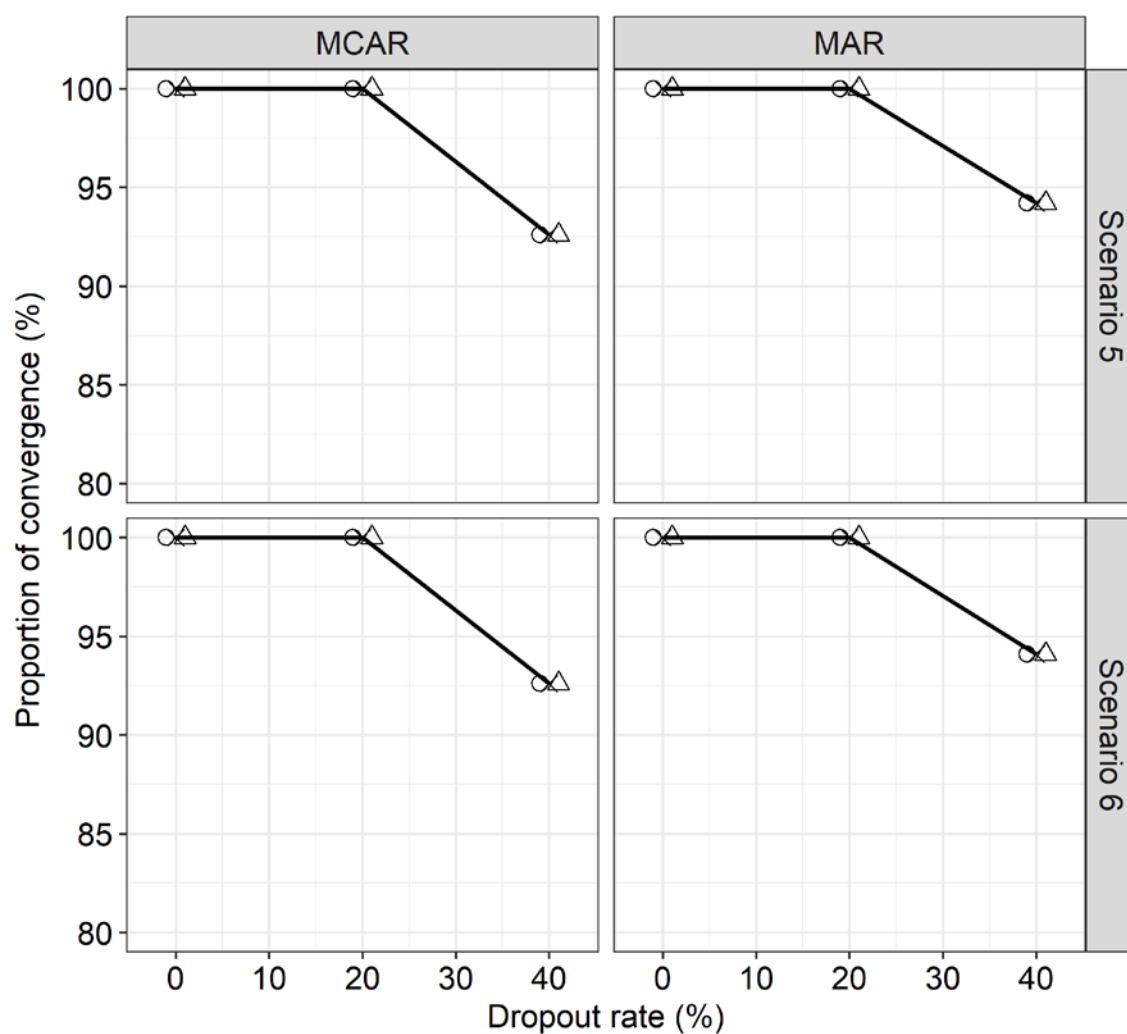
<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定



Statistical test method ○ MMRM<sub>SAT</sub> × MMRM<sub>KR</sub> △ MMRM<sub>BW</sub>

図 8.3-14 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回)

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定



Statistical test method ○ MMRM<sub>SAT</sub> × MMRM<sub>KR</sub> △ MMRM<sub>BW</sub>

図 8.3-15 収束割合 (%) のシミュレーション実験結果<sup>†</sup> (シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 20 例, 反復回数 10000 回)

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

### 8.3.3 (固定効果の) 標準誤差

表 8.3-7 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup> (シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 5 例, 反復回数 10000 回)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	4.10	---	---	4.10	---	---
	MMRM <sub>KR</sub>	4.16	---	---	4.16	---	---
	MMRM <sub>BW</sub>	4.10	---	---	4.10	---	---
2	MMRM <sub>SAT</sub>	4.10	---	---	4.10	---	---
	MMRM <sub>KR</sub>	4.16	---	---	4.16	---	---
	MMRM <sub>BW</sub>	4.10	---	---	4.10	---	---
3	MMRM <sub>SAT</sub>	4.10	---	---	4.10	---	---
	MMRM <sub>KR</sub>	4.15	---	---	4.15	---	---
	MMRM <sub>BW</sub>	4.10	---	---	4.10	---	---
4	MMRM <sub>SAT</sub>	4.10	---	---	4.10	---	---
	MMRM <sub>KR</sub>	4.15	---	---	4.15	---	---
	MMRM <sub>BW</sub>	4.10	---	---	4.10	---	---
5	MMRM <sub>SAT</sub>	4.10	---	---	4.10	---	---
	MMRM <sub>KR</sub>	4.16	---	---	4.16	---	---
	MMRM <sub>BW</sub>	4.10	---	---	4.10	---	---
6	MMRM <sub>SAT</sub>	4.10	---	---	4.10	---	---
	MMRM <sub>KR</sub>	4.16	---	---	4.16	---	---
	MMRM <sub>BW</sub>	4.10	---	---	4.10	---	---

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

表 8.3-8 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup> (シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回)

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	2.95	3.22	3.61	2.95	3.32	3.84
	MMRM <sub>KR</sub>	2.95	3.33	3.95	2.95	3.47	4.36
	MMRM <sub>BW</sub>	2.95	3.22	3.61	2.95	3.32	3.84
2	MMRM <sub>SAT</sub>	2.95	3.21	3.58	2.95	3.34	3.87
	MMRM <sub>KR</sub>	2.95	3.32	3.91	2.95	3.50	4.38
	MMRM <sub>BW</sub>	2.95	3.21	3.58	2.95	3.34	3.87
3	MMRM <sub>SAT</sub>	2.95	3.21	3.62	2.95	3.31	3.85
	MMRM <sub>KR</sub>	2.95	3.33	3.98	2.95	3.46	4.35
	MMRM <sub>BW</sub>	2.95	3.21	3.62	2.95	3.31	3.85
4	MMRM <sub>SAT</sub>	2.95	3.21	3.62	2.95	3.32	3.87
	MMRM <sub>KR</sub>	2.95	3.33	3.98	2.95	3.47	4.38
	MMRM <sub>BW</sub>	2.95	3.21	3.62	2.95	3.32	3.87
5	MMRM <sub>SAT</sub>	2.95	3.21	3.62	2.95	3.26	3.77
	MMRM <sub>KR</sub>	2.95	3.33	3.99	2.95	3.39	4.21
	MMRM <sub>BW</sub>	2.95	3.21	3.62	2.95	3.26	3.77
6	MMRM <sub>SAT</sub>	2.95	3.21	3.62	2.95	3.26	3.77
	MMRM <sub>KR</sub>	2.95	3.33	3.98	2.95	3.38	4.19
	MMRM <sub>BW</sub>	2.95	3.21	3.62	2.95	3.26	3.77

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

表 8.3-9 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 20 例, 反復回数 10000 回）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	2.11	2.28	2.53	2.11	2.32	2.66
	MMRM <sub>KR</sub>	2.11	2.32	2.65	2.11	2.37	2.84
	MMRM <sub>BW</sub>	2.11	2.28	2.53	2.11	2.32	2.66
2	MMRM <sub>SAT</sub>	2.11	2.28	2.51	2.11	2.34	2.70
	MMRM <sub>KR</sub>	2.11	2.32	2.63	2.11	2.39	2.89
	MMRM <sub>BW</sub>	2.11	2.28	2.51	2.11	2.34	2.70
3	MMRM <sub>SAT</sub>	2.11	2.28	2.54	2.11	2.32	2.65
	MMRM <sub>KR</sub>	2.11	2.32	2.67	2.11	2.37	2.83
	MMRM <sub>BW</sub>	2.11	2.28	2.54	2.11	2.32	2.65
4	MMRM <sub>SAT</sub>	2.11	2.28	2.54	2.11	2.32	2.68
	MMRM <sub>KR</sub>	2.11	2.32	2.67	2.11	2.37	2.86
	MMRM <sub>BW</sub>	2.11	2.28	2.54	2.11	2.32	2.68
5	MMRM <sub>SAT</sub>	2.11	2.28	2.54	2.11	2.30	2.62
	MMRM <sub>KR</sub>	2.11	2.32	2.67	2.11	2.34	2.76
	MMRM <sub>BW</sub>	2.11	2.28	2.54	2.11	2.30	2.62
6	MMRM <sub>SAT</sub>	2.11	2.28	2.54	2.11	2.30	2.62
	MMRM <sub>KR</sub>	2.11	2.32	2.67	2.11	2.34	2.77
	MMRM <sub>BW</sub>	2.11	2.28	2.54	2.11	2.30	2.62

