



製薬協

# ベイズ流試験デザインにおけるシミュレーションの利用法

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  
2024年度タスクフォース3-2  
STAT 新時代

Ver 1.0  
2025年10月

## 目次

1.	はじめに.....	6
1.1	背景と目的 .....	6
1.2	本稿の構成 .....	7
1.3	参考文献 .....	7
2.	ベイズ流解析の基本的な考え方とシミュレーションのポイント.....	9
2.1	ベイズ流解析のメリット及び課題.....	9
2.2	ベイズ流解析のシミュレーション.....	10
2.2.1	解析事前分布とデザイン事前分布.....	10
2.2.2	シミュレーションのポイント .....	13
2.2.2.1	第一種の過誤確率におけるシナリオの範囲.....	13
2.2.2.2	第一種の過誤確率におけるシミュレーションの繰り返し数.....	13
2.2.2.3	シミュレーションで用いる乱数 .....	13
2.2.2.4	検出力やその他の動作特性の留意事項.....	14
2.3	シンプルなベイズ流解析を用いた試験.....	14
2.3.1	Thall and Simon の方法.....	14
2.3.2	Case simple example.....	15
2.3.3	頻度論的動作特性の評価 .....	17
2.3.4	まとめ.....	18
2.3.5	プログラムの利用方法 .....	19
2.4	参考文献 .....	19
3.	外部データの利用.....	21
3.1	小児外挿の場合 .....	22
3.1.1	本試験のデザイン .....	22
3.1.2	解析方法の詳細.....	23
3.1.3	シミュレーション .....	24
3.1.3.1	シミュレーションの目的 .....	24
3.1.3.2	シミュレーションの設定 .....	24
3.1.3.3	シミュレーションの結果 .....	26
3.1.3.4	シミュレーションの結果の考察 .....	27
3.1.3.5	Rを用いたコードのポイント.....	28
3.1.3.6	SASを用いたコードのポイント .....	31
3.2	外部対照の場合 .....	33
3.2.1	想定される試験の状況 .....	33
3.2.2	試験デザイン案.....	34
3.2.3	シミュレーション実験 .....	36

3.2.3.1	シミュレーションの目的 .....	36
3.2.3.2	シミュレーションの設定 .....	36
3.2.3.3	シミュレーションの結果 .....	39
3.2.3.4	シミュレーションの結果の考察 .....	40
3.2.4	R package の紹介 (psborrow, psborrow2) .....	41
3.2.4.1	psborrow と psborrow2 の整理.....	41
3.2.4.2	psborrow package の使用例 .....	42
Flow 1:	Set treatment arm indicators and covariates.....	42
Flow 2:	Set trial operational parameters and Simulate survival times .....	44
Flow 3:	Choose borrowing method and run Bayesian analysis.....	45
Flow 4:	Summerize results to assess impact from borrowing .....	47
3.2.4.3	psborrow2 package の使用例 .....	48
Flow 1:	Set treatment arm indicators and covariates.....	48
Flow 2:	Set trial operational parameters and Simulate survival times .....	49
Flow 3:	Choose borrowing method and run Bayesian analysis.....	50
3.2.5	Case simple example results.....	53
3.2.5.1	本項の位置づけ .....	53
3.2.5.2	本項で使用する仮想データ .....	53
3.2.5.3	ベイズ流ワイブル回帰モデルに基づく CMP 法とその実装 .....	58
3.2.5.4	Effective Historical Sample Size (EHSS) とその実装.....	63
3.2.5.5	結果の解釈と課題 .....	65
3.2.5.6	プログラムの利用方法 .....	65
3.2.5.7	謝辞 .....	66
3.3	参考文献 .....	66
4.	用量反応試験.....	68
4.1	緒言 .....	68
4.1.1	本試験のデザイン .....	68
4.1.2	解析方法.....	69
4.1.2.1	NDLM のモデリング .....	69
4.1.2.2	BHM のモデリング .....	70
4.1.2.3	NHM のモデリング .....	71
4.1.2.4	効果判定 .....	71
4.1.3	シミュレーションの目的.....	71
4.2	シミュレーション設定.....	71
4.2.1	試験デザイン案.....	71
4.2.2	試験の設定.....	71

4.2.3	データ発生及びシナリオ .....	72
4.2.4	評価指標 .....	73
4.2.5	シミュレーションの繰り返し数と乱数シード .....	73
4.2.6	Stan コードのポイント .....	74
4.2.6.1	NDLM .....	74
4.2.6.2	BHM .....	75
4.3	シミュレーション結果 .....	76
4.3.1	検出力、被験者数、最大効果を持つ用量を正しく特定できる確率 .....	76
4.3.2	治療効果の推定値、バイアス、RMSE .....	77
4.4	要約及び推奨事項 .....	80
4.4.1	考察 .....	80
4.4.1.1	検出力 .....	80
4.4.1.2	平均被験者数 .....	80
4.4.1.3	最大効果を持つ用量を正しく特定できる確率 .....	80
4.4.1.4	バイアス .....	80
4.4.1.5	RMSE .....	80
4.4.1.6	推奨される試験デザイン .....	80
4.5	Case Example Result .....	81
4.6	参考文献 .....	81
4.7	Appendix .....	82
5.	バスケット試験 .....	88
5.1	本試験の目的 .....	88
5.2	シミュレーションの目的 .....	89
5.3	シミュレーション入力設定 .....	89
5.3.1	試験デザイン案 .....	89
5.3.2	中間解析及び判定基準 .....	90
5.3.3	仮想データの生成モデル .....	90
5.3.4	各がん種での発現割合及び治療効果の仮定 .....	90
5.3.5	事前分布の設定 .....	90
5.3.6	評価指標 .....	91
5.3.7	シミュレーションの繰り返し数と乱数シード .....	91
5.3.8	コードのポイント .....	91
5.4	シミュレーション結果 .....	92
5.4.1	シミュレーション結果の概要 .....	92
5.5	要約及び推奨事項 .....	93
5.5.1	考察 .....	93

5.6	参考文献 .....	94
6.	がん第 I 相試験のデザイン .....	95
6.1	がん第 I 相試験のシミュレーションの留意点 .....	95
6.1.1	デザインの概要 .....	95
6.1.2	がん第 I 相試験のシミュレーションにおける留意点 .....	95
6.1.3	参考資料 .....	97
6.2	解析で用いる統計モデル .....	98
6.2.1	BLRM .....	98
6.2.2	BHLRM .....	99
6.2.3	BHLRM + EXNEX .....	101
6.2.4	事前分布の Tips .....	102
6.2.5	まとめ .....	104
6.2.6	参考文献 .....	108
6.3	シミュレーション .....	109
6.3.1	シミュレーションの目的 .....	109
6.3.2	試験の設定 .....	109
6.3.2.1	仮想的な試験設定 .....	109
6.3.2.2	<b>統計モデル</b> .....	110
6.3.3	シミュレーション設定 .....	113
6.3.3.1	シナリオ設定 .....	113
6.3.3.2	シミュレーションプログラム .....	113
6.3.3.2.1	シミュレーションプログラムの設定 .....	114
6.3.4	シミュレーション結果 .....	117
6.3.5	交換可能事前分布の混合割合を変えた場合の動作特性 .....	119
6.3.6	事後分布の生成による仮想的な用量漸増シナリオ .....	121
6.3.7	参考文献 .....	123
7.	おわりに .....	124
Appendix1	: ベイズ解析の基礎 .....	125
執筆者	・ タスクフォースメンバー .....	127