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

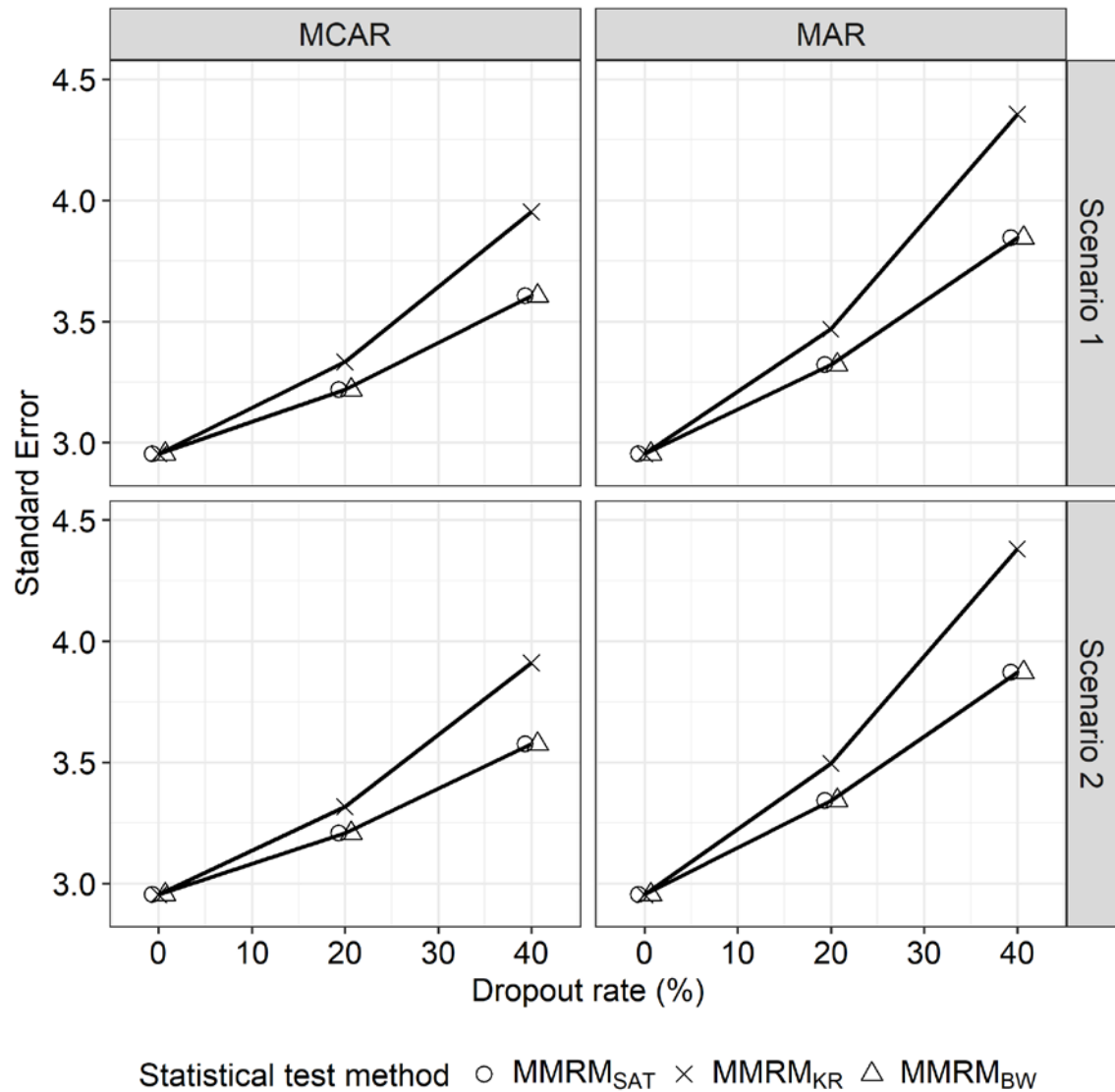


図 8.3-16 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

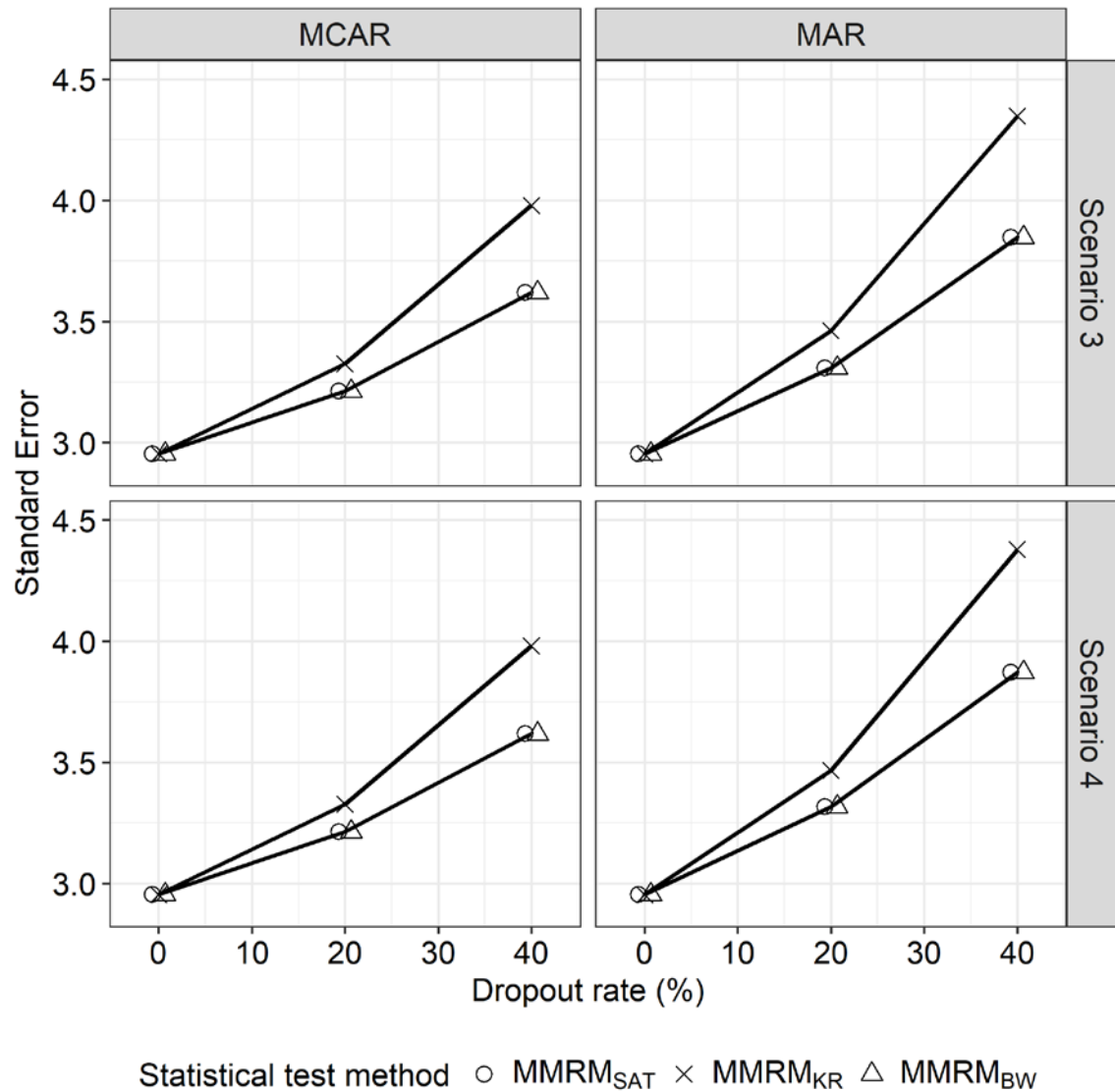


図 8.3-17 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定



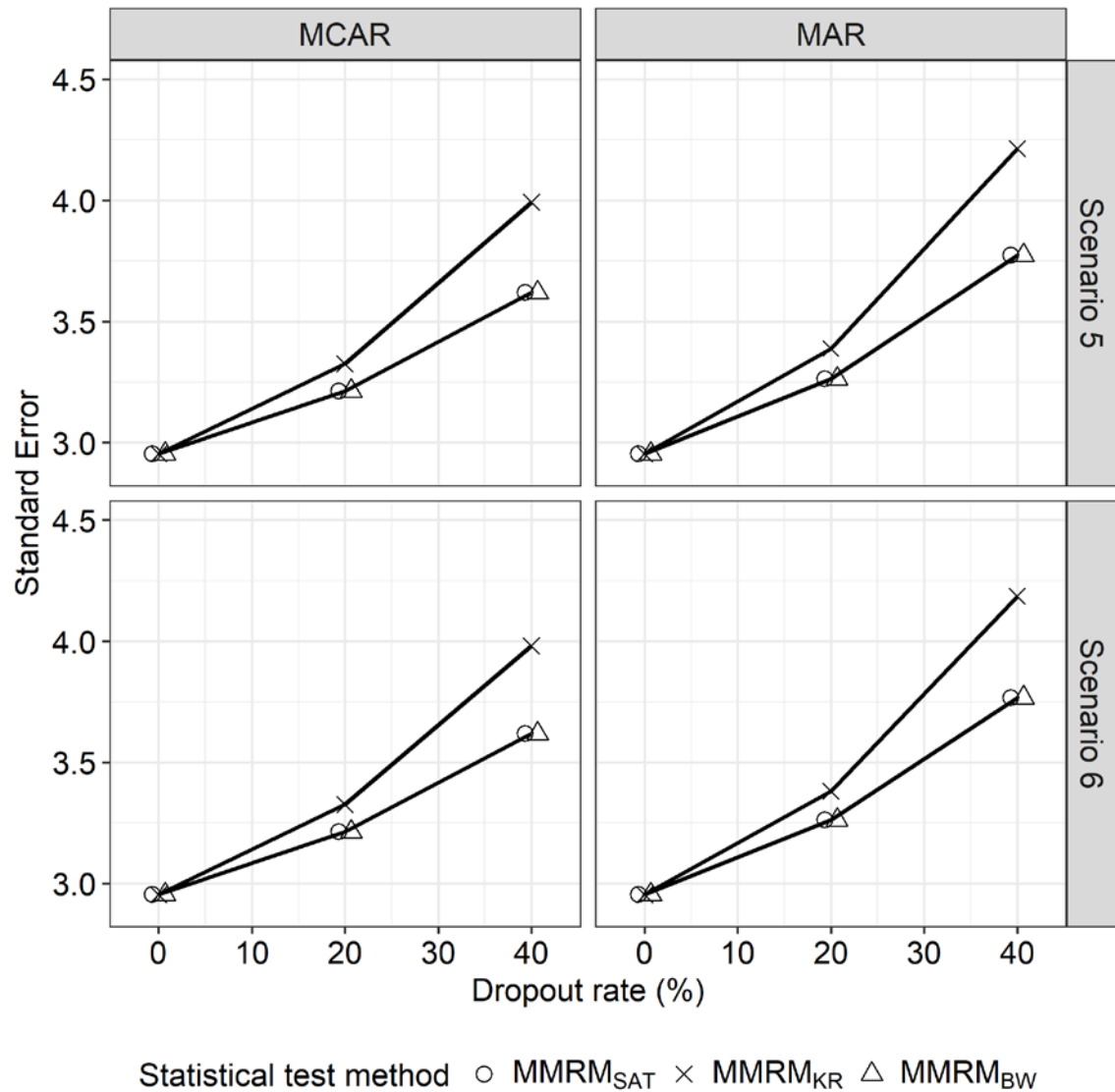


図 8.3-18 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

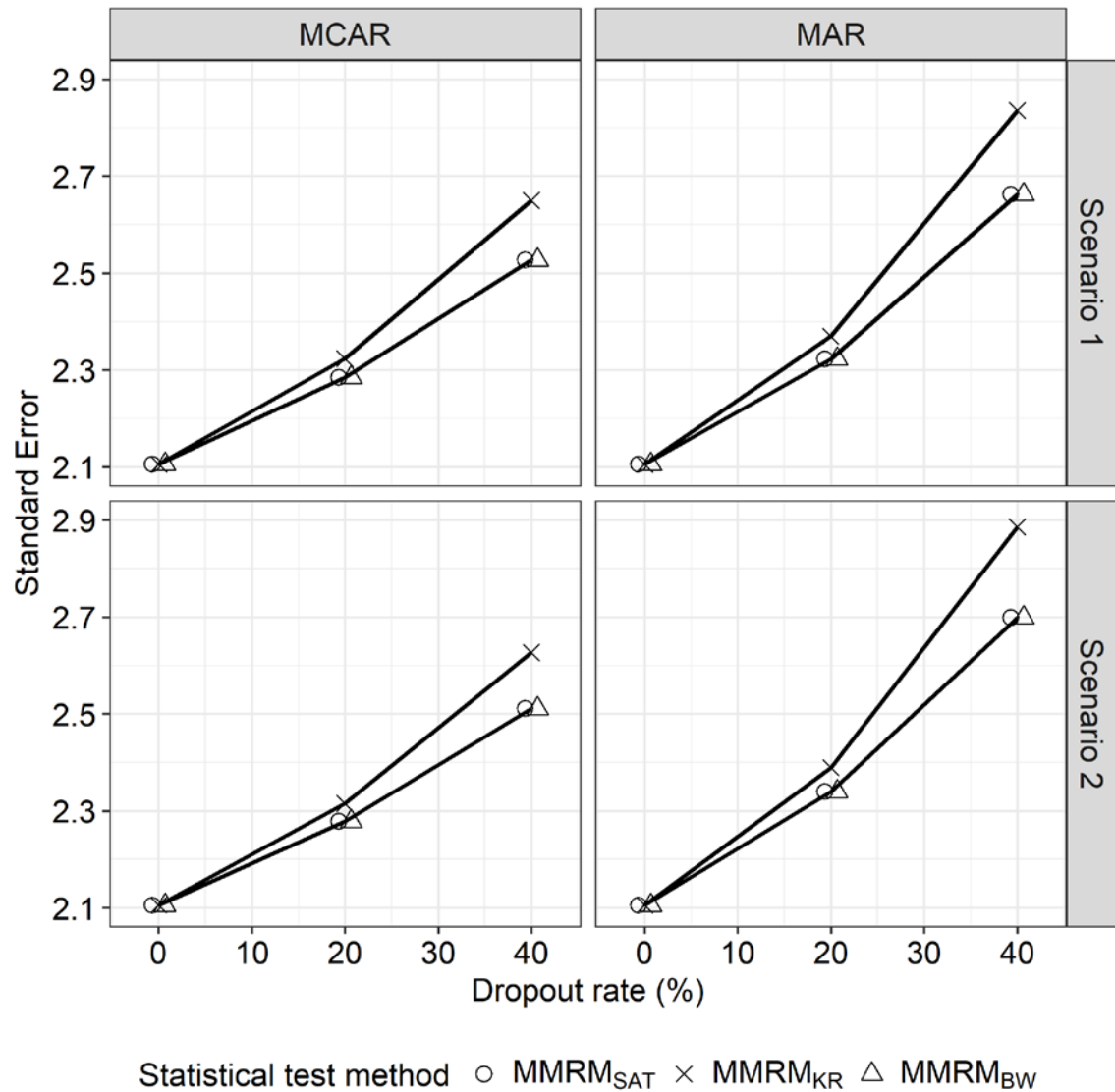


図 8.3-19 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 20 例, 反復回数 10000 回）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

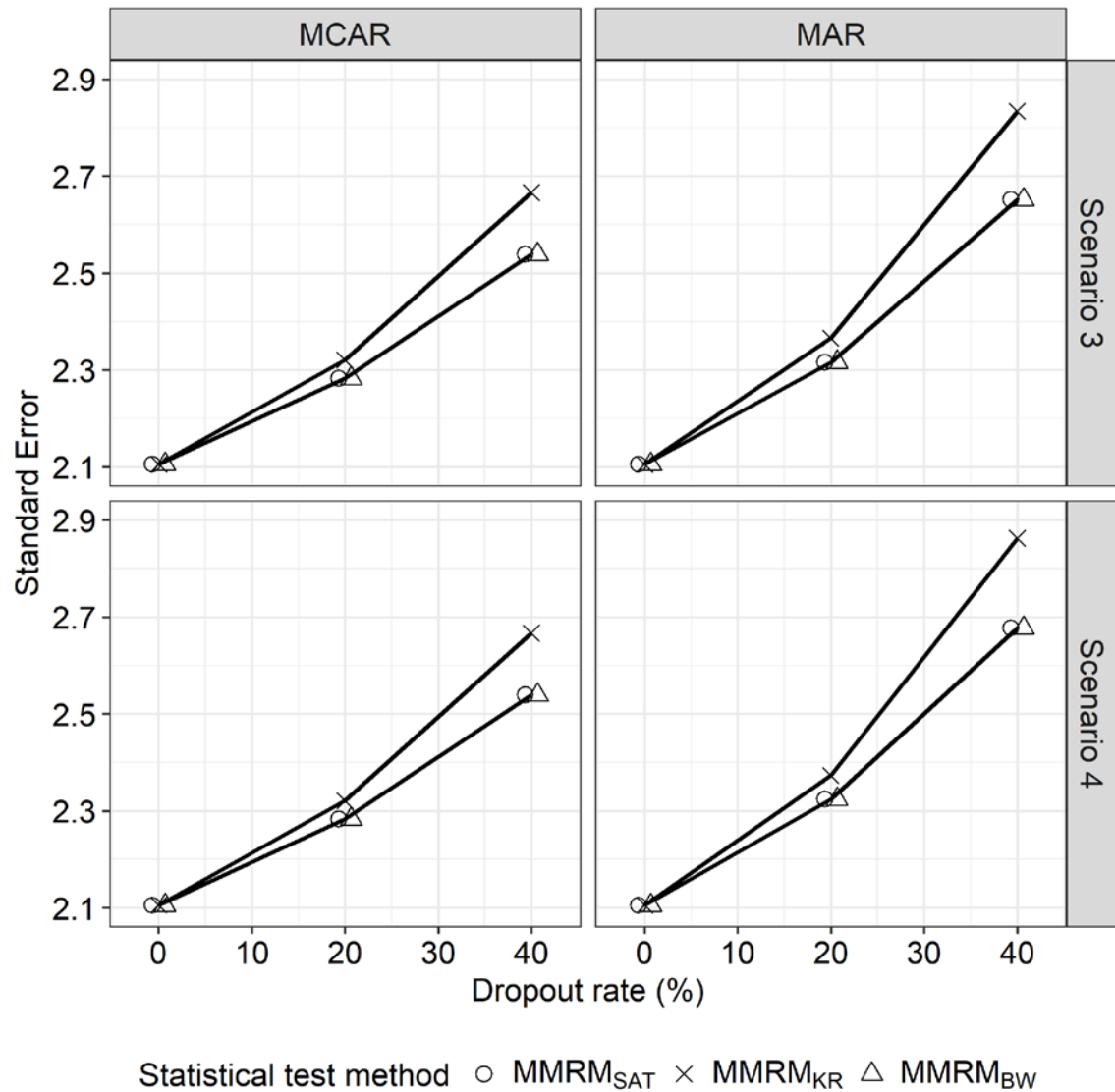


図 8.3-20 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 20 例, 反復回数 10000 回）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

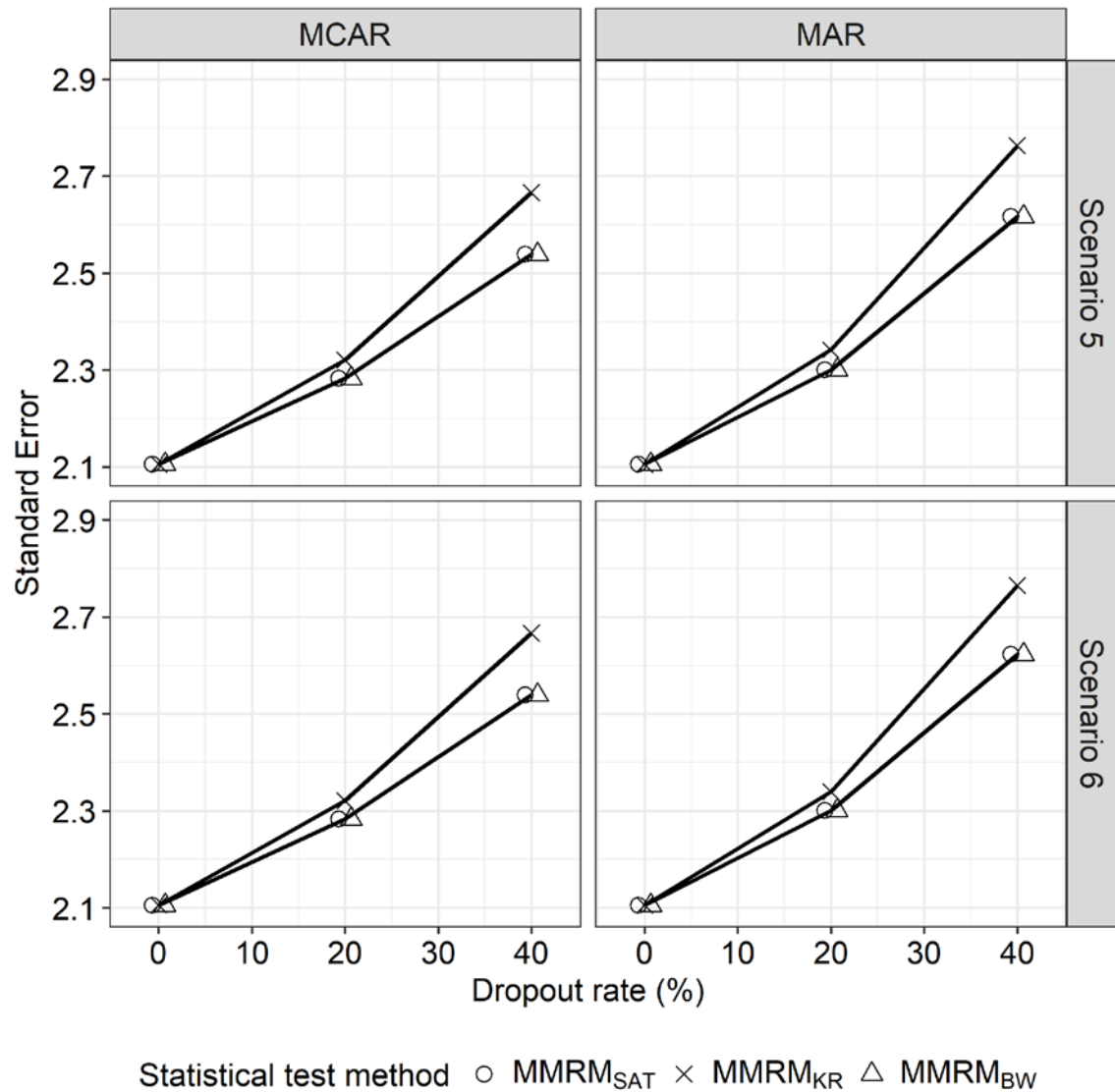


図 8.3-21 固定効果の標準誤差のシミュレーション実験結果<sup>†</sup>（シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 20 例, 反復回数 10000 回）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

### 8.3.4 ( $t$ 検定の) 自由度

#### 8.3.4.1 反復回数 1000 回の結果

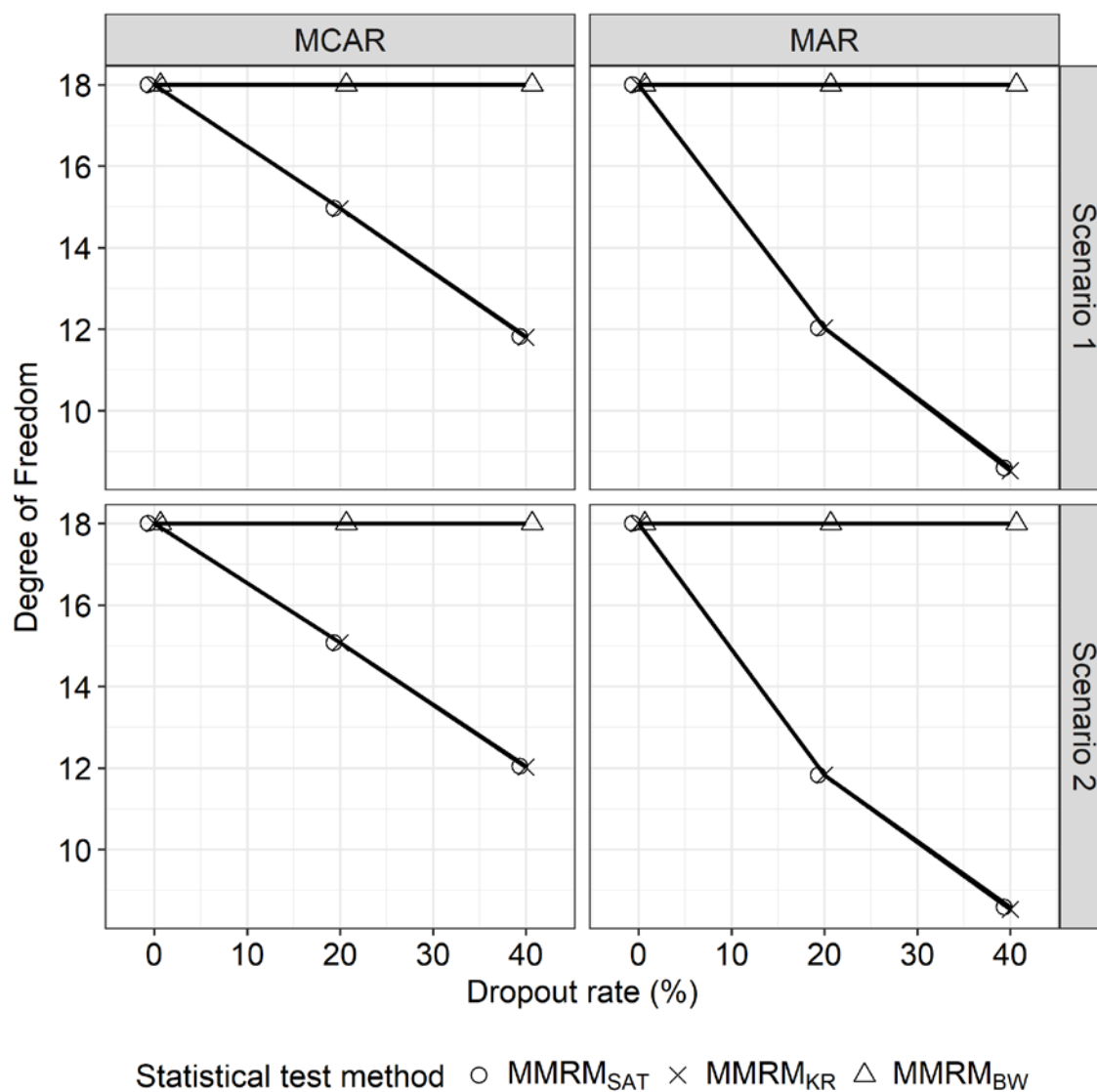


図 8.3-22  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup> (シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 10 例)