## 1. はじめに

### 1.1 背景と目的

ベイズ流の解析手法は古くからよく知られているが、臨床試験の計画と解析においては、頻度論的立場からの統計手法に基づく方法が主流である。ICH E9<sup>1</sup>では「ベイズ流の手法や他の手法の使用も、それらの使用の理由が明らかであり、異なる仮定の下でも結果として得られる結論が十分に安定している場合には検討することができる。」と記載されているものの、ベイズ流の解析手法が臨床試験で適用される状況は限られていた。

2020年にFDAより、「Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products<sup>2</sup>」が発出され、アダプティブデザイン、外部対照、マスタープロトコルなどの特殊な臨床試験デザインにおいてベイズ流の方法の活用が議論されている。ベイズ流アプローチは、特に複雑なアダプテーションや予測モデルが使用される場合に、試験のデザインと解析に柔軟性を提供でき、ベイズ推定は複数のエビデンスの情報源を体系的に組み合わせる状況で適切である可能性があるため、その利用に関心がより高まっている。また、複雑な状況だけでなく、FDAではBayesian Supplemental Analysis Demonstration Project (<https://www.fda.gov/about-fda/cder-center-clinical-trial-innovation-c3ti/bayesian-supplemental-analysis-bsa-demonstration-project>) が開始され、単純な状況でのベイズ流の補足的解析についても議論が開始されている。また、「Pediatric Inflammatory Bowel Disease: Developing Drugs for Treatment<sup>3</sup>」のドラフトガイダンスでも、小児開発におけるベイズ流の手法を用いた成人データの借用について言及されている他、「Use of Bayesian Methodology in Clinical Trials of Drug and Biological Products」の発出も2025年中に計画されている<sup>4</sup>。クロストリジオイデス・ディフィシル腸炎を適応としたREBYOTAの添付文書<sup>5</sup>ではベイズ流の解析に基づく95%信用区間が有効性の結果に記載されるなど、ベイズ流の解析を用いた事例も増えてきている。

日本においても、AMED医薬品等規制調和・評価研究事業で「希少疾患領域の臨床試験におけるベイズ流アプローチの適用に関する基本指針<sup>6</sup>」がまとめられ、2024年6月に発出された「医薬品開発等におけるマスタープロトコル試験の活用に関する留意事項<sup>7</sup>」でもベイズ流の解析方法について言及されており、その考え方が整備されてきている。

製薬協では、ベイズ流の解析を用いた臨床試験・解析手法について「Rare diseaseの治療効果の推測法<sup>8</sup>」の中で、外部対照や小児での開発における成人データの借用といったテーマでベイズ流の解析手法や事例の紹介を行った。Complex Innovative Trial Designs (CID) に関しては、製薬協が2024年4月に発行した「Complex Innovative Trial Designの現状と事例の整理<sup>9</sup>」の中で、FDAのCIDに関連した取り組み及びCIDを採用した臨床試験の事例の詳細をまとめ、革新的な要素や規制当局と議論となりうるポイントを整理している。また抗悪性腫瘍薬では、第I相試験でベイズ流の手法が使われることがあり、それらの手法は2022年に発行した「近年のがん第I相試験デザインとその選択<sup>10</sup>」で紹介された。

ベイズ流の解析を用いた試験では、途中の試験結果に基づく意思決定により試験に柔軟性を持たせたり、外部の情報を借用することで効率的な医薬品開発が可能となる場合がある。しかし、事前分布や意思決定の基準によっては誤った結論 (erroneous conclusion) を

導く可能性の増加があり得るため、試験実施前にその性能を慎重に評価する必要がある。そのため、ベイズ流試験デザインの性能を評価するためにはシミュレーションが重要な役割を担うが、シミュレーション時の着目点やその実施のためのプログラミングコードについてまとめたものがないのが現状である。

本成果物では、臨床試験の統計解析担当者向けに、ベイズ流試験デザインの動作特性を測るためのシミュレーションの方法を検討し、担当者が必要な状況において適切なベイズ流試験を計画できることを目的とする。さらに、理論の理解を助けるために、本成果物で行ったシミュレーションの R 又は SAS コードは GitHub (<https://github.com/JPMA-DS/2025TF3-2Bayesian.git>) にて公開した。ファイル名の冒頭の数字が、本成果物での章番号を示している。読者が手元で実行し、理解を深めることを期待するが、公開されたプログラムおよび得られた結果の利用責任は、利用者にあることを留意いただきたい。なお、本報告書で紹介した試験デザイン及び解析方法に関しては、当該薬剤の開発状況や対象疾患等によって、規制当局の受け入れ可能性は異なると考えられるため、各状況に応じて判断していただきたい。

## 1.2 本稿の構成

本稿の構成は以下の通りである。第2章では、ベイズ流試験の場合のシミュレーションの流れと評価すべきポイントについて述べ、比較的単純な状況下でのベイズ流解析でのシミュレーション事例を紹介する。第3章から第5章では、より複雑な状況下でのシミュレーション事例を紹介する。第6章では、抗悪性腫瘍薬第 I 相試験でのシミュレーションで評価すべきポイントやその事例を紹介する。Appendix にはベイズ解析の基礎についても触れた。いずれのシミュレーション事例も、コードは GitHub にて公開した。

## 1.3 参考文献

1. 厚生省医薬安全局審査管理課長.(1998).「臨床試験のための統計的原則」について。医薬審第 1047号. <https://www.pmda.go.jp/files/000156112.pdf>
2. Food and Drug Administration. (2020). Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/interacting-fda-complex-innovative-trial-designs-drugs-and-biological-products>
3. Food and Drug Administration. (2024). Pediatric Inflammatory Bowel Disease: Developing Drugs for Treatment. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pediatric-inflammatory-bowel-disease-developing-drugs-treatment>
4. CDER Guidance Agenda New and Revised Draft Guidances Planned for Publication in Calendar Year 2025. <https://www.fda.gov/media/185228/download?attachment>
5. Ferring Pharmaceuticals Inc. (2022). Rebyota. Package Insert. <https://www.fda.gov/media/163587/download>
6. AMED 医薬品等規制調和・評価研究事業.(2021). 希少疾患領域の臨床試験におけるベイズ流アプローチの適用に関する基本指針。

[https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS\\_202302\\_hirakawa1.pdf](https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS_202302_hirakawa1.pdf)

7. AMED 医薬品等規制調和・評価研究事業. (2024). 医薬品開発等におけるマスタープロトコル試験の活用に関する留意事項.

<https://www.mhlw.go.jp/hourei/doc/tsuchi/T240621I0030.pdf>

8. 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. (2022) . Rare disease の治療効果の推測法.

[https://www.jpma.or.jp/information/evaluation/results/allotment/ds\\_202212\\_rare.html](https://www.jpma.or.jp/information/evaluation/results/allotment/ds_202212_rare.html)

9. 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. (2024) . Complex Innovative Trial Design の現状と事例の整理.

[https://www.jpma.or.jp/information/evaluation/results/allotment/DS\\_202404\\_CID.html](https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202404_CID.html)

10. 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. (2022) . 近年のがん第 I 相試験デザインとその選択.