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定 ; MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

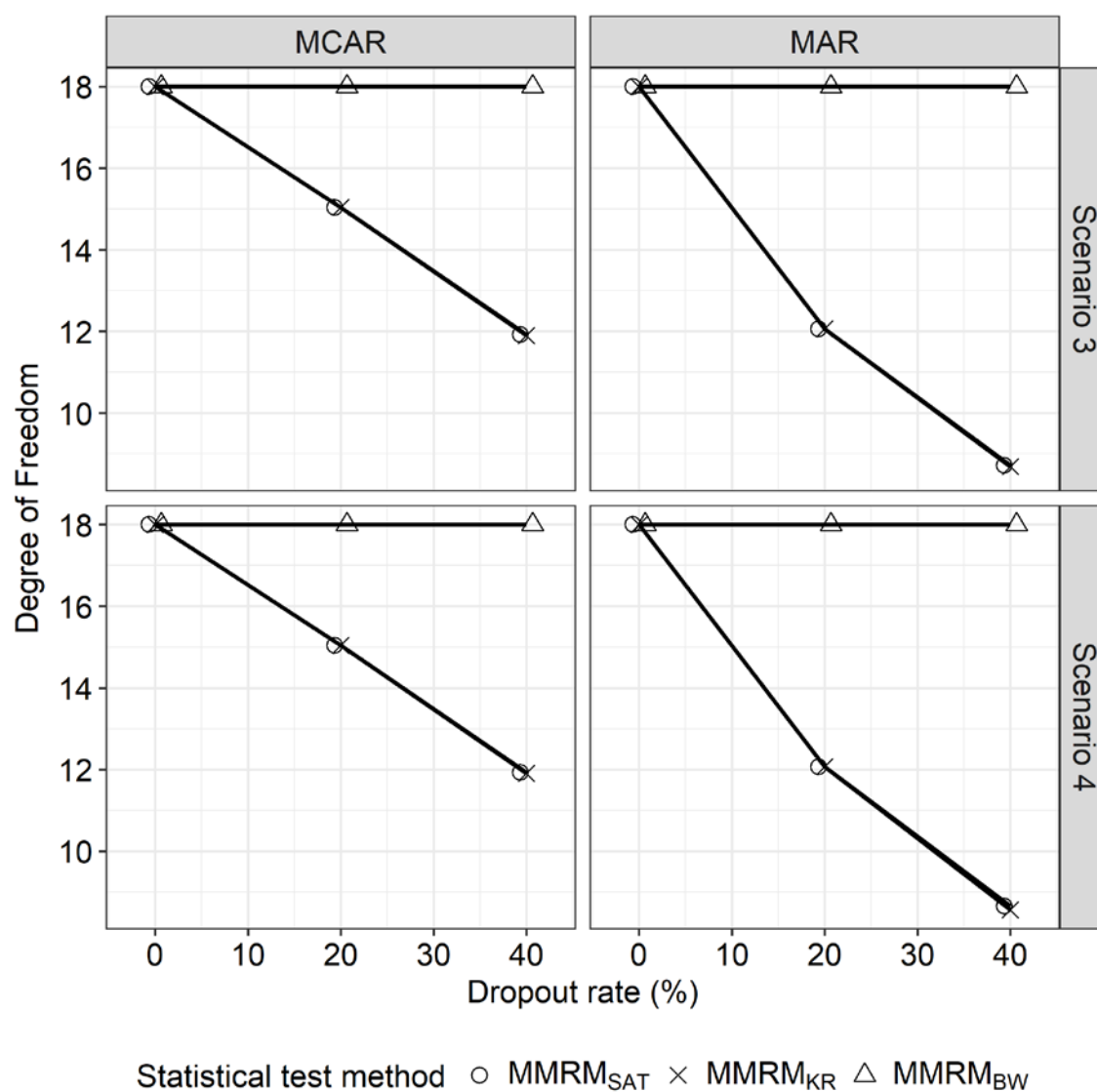


図 8.3-23  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup>（シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 10 例）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

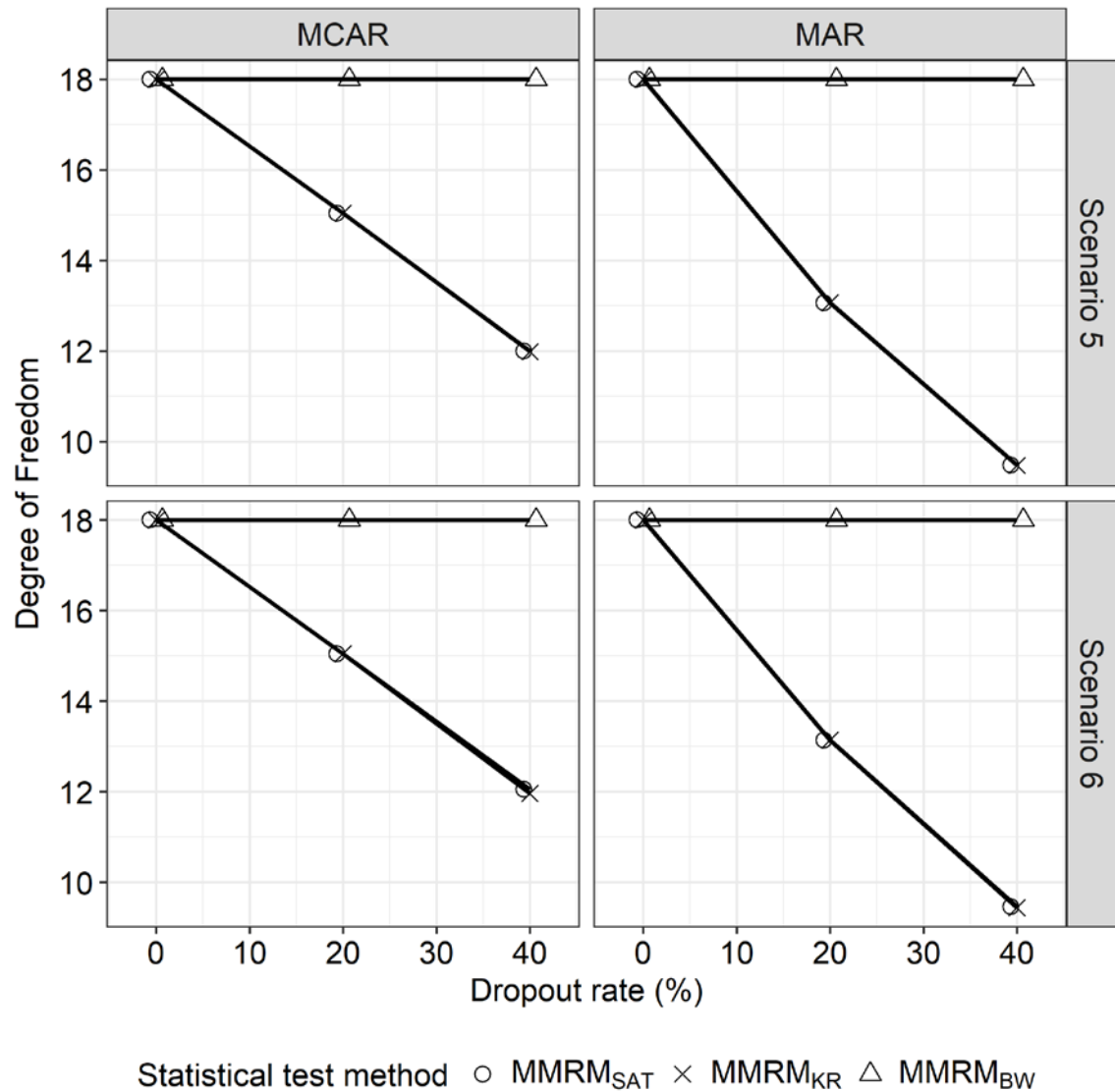


図 8.3-24  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup>（シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 10 例）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

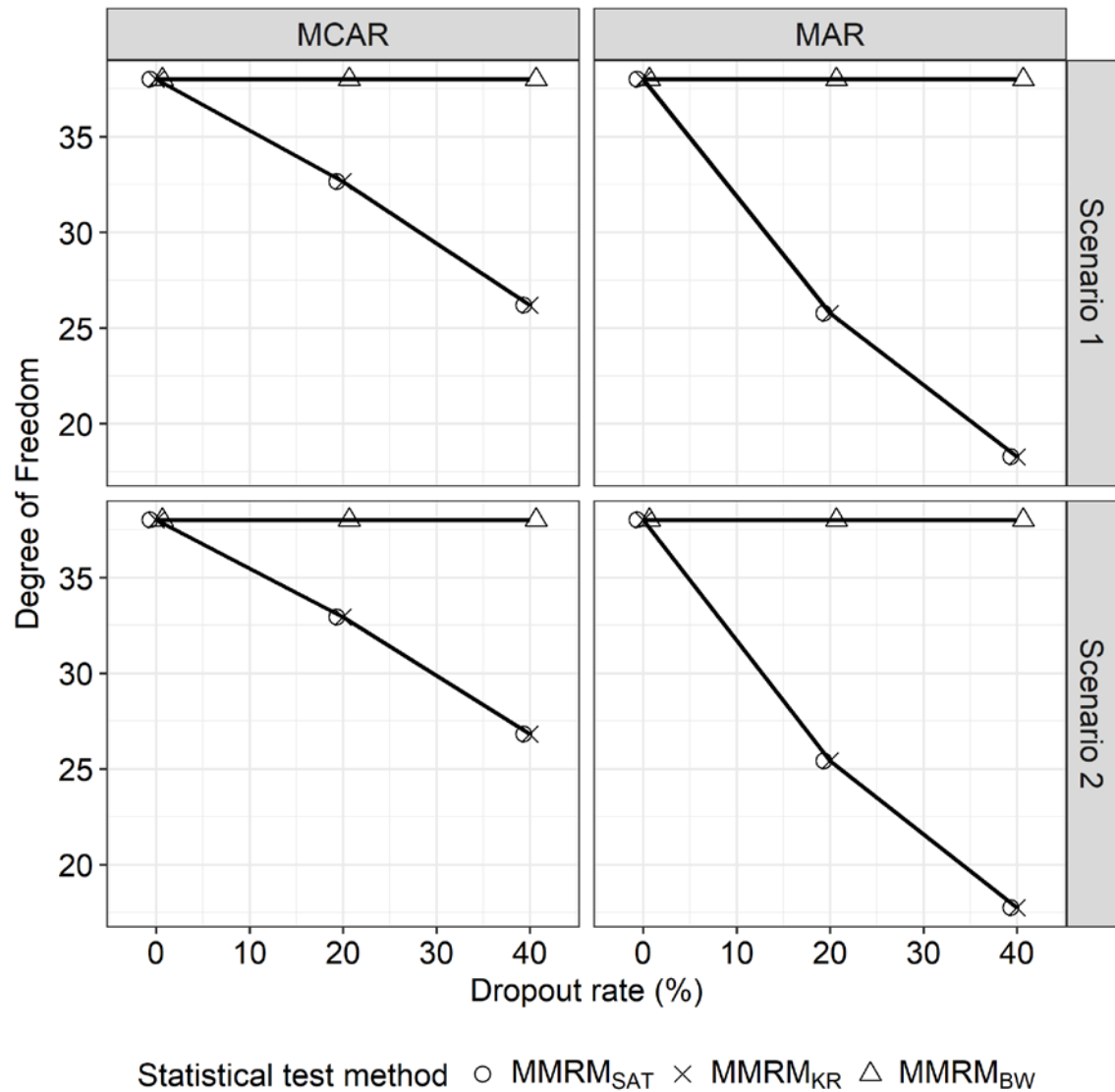


図 8.3-25  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup>（シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 20 例）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定