[https://www.jpma.or.jp/information/evaluation/results/allotment/DS\\_202306\\_oncoP1DE.html](https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202306_oncoP1DE.html)

## 2. ベイズ流解析の基本的な考え方とシミュレーションのポイント

### 2.1 ベイズ流解析のメリット及び課題

ベイズ流解析を用いる試験デザインのメリットとして、以下の点が挙げられる<sup>1</sup>。

- ・ 事前情報を取り入れることによって現行試験の情報は拡大され、その精度が増す可能性がある。
- ・ 用いる事前分布によっては、試験に必要なサンプルサイズを頻度流デザインと比べて小さくできる可能性がある。
- ・ アダプティブデザインにより標本サイズを減少させることができる。
- ・ 試験の途中での変更に対応することができる。

一方で、課題としては以下の点が挙げられる。

- ・ 事前情報の確率分布及び関連する統計モデルの設定を事前に決定する必要がある。試験の後半で事前情報や統計モデルを変更すると、試験結果の科学的妥当性が危うくなるため、これらは規制当局と事前に合意形成する必要がある<sup>2</sup>。
- ・ 頻度流アプローチに基づく第一種の過誤確率の厳密な制御が困難になることがある<sup>2</sup>。

シンプルなベイズ流解析を用いた事例として、切除不能の明細胞肉腫 (clear cell sarcoma; CCS) 及び胞巣用軟部肉腫 (alveolar soft-part sarcoma; ASPS) に対するニボルマブの医師主導治験がある<sup>3</sup>。本試験は、オープンラベル、多施設、単群、Phase II 試験であり、ニボルマブの有効性評価に Thall and Simon (1994) の方法<sup>4</sup>が用いられた。主要評価項目は、Response Evaluation Criteria in Solid Tumors (RECIST) に基づく奏効割合である。被験者の組み入れは2016年11月から開始し、組み入れ期間は2年間とした。頻度流のアプローチにより、 $P_0$ (帰無仮説の反応率) = 0.05、 $P_1$ (対立仮説の反応率) = 0.30、 $\alpha = 0.05$ 、 $1 - \beta = 0.8$ の条件で Fisher's Exact test に基づき必要被験者数を見積もったところ、18例の CCS 及び ASPS 患者が必要となったが、本試験での対象が希少疾患であるため、研究者らは2年間で18例の患者を組み入れることは困難と判断した。一方で、CCS 及び ASPS の希少性から、少なくとも15例以上の被験者で奏効割合を評価することが妥当と考え、それを叶えるために Thall and Simon (1994) の方法に基づく試験デザインが採用された。この方法は、二項分布に対する共役な事前分布であるベータ分布を利用する。詳細な理論及びシミュレーションについては、第3章で説明する。CCS 及び ASPS に対する標準治療は確立していないことを踏まえ、無治療における奏効割合の事前分布は、一般的な抗悪性腫瘍薬の試験で設定されている5%を参考に  $Beta(10,190)$  を採用した(事前分布の平均値:  $\frac{10}{10+190} = 0.05$ )。一方、ニボルマブにおける奏効割合の事前分布は、CCS 及び ASPS に対する被験薬の投与経験はないものの、他癌腫での奏効割合30%を参考に  $Beta(0.6, 1.4)$  を採用した(事前分布の平均値:  $\frac{0.6}{0.6+1.4} = 0.3$ )。これらの事前分布の下で、ニボルマブの奏効割合無治療の奏効割合を上回る確率が95%を超える場合にニボルマブは有効と判断することとした。これにより、15例の被験者を組み入れた時点で、3例以上が奏効していれば有効中止ができる試

験計画となり、頻度流アプローチに基づく試験よりも、ベイズ流アプローチに基づく試験の方が被験者数を低く抑えることが可能となった。なお、事前に10,000回のシミュレーション実験を行い、第一種の過誤確率及び検出力はそれぞれ0.076及び0.977であり、いずれも PMDA に合意された。

さらに発展させたベイズ流解析としては、バスケット試験や外部データを利用した試験があるが、これらの詳細は AMED 医薬品等規制調和・評価研究事業の平川班が作成した「希少疾患領域の臨床試験におけるベイズ流アプローチの適用に関する基本指針<sup>5</sup>」にまとめられている。

## 2.2 ベイズ流解析のシミュレーション

### 2.2.1 解析事前分布とデザイン事前分布

コンピュータシミュレーションに基づく頻度論的動作特性の評価に先立ち、解析事前分布とデザイン事前分布の2つの事前分布を用いて意思決定に伴うエラーを評価することを考える。本項では、これらの考え方を **case example** とともに整理する。詳細は手良向 (2023) 等を参照されたい<sup>6</sup>。まずは、解析事前分布とデザイン事前分布についてそれぞれの定義を以下に示す。

- **解析事前分布** : 解析時に事後分布を計算するために設定される分布である。臨床試験においては一般に客観的な証拠に基づいて特定される。
- **デザイン事前分布** : 計画段階で関心のあるパラメータに関する不確実性にモデルを与え、観測データの事前予測分布を計算するために設定される分布である。解析事前分布に対して、デザイン事前分布はある程度の主観性が許容される<sup>6</sup>。

これら2つの事前分布の役割の整理を目的として、がん領域を想定した奏効の有無に関するバイナリー型アウトカム ( $Y$ ) を評価項目とした事例について考えることとする。ただし、 $Y$  は  $N$  人の被験者のうち奏効例数である。具体的には、単群試験の実施を想定し、試験デザイン及び成功の定義を以下のように与える。

- 奏効割合 ( $p_A$ ) の事前分布 :  $Beta(0.5, 0.5)$
- 予定登録例数 ( $N$ ) : 60例
- 成功の定義 :  $\Pr(p_A \geq 0.2 | Y = y) \geq 0.95$

$p_A$  に対する事前分布である  $Beta(0.5, 0.5)$  が解析事前分布である。解析事前分布である点を強調するため、本章では  $p_A$  と下付き文字  $A$  を添えて表記する点に注意する。また、未観測のデータに関して、 $Y$  を  $N$  人の被験者のうち奏効例数とすると、 $Y$  の従う分布は二項分布に従い、 $Bin(N, p_A)$  である。以上の設定下で、 $Y = y$  を得た後の  $p_A$  の事後分布に基づき、成功の定義を満たす境界例数は18例と算出される。詳細は対応する R スクリプト (04\_DesignPrior\_analysisPrior\_example.R) を参照のこと。

表 2.1 奏効例数と成功の定義を満たす確率

奏効例数	16例	17例	18例	19例
奏効割合の点推定値	26.67%	28.33%	<b>30.00%</b>	31.67%
$\Pr(p_A \geq 0.2   Y = y) \geq 0.95$	89.87%	94.21%	<b>96.90%</b>	98.45%

次に、帰無仮説及び対立仮説下における確率分布を想定し、上記の意思決定基準におけるエラーを評価する。未観測のデータに関して、 $Y$ の従う分布は $Bin(N, p_D)$ であり、 $p_D$ に対して事前分布を仮定するとき、この分布をデザイン事前分布と呼ぶ。なお、デザイン事前分布である点を強調するため、同様に $p_D$ と下付き文字 $D$ を添えて表記した。二項分布との共役性により、デザイン事前分布を $Beta(\alpha_D, \beta_D)$ とする。なお、解析事前分布と明示的に区別するため、デザイン事前分布の超パラメータを $\alpha_D, \beta_D$ と表記する。 $Beta(\alpha_D, \beta_D)$ に基づいて、 $N$ 人の被験者のデータを観測した後の事前予測分布はベータ二項分布

$$Y \sim \text{Beta} - \text{Bin}(N, \alpha_D, \beta_D)$$

に従い、以下の確率質量関数を持つ。

$$\Pr(Y = y) = \binom{N}{y} \frac{B(\alpha_D + y, \beta_D + N - y)}{B(\alpha_D, \beta_D)}$$

ここに、 $B(\cdot)$ はベータ関数である。 $p_D$ の事前分布にただ1つの値のみを取る確率分布、すなわち退化分布をおいた場合、事前予測分布は二項分布に帰着する。以上の設定に基づき、閾値奏効割合及び期待奏効割合のそれぞれに対して、複数のデザイン事前分布を設定する。

**帰無仮説下**

- 退化分布 ( $p_D$ ) : 20%のみをとる確率分布
- Informative な事前分布 ( $p_D$ ) :  $Beta(20, 80)$
- 曖昧な事前分布 ( $p_D$ ) :  $Beta(2, 8)$

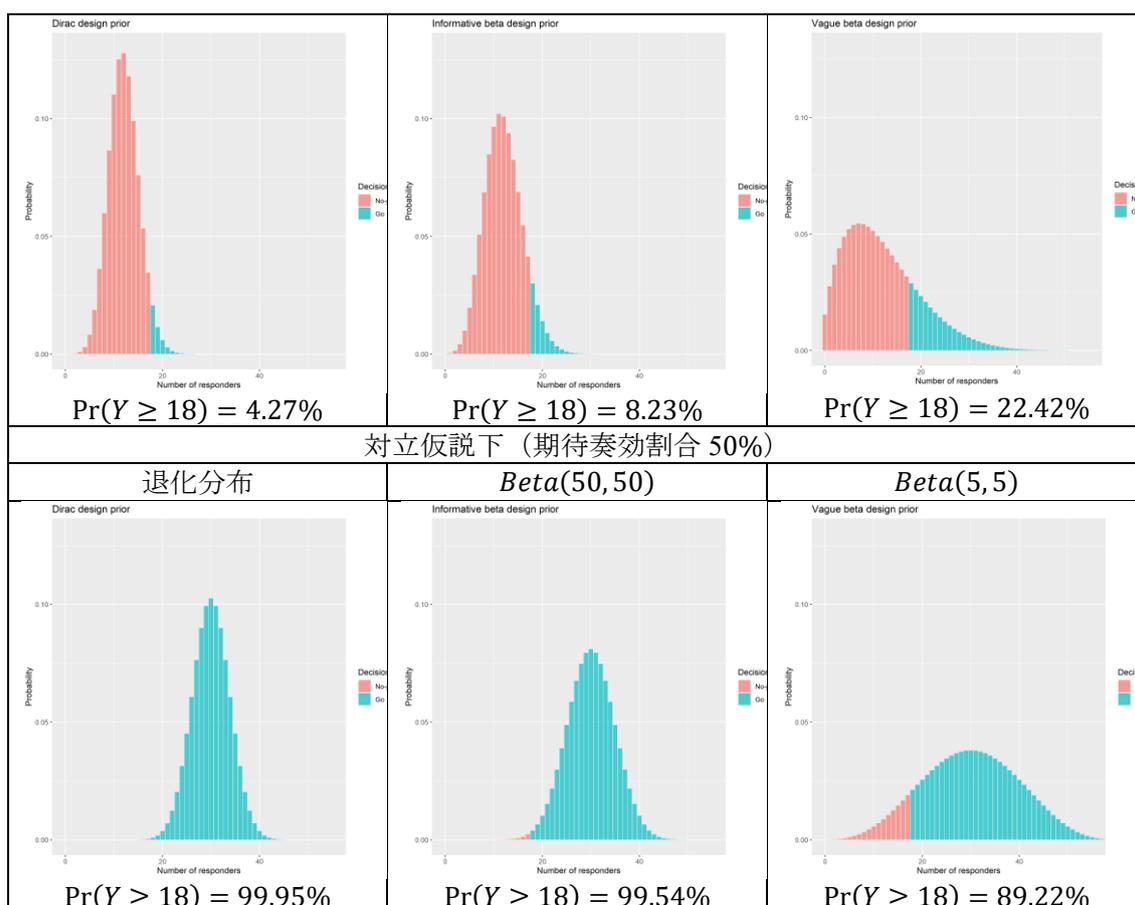
**対立仮説下**

- 退化分布 ( $p_D$ ) : 50%のみをとる確率分布
- Informative な事前分布 ( $p_D$ ) :  $Beta(50, 50)$
- 曖昧な事前分布 ( $p_D$ ) :  $Beta(5, 5)$

以上の3つのデザイン事前分布に対して、事前予測分布及び成功の定義を満たす確率を評価した結果を表 2.2に示す (R スクリプト : 04\_DesignPrior\_analysisPrior\_example.R)。分布の青色で示された面積が成功の定義を満たす確率であり、赤色の面積が成功の定義を満たさない確率である。ここで、対立仮説下で成功の定義を満たす確率をベイズ流検出力と呼ぶ (手良向、2023) 6。

表 2.2 事前予測分布及び成功の定義を満たす確率

帰無仮説下 (閾値奏効割合 20%)		
退化分布	$Beta(20, 80)$	$Beta(2, 8)$



以上の2つの事前分布を用いた推論上の性能評価において、Best et al. (2023) では以下に定義される average type I error の考えを整理している<sup>3</sup>。

$$\int \Pr(\text{Study success} | \theta_c, \theta_t = \theta_c + \delta^*) p(\theta_c) d\theta_c.$$

- $\theta_c, \theta_t$  : 標準治療及び試験治療に対する関心のあるパラメータ
- $p(\theta_c)$  :  $\theta_c$  のデザイン事前分布
- $\delta^*$  : 標準治療に対する試験治療の上乗せ効果を示す定数

$\Pr(\text{Study success} | \theta_c, \theta_t = \theta_c + \delta^*)$  は、ある  $\theta_c$  で条件付けたときの成功確率である。なお、数式上解析事前分布は明示されていないが、上述の case example の成功の定義と同様に、 $\Pr(\text{Study success} | \theta_c, \theta_t = \theta_c + \delta^*)$  において暗に考慮されている。さらに、Best et al. (2023) では解析事前分布に informative な事前分布を設定する場合における第一種の過誤確率の増大に伴う種々の課題に対して、average type I error に基づく新たな評価指標についても整理している<sup>3</sup>。具体的には、比較における治療効果 (オッズ比、ハザード比等) に対して informative な事前分布を設定する場合に対して、対応するデザイン事前分布を治療効果がない、あるいは悪化させる範囲でトランケートすることで average type I error を名目有意水準以内に抑えるというアプローチである。これは、type I error を考慮すべき状況が真に治療効果がない、あるいは悪化させる状況に限定されることに基づく。当該評価指標の活

用可能性は高いと考えられるものの、利用にあたっては規制当局との事前に合意は必須であろう。Average type I error とその周辺の詳細は、Best et al. (2023) を参照されたい<sup>2</sup>。