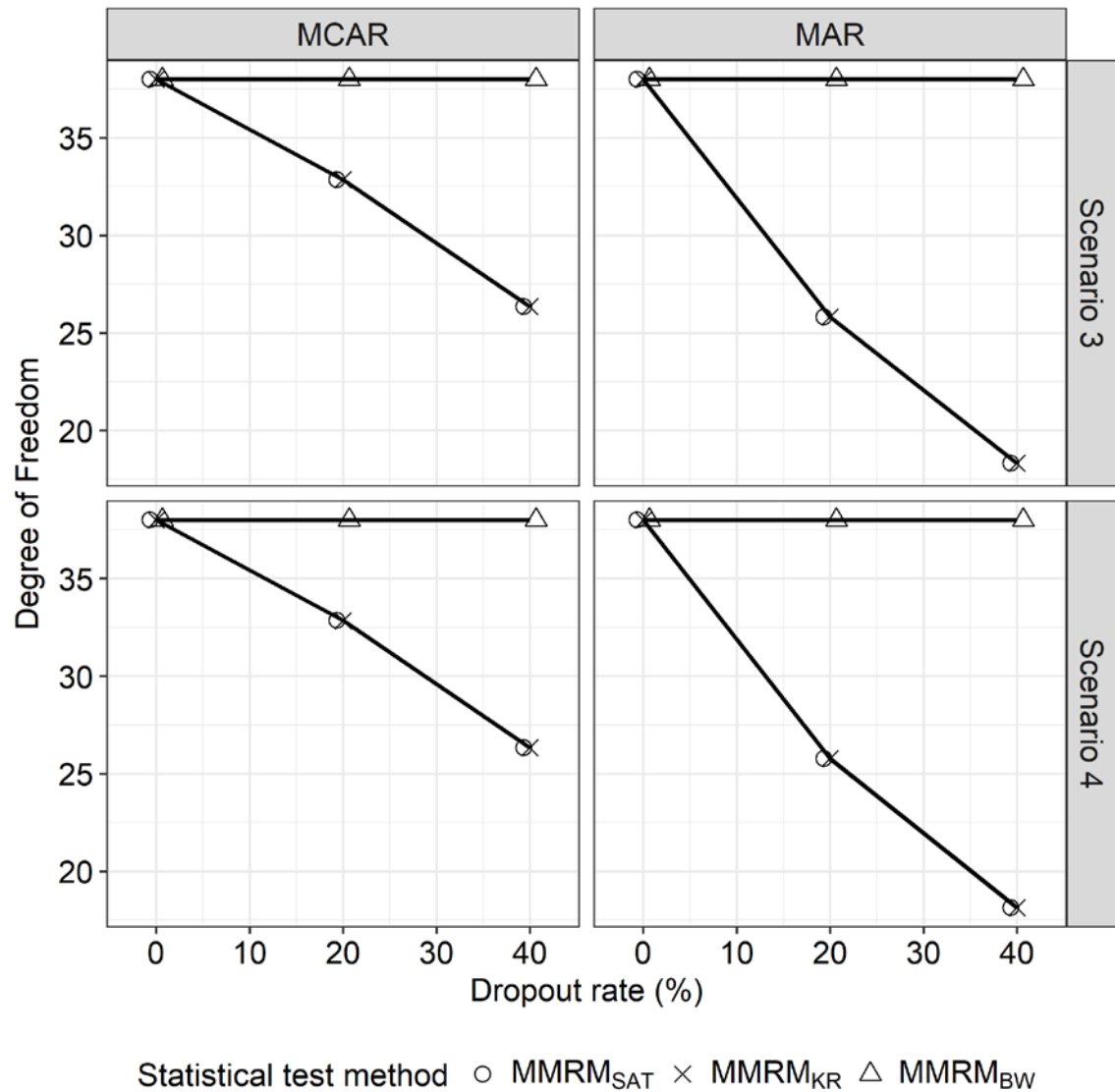


図 8.3-26  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup>（シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 20 例）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

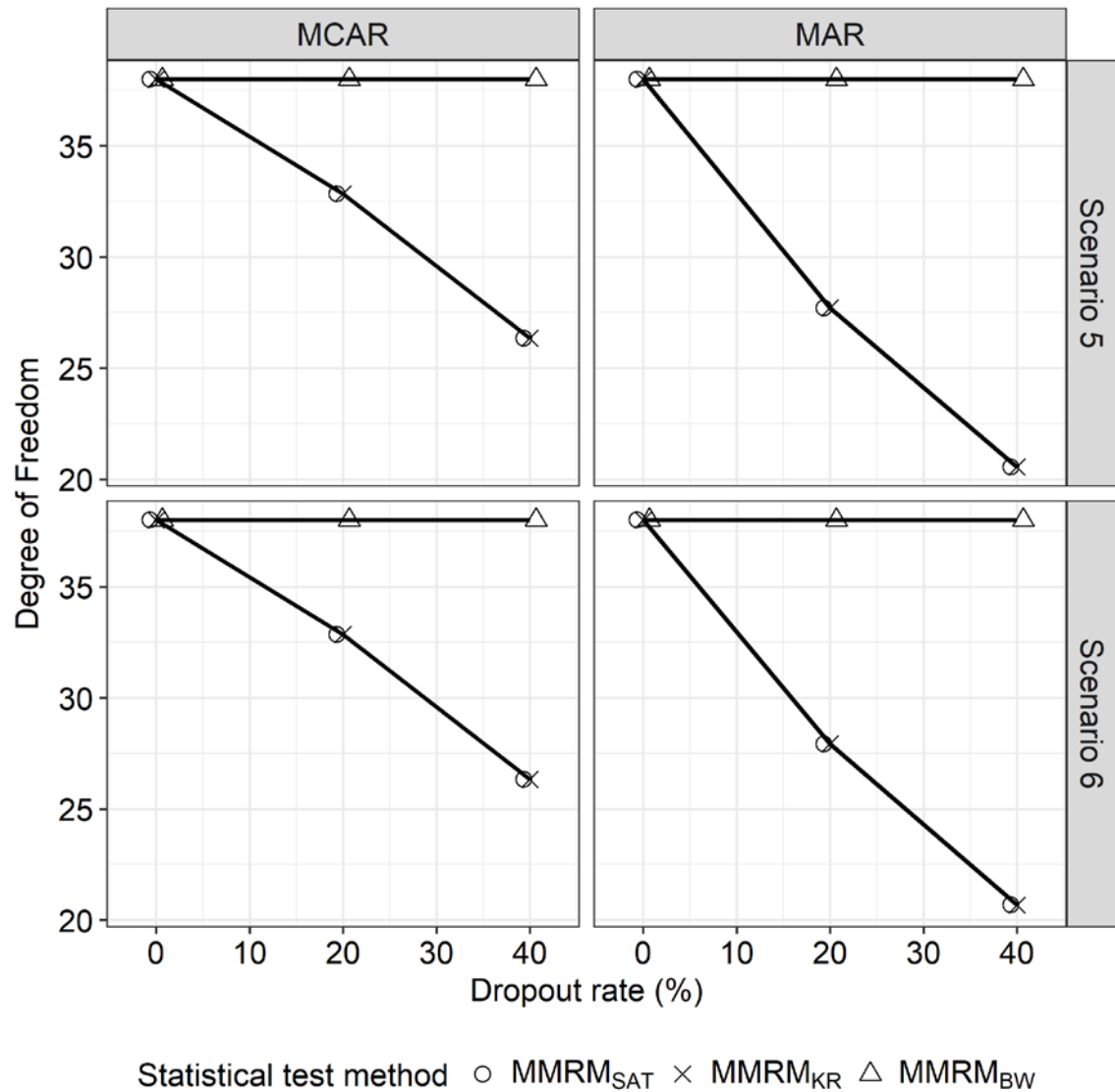


図 8.3-27  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup>（シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 20 例）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

### 8.3.4.2 反復回数 10000 回の結果

表 8.3-10  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup>（シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 5 例, 反復回数 10000 回）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	8.00	---	---	8.00	---	---
	MMRM <sub>KR</sub>	7.97	---	---	7.97	---	---
	MMRM <sub>BW</sub>	8.00	---	---	8.00	---	---
2	MMRM <sub>SAT</sub>	8.00	---	---	8.00	---	---
	MMRM <sub>KR</sub>	7.97	---	---	7.97	---	---
	MMRM <sub>BW</sub>	8.00	---	---	8.00	---	---
3	MMRM <sub>SAT</sub>	8.00	---	---	8.00	---	---
	MMRM <sub>KR</sub>	7.97	---	---	7.97	---	---
	MMRM <sub>BW</sub>	8.00	---	---	8.00	---	---
4	MMRM <sub>SAT</sub>	8.00	---	---	8.00	---	---
	MMRM <sub>KR</sub>	7.97	---	---	7.97	---	---
	MMRM <sub>BW</sub>	8.00	---	---	8.00	---	---
5	MMRM <sub>SAT</sub>	8.00	---	---	8.00	---	---
	MMRM <sub>KR</sub>	7.97	---	---	7.97	---	---
	MMRM <sub>BW</sub>	8.00	---	---	8.00	---	---
6	MMRM <sub>SAT</sub>	8.00	---	---	8.00	---	---
	MMRM <sub>KR</sub>	7.97	---	---	7.97	---	---
	MMRM <sub>BW</sub>	8.00	---	---	8.00	---	---

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

表 8.3-11  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup>（シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	18.00	15.04	12.05	18.00	12.10	8.69
	MMRM <sub>KR</sub>	18.00	15.04	12.02	18.00	12.10	8.65
	MMRM <sub>BW</sub>	18.00	18.00	18.00	18.00	18.00	18.00
2	MMRM <sub>SAT</sub>	18.00	15.16	12.26	18.00	11.94	8.52
	MMRM <sub>KR</sub>	18.00	15.16	12.25	18.00	11.94	8.48
	MMRM <sub>BW</sub>	18.00	18.00	18.00	18.00	18.00	18.00
3	MMRM <sub>SAT</sub>	18.00	15.15	12.02	18.00	12.17	8.72
	MMRM <sub>KR</sub>	18.00	15.15	12.01	18.00	12.17	8.69
	MMRM <sub>BW</sub>	18.00	18.00	18.00	18.00	18.00	18.00
4	MMRM <sub>SAT</sub>	18.00	15.14	12.03	18.00	12.18	8.67
	MMRM <sub>KR</sub>	18.00	15.14	12.01	18.00	12.18	8.64
	MMRM <sub>BW</sub>	18.00	18.00	18.00	18.00	18.00	18.00
5	MMRM <sub>SAT</sub>	18.00	15.15	12.04	18.00	13.04	9.58
	MMRM <sub>KR</sub>	18.00	15.15	12.02	18.00	13.04	9.57
	MMRM <sub>BW</sub>	18.00	18.00	18.00	18.00	18.00	18.00
6	MMRM <sub>SAT</sub>	18.00	15.15	12.06	18.00	13.11	9.65
	MMRM <sub>KR</sub>	18.00	15.15	12.03	18.00	13.11	9.63
	MMRM <sub>BW</sub>	18.00	18.00	18.00	18.00	18.00	18.00