以降、特に断りが無い限り、本文書では解析事前分布を事前分布と表現する。

## 2.2.2 シミュレーションのポイント

本成果物で紹介するような複雑で革新的な試験デザイン (Complex Innovative Trial Design: CID) は、いずれもベイズ流解析を利用する。試験の実施に先立ち FDA に CID 提案を行う場合には、試験の動作特性をシミュレーションにより評価したシミュレーション報告書を提出する必要がある<sup>7,8</sup>。FDA のアダプティブデザインに関する企業向けガイダンス「Adaptive Design Clinical Trials for Drugs and Biologics Guidance for Industry<sup>9,10</sup>」には、以下のシミュレーションのポイント (2.2.2.1、2.2.2.2、2.2.2.3及び2.2.2.4) が記載されている。さらに、FDA のガイダンス「Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products<sup>7,8</sup>」には事前分布の選択に対する試験の動作特性の感度を評価することは有益であると言及されている。事前分布を慎重に設定せずにベイズ流手法を使用すると、誤った結論を導く可能性が高くなるため、事前分布に関する議論は、ベイズ流の提案の評価にとって特に重要である。

### 2.2.2.1 第一種の過誤確率におけるシナリオの範囲

第一種の過誤確率は、妥当なグリッド上 (試験を実施する上で想定される様々なシナリオや条件) でシミュレーションを実施し算出する。

### 2.2.2.2 第一種の過誤確率におけるシミュレーションの繰り返し数

シミュレーションにより精度の高い動作特性を得るために、各シナリオで繰り返し回数100,000回を推奨する。これは、各シナリオにおける第一種の過誤確率の95%の信頼区間の幅が±0.1%に収まることを保証し、第一種の過誤確率の推定値のごくわずかな差を特定できるようにするためである。ただし、細かいグリッド上を探索し、すべてのシナリオで第一種の過誤確率の推定値が目標とする水準を下回っている場合は、10,000回の繰り返し回数で十分である。また、第一種の過誤確率の推定値の95%信頼区間の上限が目標とする水準を下回っている場合は、その繰り返し数で十分な精度があることが期待されることから、一般により少ない回数のシミュレーションを行うことができる。

### 2.2.2.3 シミュレーションで用いる乱数

シミュレーションのシナリオごとに異なる乱数シードを用いる。これは、シミュレーション結果が生成されたデータセットに依存することを避けるためである。つまり、偶然にも偏ったデータが生成された時に、それが全シナリオに影響するリスクを避けるために、シナリオごとに異なる乱数シードを用いることが推奨される。

### 2.2.2.4 検出力やその他の動作特性の留意事項

複雑なアダプティブデザインでは、検出力やその他の動作特性（期待被験者数、期待試験期間、治療効果の推定値のバイアス等）を推定するために、多数の臨床試験の条件に基づいたシミュレーションが必要になる。患者の不利益の観点から第一種の過誤確率の高い精度が求められる。一方でその他の動作特性は第一種の過誤確率ほどの精度は求められず、さらに実施する試験の性能を詳細に評価するために、第一種の過誤確率と比べてより少ない繰り返し回数でより広範なシナリオについて検討することが推奨される。

## 2.3 シンプルなベイズ流解析を用いた試験

本項では、シンプルなベイズ流解析を用いた事例として、2.1項で紹介した **Thall and Simon (1994)**の方法に基づく試験デザイン（以降、TS デザイン）の頻度論的動作特性の評価方法を概説する<sup>4</sup>。なお、2.1項でも触れたとおり、希少疾患の開発において TS デザインの適用事例があり、「希少疾患領域の臨床試験におけるベイズ流アプローチの適用に関する基本指針」では、「無情報事前分布を用いたとしても、事後分布を用いた治療効果の定量的評価中間モニタリングによる柔軟な意思決定が可能になる点で、ベイズ流アプローチは有用である」と言及している<sup>3,5</sup>。TS デザインは、必要サンプルサイズを固定する頻度流の方法と比較して、限られた登録期間でサンプルサイズの着地点に幅を持たせて柔軟に試験を実施することが可能であり、当該領域において強力な方法であると考えられる。以降、理論及び頻度論的動作特性とその実装方法について説明する。TS デザインの詳細は、**Thall and Simon (1994)** 及び **Yin (2012)**等を参照されたい<sup>4,13</sup>。

### 2.3.1 Thall and Simon の方法

TS デザインは、がん領域を対象とした抗腫瘍効果に対する二値アウトカムを評価するベイズ流適応的デザインであり、単群試験を前提に開発された手法である。本項では、がん領域における奏効の有無を評価項目として取り扱うが、二値アウトカムを対象とした単群試験を想定する場合、特に疾患領域は問わず適用できる。TS デザインは、標準治療（Standard therapy）と試験治療（Experimental drug）における奏効割合の事前分布にベータ分布を仮定する。次に、試験治療において単群臨床試験から得られたデータ（尤度）で奏効割合の事後分布を求める。最後に、標準治療の奏効割合の事前分布と試験治療の奏効割合の事後分布を比較する。

以降、具体的な数理について説明する。まず、 $p_E$ と $p_S$ をそれぞれ試験治療と標準治療の奏効割合とし、独立にベータ分布に従うものとする。

$$p_E \sim \text{Beta}(\alpha_E, \beta_E), \quad p_S \sim \text{Beta}(\alpha_S, \beta_S).$$

ここに、 $\alpha_E$ 、 $\beta_E$ 、 $\alpha_S$ 、 $\beta_S$ は試験計画時に設定する超パラメータである。標準治療に関しても参照する奏効割合の成績に不確実性を持たせて分布を仮定する点に注意する。組み入れ予定の最大被験者数 $N_{max}$ 例のうち、 $n$ 例時点で中間モニタリングを実施する場合に、 $X$ を $n$ 例の被験者に対する奏効例数とすると、確率変数 $X$ の従う確率分布は二項分布であり、

$X \sim \text{Bin}(n, p_E)$ である。ベータ分布と二項分布の共役性により、 $X = x$ を得た $p_E$ の事後分布もまたベータ分布に従い、

$$p_E | x \sim \text{Beta}(\alpha_E + x, \beta_E + n - x)$$

となる。また、標準治療に対する試験治療の奏効割合の最低許容増分を $\delta$ とすると、試験の成功確率は以下のように定義される。

$$\Pr(p_E > p_S + \delta | x).$$

$p_E$ と $p_S$ はそれぞれ独立であるから、 $p_E$ の事後分布と $p_S$ の事前分布に関する同時分布の確率密度関数は

$$f(p_E; \alpha_E + x, \beta_E + n - x) f(p_S; \alpha_S, \beta_S)$$

となる。ただし、 $f(p; \alpha, \beta)$ は $\text{Beta}(\alpha, \beta)$ の確率密度関数を表す。また、固定したある $p_S$ に対して $p_E > p_S + \delta$ となる確率は、

$$\begin{aligned} \int_{p_S + \delta}^1 f(p_E; \alpha_E + x, \beta_E + n - x) f(p_S; \alpha_S, \beta_S) dp_E \\ = f(p_S; \alpha_S, \beta_S) [1 - F(p_S + \delta; \alpha_E + x, \beta_E + n - x)] \end{aligned}$$