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

表 8.3-12  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup>（シナリオ 1, 2, 3, 4, 5 及び 6, MCAR 及び MAR 下, 1 群 20 例, 反復回数 10000 回）

シナリオ	方法 <sup>†</sup>	累積欠測割合					
		MCAR			MAR		
		0%	20%	40%	0%	20%	40%
1	MMRM <sub>SAT</sub>	38.00	32.75	26.55	38.00	25.64	18.01
	MMRM <sub>KR</sub>	38.00	32.75	26.55	38.00	25.64	18.01
	MMRM <sub>BW</sub>	38.00	38.00	38.00	38.00	38.00	38.00
2	MMRM <sub>SAT</sub>	38.00	32.96	27.06	38.00	25.22	17.50
	MMRM <sub>KR</sub>	38.00	32.96	27.06	38.00	25.22	17.50
	MMRM <sub>BW</sub>	38.00	38.00	38.00	38.00	38.00	38.00
3	MMRM <sub>SAT</sub>	38.00	32.91	26.55	38.00	25.69	18.08
	MMRM <sub>KR</sub>	38.00	32.91	26.55	38.00	25.69	18.08
	MMRM <sub>BW</sub>	38.00	38.00	38.00	38.00	38.00	38.00
4	MMRM <sub>SAT</sub>	38.00	32.91	26.55	38.00	25.69	17.95
	MMRM <sub>KR</sub>	38.00	32.91	26.55	38.00	25.69	17.95
	MMRM <sub>BW</sub>	38.00	38.00	38.00	38.00	38.00	38.00
5	MMRM <sub>SAT</sub>	38.00	32.91	26.55	38.00	27.59	20.56
	MMRM <sub>KR</sub>	38.00	32.91	26.55	38.00	27.59	20.56
	MMRM <sub>BW</sub>	38.00	38.00	38.00	38.00	38.00	38.00
6	MMRM <sub>SAT</sub>	38.00	32.91	26.55	38.00	27.76	20.64
	MMRM <sub>KR</sub>	38.00	32.91	26.55	38.00	27.76	20.64
	MMRM <sub>BW</sub>	38.00	38.00	38.00	38.00	38.00	38.00

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

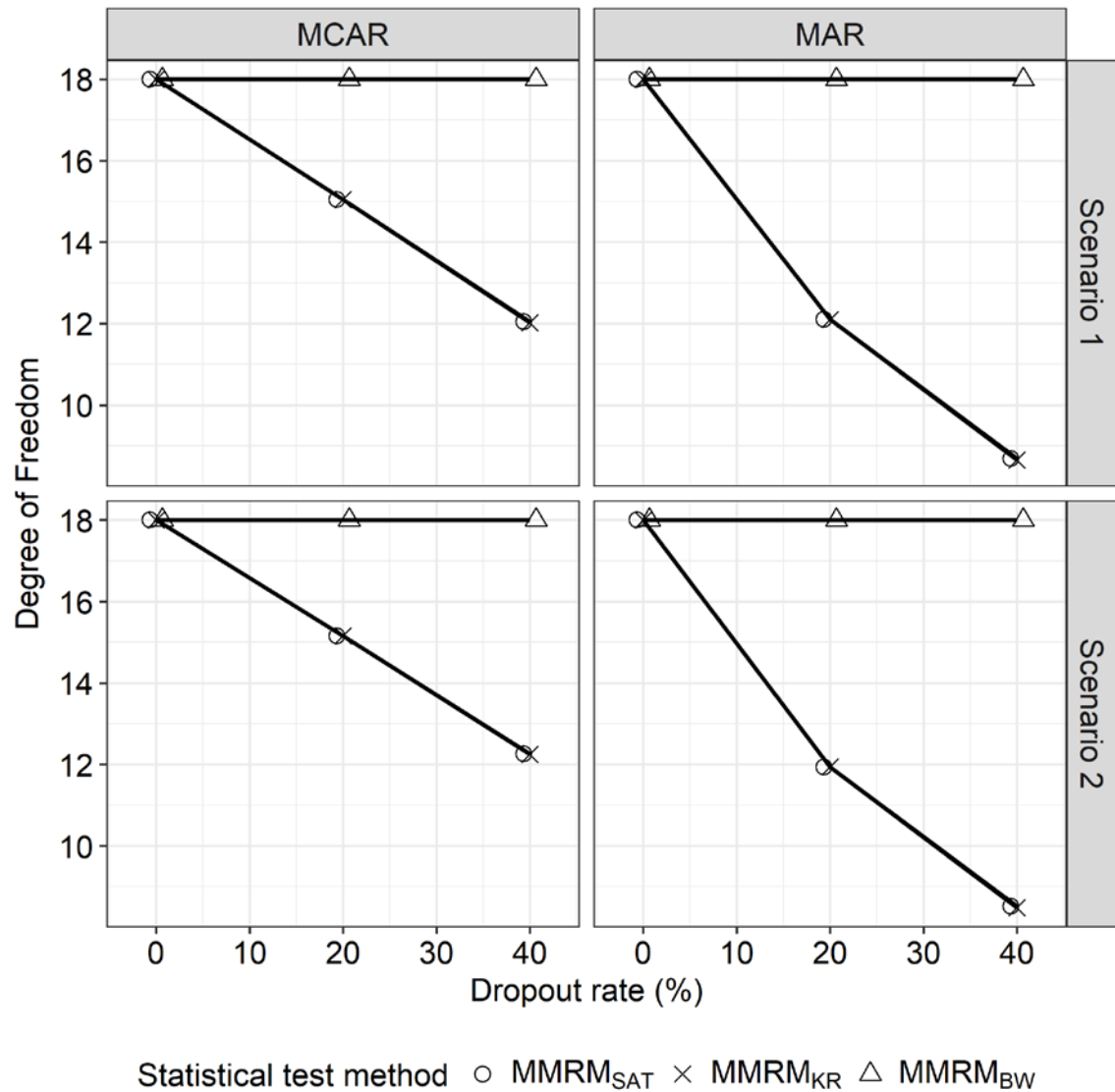


図 8.3-28  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup>（シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

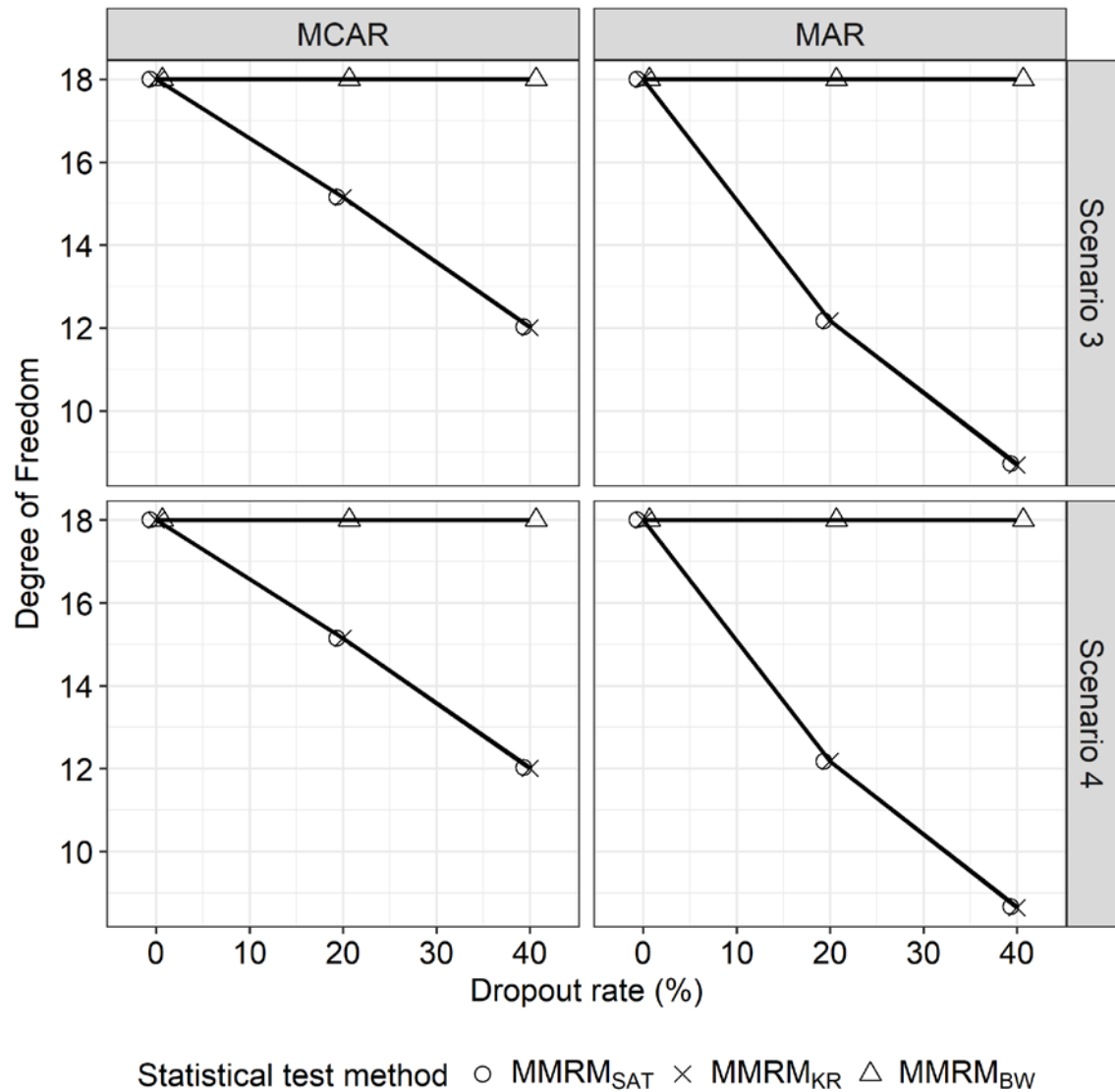


図 8.3-29  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup>（シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