となる。ただし、 $F(p; \alpha, \beta) = \int_0^p f(x; \alpha, \beta) dx$ である。最後に、 $p_S$ に対して周辺化することで以下を得る。

$$\Pr(p_E > p_S + \delta | x) = \int_0^{1-\delta} f(p_S; \alpha_S, \beta_S) [1 - F(p_S + \delta; \alpha_E + x, \beta_E + n - x)] dp_S.$$

$\theta_U$ と $\theta_L$ をそれぞれ事前に規定した事後確率の目標値とし、以下のルールに基づいて $n$ 例時点の境界奏効例数を設定する。

- $\Pr(p_E > p_S + \delta | x) \geq \theta_U$ を満たす奏効例数 $x$ の最小の整数を $U_n$ とする。
- $\Pr(p_E > p_S + \delta | x) \leq \theta_L$ を満たす奏効例数 $x$ の最大の整数を $L_n$ とする。

中間モニタリング時点での中止規則を以下のように記述できる。この操作を予定最小例数 ( $N_{min}$ ) から $N_{max}$ まで評価を継続して意思決定する。

- $x \geq U_n$ の場合、試験治療は有望であると判断し、有効中止する。
- $x \leq L_n$ の場合、試験治療は有望でないと判断し、無益性中止する。
- $L_n < x < U_n$ かつ $n < N_{max}$ の場合、試験を継続して追加で被験者を組み入れる。

### 2.3.2 Case simple example

本項では、case example を与え、TS デザインを用いた試験デザインについて説明する。以下にデザインパラメータの概要を示す。

- 試験治療の奏効割合 ( $p_E$ ) の事前分布 :  $\text{Beta}(0.5, 0.5)$
- 標準治療の奏効割合 ( $p_S$ ) の事前分布 :  $\text{Beta}(34.4, 137.6)$
- 予定登録例数 ( $N_{max}$ ) : 15例
- 予定最小例数 ( $N_{min}$ ) : 10例
- 事後確率の目標値 ( $\theta_U$ ) : 95%
- 最低許容増分 ( $\delta$ ) : 0

標準治療の奏効割合の事前分布は Thall and Simon (1994)を参考に、平均が0.2で、かつベータ分布の5%点と95%点の幅が0.1となるような超パラメータを選択した<sup>12</sup>。試験治療の奏効割合の事前分布は Jeffreys の無情報事前分布を採用した。事後確率の目標値は、 $\theta_L$ を0%、 $\theta_U$ を95%とした。算出方法は別添の01\_TSdesign\_PriorElicitation.Rを参照のこと。

以上の設定をまとめた採用する TS デザインのシェーマを図 2-1に示す。

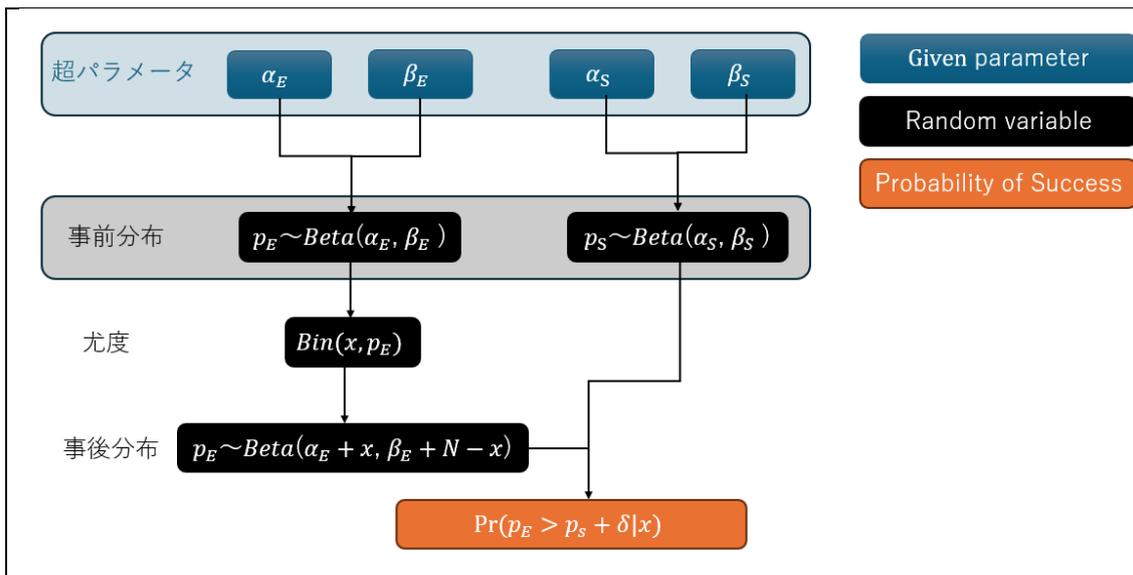


図 2-1 TS デザインのシェーマ

以上の設定の下、試験治療の奏効割合の事後分布 ( $p_E$ ) が標準治療の奏効割合の事前分布 ( $p_S$ ) を超える確率が95%以上となる必要最小奏効例数を以下に示す。

登録例数	10例	11例	12例	13例	14例	15例
必要最小奏効例数	5例	5例	5例	6例	6例	6例

例えば、10例時点での奏効例数が5例以上であれば試験治療は有効と判断し、試験を中止する。なお、上記の必要最小奏効例数は、ph2bayes package の stopbound\_post 関数を用いて容易に導出できる<sup>11</sup>。以下に算出のための R スクリプトを示す。

### R Script

```
n_max <- 15
alpha_e <- 0.5
beta_e <- 0.5
theta <- 0.95
alpha_s <- 34.4
beta_s <- 137.6

boundary_ <- ph2bayes::stopbound_post(theta = theta,
                                     type = "superiority",
                                     nmax = n_max,
                                     alpha_e = alpha_e,
                                     beta_e = beta_e,
                                     alpha_s = alpha_s,
                                     beta_s = beta_s,
```

```

delta = 0)
boundary <- dplyr::full_join(data.frame(n = c(1:n_max)), boundary_, by = "n") %>%
tidyr::fill(bound)
    
```

### 2.3.3 頻度論的動作特性の評価

2.3.2項で示したデザインの動作特性を評価する。具体的には、コンピュータシミュレーションを用いて帰無仮説及び対立仮説下での頻度論的動作特性を評価する。以下に動作特性に必要な設定を示す。

- 閾値奏効割合の真値 : 20%
- 期待奏効割合の真値 : 50%
- シミュレーション回数 : 100,000回
- 評価のタイミング : 1例ごと (すなわち、意思決定回数は計6回)
- 名目有意水準 : 10%

以上の設定の下で、指定したシミュレーション回数の臨床試験を仮想的に実施し、得られた結果に基づいて第一種の過誤確率、検出力、期待被験者数を評価する。特に、第一種の過誤確率は帰無仮説下で試験治療を有効と判断する確率、検出力は対立仮説下で試験治療を有効と判断する確率と定義する。以下に頻度論的動作特性の結果を示す。

帰無仮説		対立仮説	
平均被験者数	第一種の過誤確率	平均被験者数	検出力
14.69例	8.79%	11.21例	87.27%

シミュレーションに基づく第一種の過誤確率は8.79%であり、名目有意水準である10%を下回ることが確認できる。また、検出力も80%を超える十分高い確率を担保できていることから、検討した試験デザインの性能は許容範囲内であると判断できる。また、上記の結果を得るために必要となる R スクリプトを以下に示す。なお、境界例数は2.3.2項の結果を利用した。状況に応じてデータの生成並びに適用するベイズ流アプローチと成功の定義を変更する必要があるものの、基本的なシミュレーションのプロセスは以下の流れに基づく。