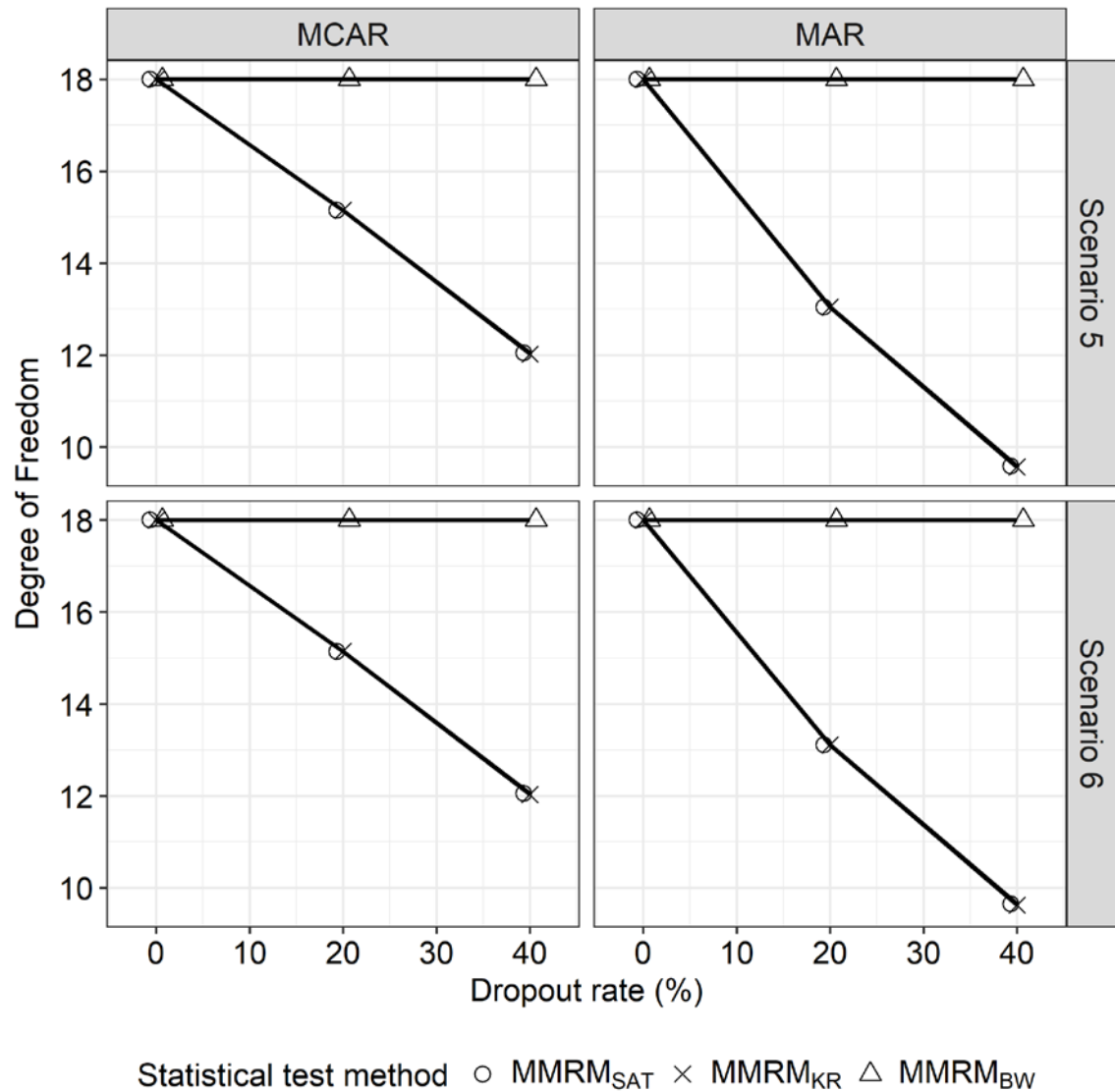


図 8.3-30  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup>（シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 10 例, 反復回数 10000 回）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

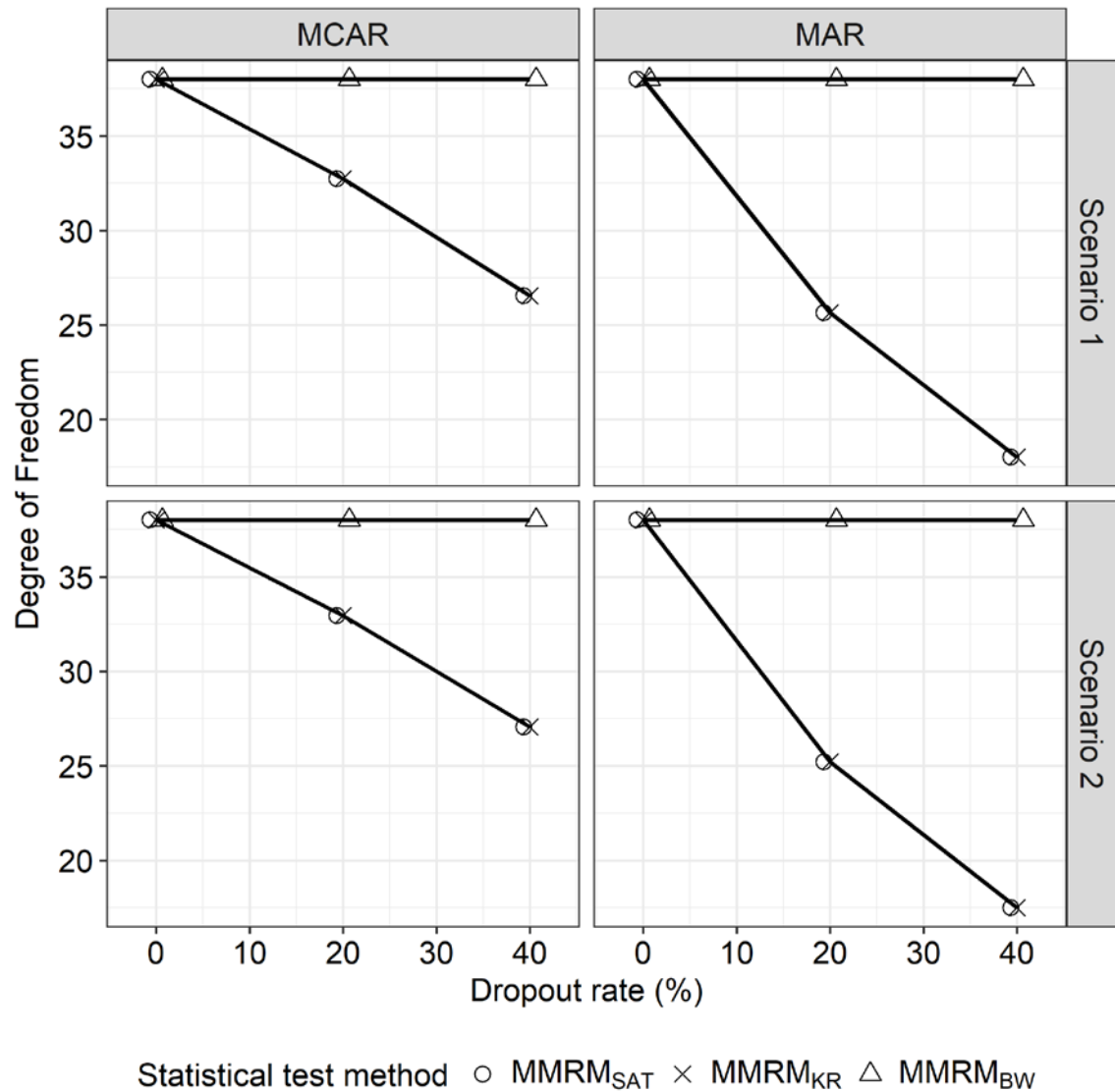


図 8.3-31  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup>（シナリオ 1 及び 2, MCAR 及び MAR 下, 1 群 20 例, 反復回数 10000 回）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定



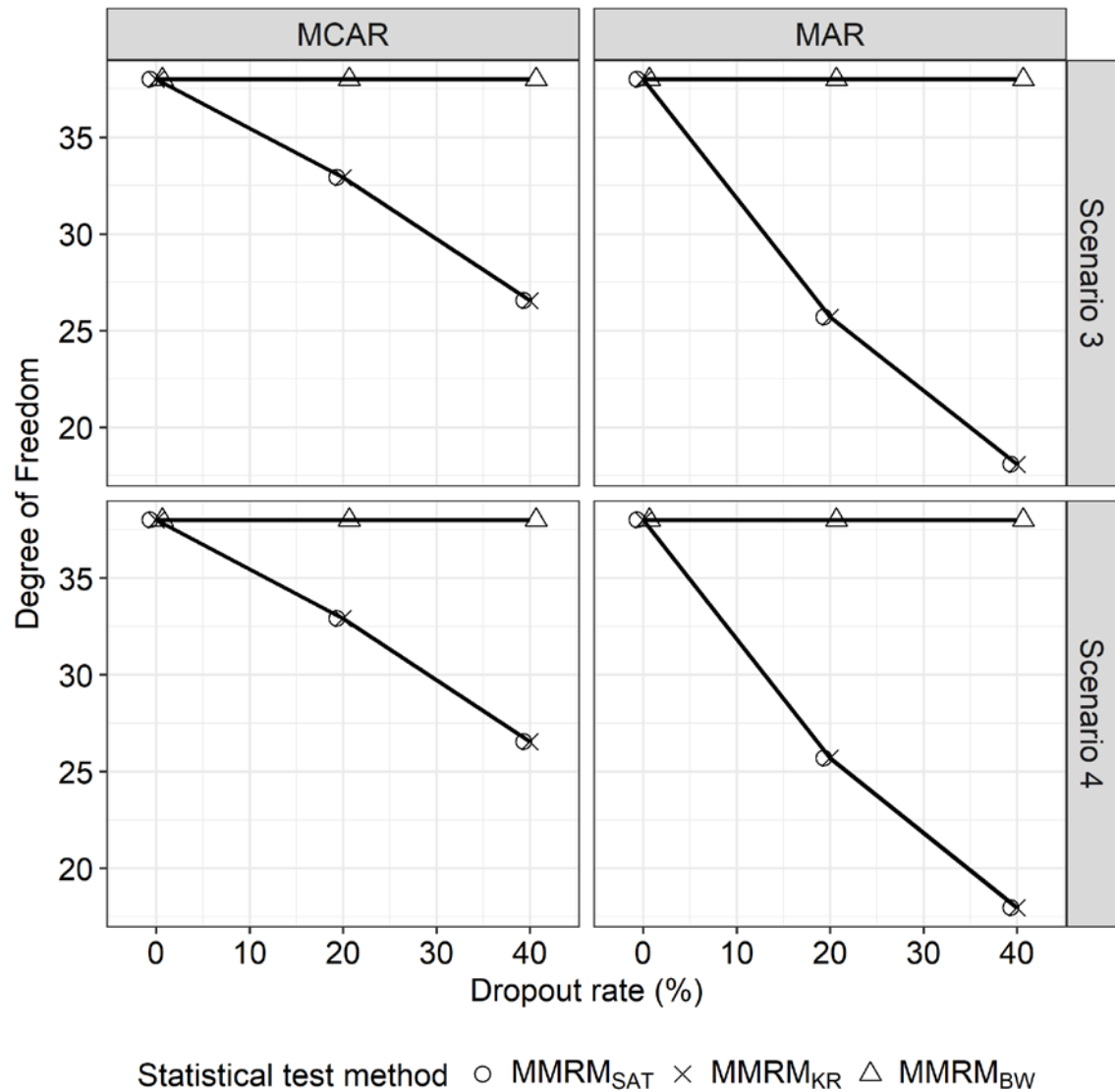


図 8.3-32  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup>（シナリオ 3 及び 4, MCAR 及び MAR 下, 1 群 20 例, 反復回数 10000 回）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