1. シミュレーション設定に基づき、仮定した確率分布に従う疑似乱数を生成し、1回の臨床試験データを生成する。
2. 1.のデータに対してベイズ流デザインを適用し、事後分布に基づく要約統計量並びに試験の成功の有無を評価する。
3. 上記を複数回繰り返し行い、得られた結果に基づいて第一種の過誤確率、検出力、推定値のバイアス等の動作特性の評価指標を算出する。

本章では、後述する複雑なベイズ流手法のシミュレーションの解説に先立ち、シンプルな評価事例を与えることを目的としている。このため、全体の構成を示すため全スクリプト

トを本文中に記載する。また、本資料の別添として SAS スクリプト (2\_3\_Thall and Simon.sas) も添付した。

### R script

```
N_min <- 10
boundary <- boundary # from "02_TSdesign_Boundary.R"
nsim <- 100000
p_H0 <- 0.2
p_H1 <- 0.5

TS_sim <- function(sim, p, boundary, n_min, n_max){
  success_FL <- 0
  endflg <- 0
  true_p <- p
  responder <- rbinom(1, n_min, true_p)
  act_num <- n_min

  if (responder >= boundary[n_min, ]$bound){
    success_FL <- 1
    endflg <- 1
  }else{
    while(all(endflg == 0, act_num < n_max)){
      x <- rbinom(1, 1, true_p)
      responder <- responder + x
      act_num <- act_num + 1
      if (responder >= boundary[act_num, ]$bound){
        success_FL <- 1
        endflg <- 1
        break()
      }else {
        endflg <- 0
      }
    }
  }
  }
  tidy::tibble(sim, true_p, act_num, responder, success_FL)
}

TS_OC <- function(nsim, theta, n_min, n_max, alpha_e, beta_e, alpha_s, beta_s, p_H0,
p_H1){
  result_H0 <- purrr::map2(c(1:nsim), p_H0, TS_sim, boundary = boundary,
n_min = n_min, n_max = n_max) %>% do.call(rbind,.)
  result_H1 <- purrr::map2(c(1:nsim), p_H1, TS_sim, boundary = boundary,
n_min = n_min, n_max = n_max) %>% do.call(rbind,.)
  #H0
  ASS_H0 <- mean(result_H0$act_num) #Average sample size
  alpha_err <- sum(result_H0$success_FL)/nsim #Type I error

  #H1
  ASS_H1 <- mean(result_H1$act_num) #Average sample size
  power <- sum(result_H1$success_FL)/nsim #Statistical power

  tidy::tibble(theta, n_min, n_max, alpha_e, beta_e, alpha_s, beta_s,
ASS_H0, alpha_err, ASS_H1, power)
}

result <- TS_OC(nsim = nsim, theta = theta, n_min = n_min, n_max = n_max,
alpha_e = alpha_e, beta_e = beta_e, alpha_s = alpha_s, beta_s = beta_s,
p_H0 = p_H0, p_H1 = p_H1)
result %<>% knitr::kable()

#Save results
capture.output(result, file = "./output/03_TSdesign_OC.txt")
```

- TS\_sim 関数：1回の臨床試験結果を得るための関数
- TS\_OC 関数：TS\_sim 関数を用いて100,000回の臨床試験を繰り返し実行し、結果を要約するための関数

### 2.3.4 まとめ

本項では、シンプルなベイズ流解析を用いた事例として TS デザインを取り上げ、頻度論的動作特性の評価方法を説明した。試験治療における事前分布は無情報事前分布を用い

たが、このような場合でもベイズ流アプローチの柔軟性が享受できるという点で、登録が困難な疾患領域において活用の幅は多岐にわたることが期待される。一方、事前分布に情報を持たせる場合は、得られた結果に対するバイアスの程度を追加で評価する必要があるだろう。後述の項でそのようなケースにおけるベイズ流アプローチのシミュレーション方法についても触れる。

### 2.3.5 プログラムの利用方法

本項で用意したプログラム及び実行結果は、すべて2\_3\_TSdesign フォルダに格納している。また、Rの開発環境は以下の操作により復元可能である。

1. R コンソール上で”install.packages(“renv”, type = “win.binary”)”を実行し、renv package をインストールする。
2. ディレクトリパスに2バイト文字を含まない環境で zip ファイルを展開する。
3. Rproj ファイルをダブルクリックし、RStudio を起動する。
4. R コンソール上で”renv::restore()”を実行する。

## 2.4 参考文献

1. Food and Drug Administration. (2010). Guidance for the Use of Bayesian Statistics in Medical Device in Clinical Trials. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-use-bayesian-statistics-medical-device-clinical-trials>
2. Best, N., Ajimi, M., Neuenschwander, B., Hilary, GS. and Wandel, S. (2023). Beyond the Classical Type I Error: Bayesian Metrics for Bayesian Designs Using Informative Priors. Statistics in Biopharmaceutical Research, 00.0, 1–14.
3. Hirakawa, A., Nishikawa, T., Yonemori, K., Shibata, T., Nakamura, K., Ando, M., Ueda, T., Ozaki, T., Tamura, K., Kawai, A. and Fujiwara, Y. (2018). Utility of Bayesian single-arm design in new drug application for rare cancers in Japan: A case study of phase 2 trial for sarcoma. Therapeutic Innovation & Regulatory Science, 52(3), 334–338.
4. Thall, P. F. and Simon, R. (1994). Practical Bayesian guidelines for phase IIB clinical trials. Biometrics, 50(2), 337–349.
5. AMED 医薬品等規制調和・評価研究事業. (2021). 希少疾患領域の臨床試験におけるベイズ流アプローチの適用に関する基本指針. [https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS\\_202302\\_hirakawa1.pdf](https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS_202302_hirakawa1.pdf)
6. 手良向 聡. (2023). ベイズ流臨床試験における標本サイズ設定：2つの事前分布を用いた推論上の性能に基づく接近法. 計量生物学, 44(1), 34-51.
7. Food and Drug Administration. (2020). Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/interacting-fda-complex-innovative-trial-designs-drugs-and-biological-products>
8. 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. (2024) . Complex

Innovative Trial Design の現状と事例の整理.