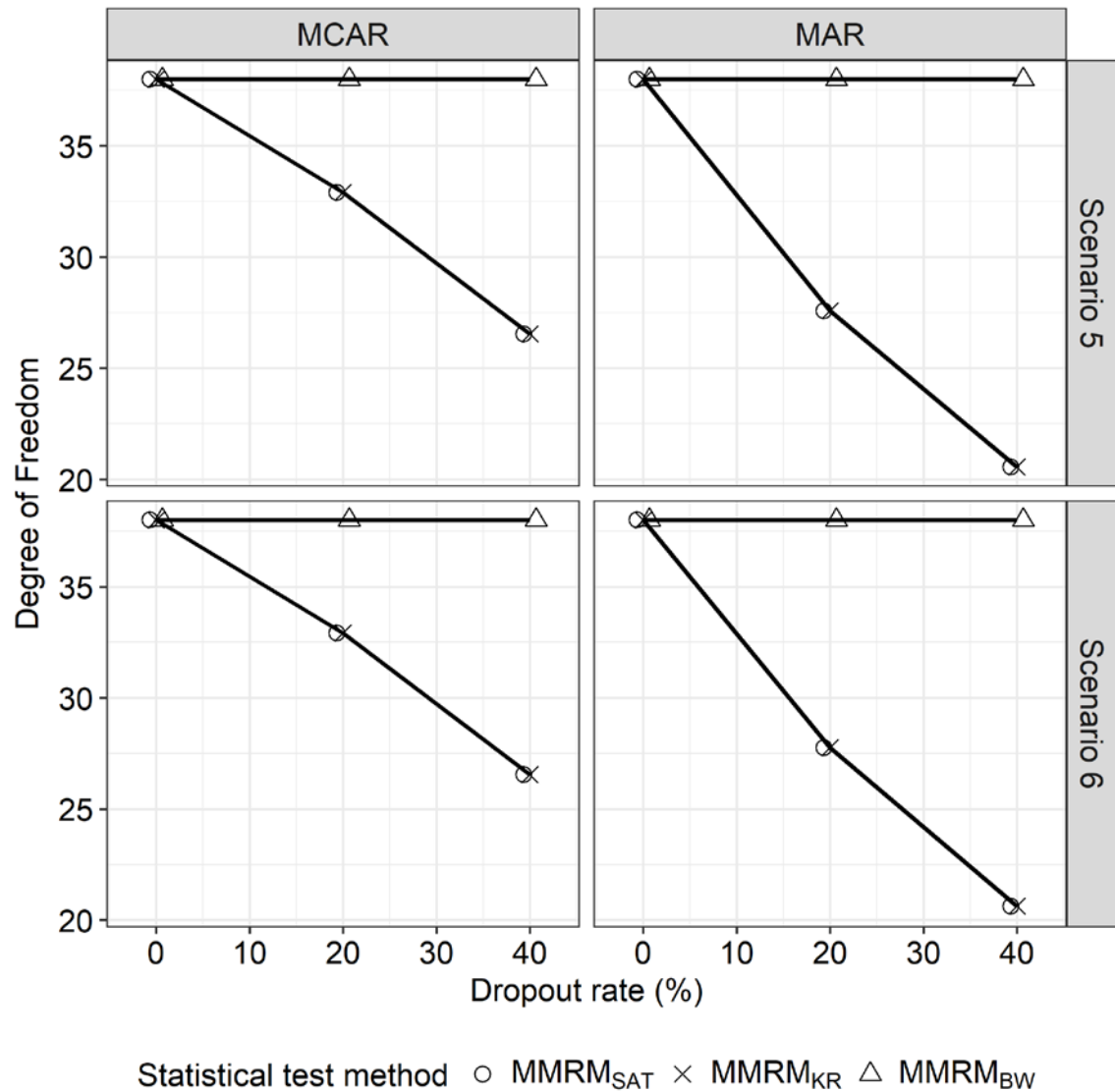


図 8.3-33  $t$ 検定の自由度のシミュレーション実験結果<sup>†</sup>（シナリオ 5 及び 6, MCAR 及び MAR 下, 1 群 20 例, 反復回数 10000 回）

<sup>†</sup>MMRM<sub>SAT</sub>, Satterthwaite 法による REML 推定を用いた Wald 型検定； MMRM<sub>KR</sub>, Kenward and Roger 法による REML 推定を用いた Wald 型検定； MMRM<sub>BW</sub>, Between-within 法による REML 推定を用いた Wald 型検定

## 8.4 産後うつ臨床試験個別値一覧

ランダム化された 61 例の EPDS スコアに関して個別値が書籍 (Everitt B. S., & Der, G., 2001) に紹介されている。個別値一覧を以下の表 8.4-1 に示す。T1 から T8 までは 1 か月ごとの時点を経時的に示しており，表中の「---」は欠測値である。本試験では，プラセボ群で約 37.0% (10/27)，エストロゲン群で約 17.6% (6/34) の被験者に，最終時点 (T8) の EPDS スコアに欠測が発生している。

表 8.4-1 産後うつ臨床試験個別値一覧 (61 例)

Treatment	T1	T2	T3	T4	T5	T6	T7	T8
0	18	18	17	18	15	17	14	15
0	25	27	26	23	18	17	12	10
0	19	16	17	14	---	---	---	---
0	24	17	14	23	17	13	12	12
0	19	15	12	10	8	4	5	5
0	22	20	19	11	9	8	6	5
0	28	16	13	13	9	7	8	7
0	24	28	26	27	---	---	---	---
0	27	28	26	24	19	13	11	9
0	18	25	9	12	15	12	13	20
0	23	24	14	---	---	---	---	---
0	21	16	19	13	14	23	15	11
0	23	26	13	22	---	---	---	---
0	21	21	7	13	---	---	---	---
0	22	21	18	---	---	---	---	---
0	23	22	18	---	---	---	---	---
0	26	26	19	13	22	12	18	13
0	20	19	19	7	8	2	5	6
0	20	22	20	15	20	17	15	13
0	15	16	7	8	12	10	10	12
0	22	21	19	18	16	13	16	15
0	24	20	16	21	17	21	16	18
0	---	17	15	---	---	---	---	---
0	24	22	20	21	17	14	14	10
0	24	19	16	19	---	---	---	---
0	22	21	7	4	4	4	3	3
0	16	18	19	---	---	---	---	---
1	21	21	13	12	9	9	13	6
1	27	27	8	17	15	7	5	7
1	24	15	8	12	10	10	6	5
1	28	24	14	14	13	12	18	15
1	19	15	15	16	11	14	12	8
1	17	17	9	5	3	6	0	2
1	21	20	7	7	7	12	9	6
1	18	18	8	1	1	2	0	1
1	24	28	11	7	3	2	2	2
1	21	21	7	8	6	6	4	4
1	19	18	8	6	4	11	7	6
1	28	27	22	27	24	22	24	23
1	23	19	14	12	15	12	9	6
1	21	20	13	10	7	9	11	11
1	18	16	17	26	---	---	---	---
1	22	21	19	9	9	12	5	7
1	24	23	11	7	5	8	2	3
1	23	23	16	13	---	---	---	---
1	24	24	16	15	11	11	11	11
1	25	25	20	18	16	9	10	6

1	15	22	15	17	12	9	8	6
1	26	20	7	2	1	0	0	2
1	22	20	12	8	6	3	2	3
1	24	25	15	24	18	15	13	12
1	22	18	17	6	2	2	0	1
1	27	26	1	18	10	13	12	10
1	22	20	27	13	9	8	4	5
1	20	17	20	10	8	8	7	6
1	22	22	12	---	---	---	---	---
1	20	22	15	2	4	6	3	3
1	21	23	11	9	10	8	7	4
1	17	17	15	---	---	---	---	---
1	18	22	7	12	15	---	---	---
1	23	26	24	---	---	---	---	---

## 8.5 認知行動療法の臨床試験個別値一覧

認知行動療法の臨床試験の対象者から抽出された 100 例分の個別値を示す。全対象者の個別値データは公開されていないため、実データへの適用においても本 100 例分のデータを対象に解析を実施した。

表 8.5-1 認知行動療法の臨床試験個別値一覧（100 例）

被験者 番号	前治療	抑うつ 期間	治療	Pre	2m	BDI 3m	5m	8m
1	N	>6m	TAU	29	2	2	NA	NA
2	Y	>6m	BtB	32	16	24	17	20
3	Y	<6m	TAU	25	20	NA	NA	NA
4	N	>6m	BtB	21	17	16	10	9
5	Y	>6m	BtB	26	23	NA	NA	NA
6	Y	<6m	BtB	7	0	0	0	0
7	Y	<6m	TAU	17	7	7	3	7
8	N	>6m	TAU	20	20	21	19	13
9	Y	<6m	BtB	18	13	14	20	11
10	Y	>6m	BtB	20	5	5	8	12
11	N	>6m	TAU	30	32	24	12	2
12	Y	<6m	BtB	49	35	NA	NA	NA
13	N	>6m	TAU	26	27	23	NA	NA
14	Y	>6m	TAU	30	26	36	27	22
15	Y	>6m	BtB	23	13	13	12	23
16	N	<6m	TAU	16	13	3	2	0
17	N	>6m	BtB	30	30	29	NA	NA
18	N	<6m	BtB	13	8	8	7	6
19	N	>6m	TAU	37	30	33	31	22
20	Y	<6m	BtB	35	12	10	8	10
21	N	>6m	BtB	21	6	NA	NA	NA
22	N	<6m	TAU	26	17	17	20	12
23	N	>6m	TAU	29	22	10	NA	NA
24	N	>6m	TAU	20	21	NA	NA	NA
25	N	>6m	TAU	33	23	NA	NA	NA
26	N	>6m	BtB	19	12	13	NA	NA
27	Y	<6m	TAU	12	15	NA	NA	NA
28	Y	>6m	TAU	47	36	49	34	NA
29	Y	>6m	BtB	36	6	0	0	2
30	N	<6m	BtB	10	8	6	3	3
31	N	<6m	TAU	27	7	15	16	0
32	N	<6m	BtB	18	10	10	6	8

33	Y	<6m	BtB	11	8	3	2	15
34	Y	<6m	BtB	6	7	NA	NA	NA
35	Y	>6m	BtB	44	24	20	29	14
36	N	<6m	TAU	38	38	NA	NA	NA
37	N	<6m	TAU	21	14	20	1	8
38	Y	>6m	TAU	34	17	8	9	13
39	Y	<6m	BtB	9	7	1	NA	NA
40	Y	>6m	TAU	38	27	19	20	30
41	Y	<6m	BtB	46	40	NA	NA	NA
42	N	<6m	TAU	20	19	18	19	18
43	Y	>6m	TAU	17	29	2	0	0
44	N	>6m	BtB	18	20	NA	NA	NA
45	Y	>6m	BtB	42	1	8	10	6
46	N	<6m	BtB	30	30	NA	NA	NA
47	Y	<6m	BtB	33	27	16	30	15
48	N	<6m	BtB	12	1	0	0	NA
49	Y	<6m	BtB	2	5	NA	NA	NA
50	N	>6m	TAU	36	42	49	47	40
51	N	<6m	TAU	35	30	NA	NA	NA
52	N	<6m	BtB	23	20	NA	NA	NA
53	N	>6m	TAU	31	48	38	38	37
54	Y	<6m	BtB	8	5	7	NA	NA
55	Y	<6m	TAU	23	21	26	NA	NA
56	Y	<6m	BtB	7	7	5	4	0
57	N	<6m	TAU	14	13	14	NA	NA
58	N	<6m	TAU	40	36	33	NA	NA
59	Y	<6m	BtB	23	30	NA	NA	NA
60	N	>6m	BtB	14	3	NA	NA	NA
61	N	>6m	TAU	22	20	16	24	16
62	N	>6m	TAU	23	23	15	25	17
63	N	<6m	TAU	15	7	13	13	NA
64	N	>6m	TAU	8	12	11	26	NA
65	N	>6m	BtB	12	18	NA	NA	NA
66	N	>6m	TAU	7	6	2	1	NA
67	Y	<6m	TAU	17	9	3	1	0
68	Y	<6m	BtB	33	18	16	NA	NA
69	N	<6m	TAU	27	20	NA	NA	NA
70	N	<6m	BtB	27	30	NA	NA	NA
71	N	<6m	BtB	9	6	10	1	0
72	N	>6m	BtB	40	30	12	NA	NA
73	N	>6m	TAU	11	8	7	NA	NA
74	N	<6m	TAU	9	8	NA	NA	NA
75	N	>6m	TAU	14	22	21	24	19
76	Y	>6m	BtB	28	9	20	18	13
77	N	>6m	BtB	15	9	13	14	10
78	Y	>6m	BtB	22	10	5	5	12
79	N	<6m	TAU	23	9	NA	NA	NA
80	N	>6m	TAU	21	22	24	23	22
81	N	>6m	TAU	27	31	28	22	14
82	Y	>6m	BtB	14	15	NA	NA	NA
83	N	>6m	TAU	10	13	12	8	20
84	Y	<6m	TAU	21	9	6	7	1
85	Y	>6m	BtB	46	36	53	NA	NA
86	N	>6m	BtB	36	14	7	15	15
87	Y	>6m	BtB	23	17	NA	NA	NA
88	Y	>6m	TAU	35	0	6	0	1
89	Y	<6m	BtB	33	13	13	10	8
90	N	<6m	BtB	19	4	27	1	2

91	N	<6m	TAU	16	NA	NA	NA	NA
92	Y	<6m	BtB	30	26	28	NA	NA
93	Y	<6m	BtB	17	8	7	12	NA
94	N	>6m	BtB	19	4	3	3	3
95	N	>6m	BtB	16	11	4	2	3
96	Y	>6m	BtB	16	16	10	10	8
97	Y	<6m	TAU	28	NA	NA	NA	NA
98	N	>6m	BtB	11	22	9	11	11
99	N	<6m	TAU	13	5	5	0	6
100	Y	<6m	TAU	43	NA	NA	NA	NA