9. Food and Drug Administration. (2019). Adaptive Design Clinical Trials for Drugs and Biologics Guidance for Industry. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics-guidance-industry>
10. 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. (2021) . アダプティブデザインに関する FDA ガイダンスの邦訳.
11. Nagashima, K. (2018). `_ph2bayes: Bayesian Single-Arm Phase II Designs_`. R package version 0.0.2.
12. Thall, P. F. and Simon, R. (1994). A Bayesian approach to establishing sample size and monitoring criteria for phase II clinical trials. *Controlled Clinical Trials*, 15(6), 463-481.
13. Yin, G. (2012). Bayesian and frequentist adaptive methods. (手良向 聡, 大門 貴志 (訳者) (2014) . 臨床試験デザインーベイズ流・頻度流の適応的方法ーメディカルパブリケーションズ)

### 3. 外部データの利用

医薬品の開発において、実際の医療環境下で取得されたデータを含む、実施する試験以外のデータソースから得られたデータの利活用を試みる動きが活発化している<sup>1</sup>。例えば、ICH E8 (R1) 臨床試験の一般指針にて、過去の臨床研究、全国的な死亡データベース、疾患及び薬剤レジストリ、医療保険請求データ、並びに日常的な医療行為からの医療及び管理の記録を含む外部データの二次利用について記載され、試験計画時に考慮すべき内容が明記された<sup>2</sup>。

外部データの利用は、生命を脅かす希少疾患や、不可逆的な転帰を辿る疾患、小児医薬品開発等、同時対照群の設定が倫理的な観点から困難な場合に薬効を評価する方法の1つとして考えられている<sup>3</sup>。

小児医薬品開発はこの外部データの利用が特に注目されている領域の1つである。小児集団は脆弱なサブグループであり、過度の危険から保護する特別な配慮が必要な集団である<sup>4,5</sup>。成人試験データ等を小児試験データに外挿することにより、小児集団全体への曝露を少しでも減らし、かつ効率的な小児医薬品開発が行えることが期待できる。特に、小児用医薬品開発における外挿に関するガイドラインである ICH-E11A (2025/1 時点 Step 4) では、mixture prior 等のベイズ流アプローチについても言及しており、小児医薬品開発における外部データの利用に対する期待はますます高まっている<sup>6</sup>。

外部データの利用に関して、ICH-E10ガイドラインでは、臨床試験における対照群の種類の一つとして試験に参加して治療を受ける集団とは別の対照集団（外部対照群）を設けることが述べられている<sup>7</sup>。臨床試験での外部対照の利用方法については、単群試験の外部対照群とすることや、事前に定義したアルゴリズムやマッチング等の統計手法を用いて、二重盲検試験の対照群と併合することが考えられる。このように実施しようとする臨床試験のデザイン・解析に外部データを対照群として利用する考え方をヒストリカルコントロールアプローチと呼び、その中でも後者の二重盲検試験の対照群と外部データを併合して対照群を構成する考え方を hybrid control アプローチと呼ぶ<sup>8</sup>。hybrid control アプローチの実装には、特にベイズ流アプローチが用いられる。

ベイズ流アプローチによる外部データの利用は、既存の情報（専門家の意見、ヒストリカルデータ等）を事前分布として導入し、情報のより効率的な利用を通じてサンプルサイズを大幅に削減、もしくは統計的検出力を向上できる可能性を提供する。この結果、試験実施にかかるコストの削減、及び対照群への割付を減らすといった倫理的側面への考慮が期待される。一方で、外部データの利用においてはいくつか課題も存在する。外部データとして利用するデータは、二次利用を前提とせずに設計・運用されていることも多く、実施する試験の目的に十分に合致しないことがある。また外部データとして利用したい参照集団と対象となる集団の治療効果が十分に類似していない状況も存在する。このような場合、データの質や信頼性等の観点から様々な潜在バイアスの存在が懸念される。これらのバイアスに対処するために、外部データを利用する試験計画時には、バイアスや第一種の過誤確率を含む試験の動作特性をシミュレーション実験にて評価する必要がある。

3.1章では、成人試験の要約データを用いて治療効果の外挿を行った小児 PAH (pulmonary

arterial hypertension：肺動脈性肺高血圧症）患者を対象とした第 III 相ランダム化二重盲検プラセボ対照の並行群間試験を、3.2章では、計画された試験の対照群データと外部対照のデータを用いた hybrid control アプローチが検討されたびまん性大細胞型 B 細胞リンパ腫を対象とした臨床試験を参考に、ベイズ流アプローチを用いる試験の動作特性をシミュレーションにて検討する方法について紹介する。

### 3.1 小児外挿の場合

本項では、成人試験の要約データを用いて治療効果の外挿を行った小児試験の事例である Ivy et al. (2021)<sup>9, 10</sup>を参考に、動作特性のシミュレーションによる検討方法について紹介する。

Ivy et al. (2021) で示された臨床試験は小児 PAH（pulmonary arterial hypertension：肺動脈性肺高血圧症）患者を対象とした第 III 相ランダム化二重盲検プラセボ対照の並行群間試験（NCT01824290）<sup>11</sup>（以下、本試験）である。本試験は6~17歳の小児患者を対象としており、tadalafil 群と placebo 群に1:1で割り付けられる。試験開始時の目標被験者数は検出力80%を確保するために各群67例と設定していたが、登録状況が芳しくなかったため試験開始後に各群17例に変更されている。主要評価項目は24週時の6分間歩行距離のベースラインからの変化量である。主解析は MMRM を用いた群間の比較を計画していたが、被験者数不足により十分な検出力が確保できていない。これに伴い、実施済みの成人臨床試験において、本試験の小児患者と背景情報が類似している成人患者データの要約データを外部データとして利用した Bayesian MMRM を行うように計画を変更している。Bayesian MMRM で用いる事前分布として本試験では懐疑的な設定の skeptical prior と、実施済みの成人臨床試験結果に基づいて構成された adult prior の混合分布に基づく mixture prior を採用している<sup>12</sup>。Mixture prior を採用することで、成人臨床試験との類似性が低い場合に、成人試験データ単体に基づいて構成された prior を用いる場合よりも type I error rate のインフレを抑えることが期待できる。

#### 3.1.1 本試験のデザイン

試験標題	肺動脈性肺高血圧症の小児第 III 相試験
目的	ERA（エンドセリン受容体拮抗薬）による治療を受けている、肺動脈性肺高血圧症6~17歳の小児患者を対象とした tadalafil の有効性の検証
試験期間	30ヶ月（二重盲検期 6ヶ月、Open-label 期 24ヶ月）
投与群	プラセボ群、試験薬群
割付比	1:1
主要評価項目	24週時の6分間歩行距離のベースラインからの変化量
解析方法	主解析：MMRM 補足的解析：Bayesian MMRM

### 3.1.2 解析方法の詳細

本項では補足的解析である Bayesian MMRM の解析方法について詳細を記述する。

解析モデル：

時点、投与群と時点の交互作用を固定効果とし、共分散構造を **unstructured** と仮定する以下のモデルを考える。

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

ここで、 $\mathbf{y}_i$ は被験者*i*の応答変数ベクトル、 $\mathbf{X}_i$ は被験者*i*のデザイン行列で、要素にカテゴリカルな説明変数の水準を表す定数か量的な説明変数の測定値を含んでいる。 $\boldsymbol{\beta}$ はパラメータベクトルであり、治療効果に関連のある $\boldsymbol{\beta}_{rt}$ と治療効果に関連のない $\boldsymbol{\beta}_{nrt}$ で構成されている。誤差ベクトル $\boldsymbol{\varepsilon}_i$ は被験者*i*の各時点の誤差 ( $\boldsymbol{\varepsilon}_i = [\boldsymbol{\varepsilon}_{i1}, \dots, \boldsymbol{\varepsilon}_{iT}]$ ) であり、平均0、共分散行列 $\boldsymbol{\Sigma}$ をもつ多変量正規分布に従うと仮定する。

事前分布：

共分散行列 $\boldsymbol{\Sigma}$ の事前分布として自由度*T*、スケールパラメータ1の逆ウィシャート分布 **Inv - Wishart<sub>T</sub>(I)**を用いる。

治療効果に関連のない項である $\boldsymbol{\beta}_{nrt}$  (Ivy et al. (2021)では時点の項) については各々独立に適当な無情報事前分布を設定する。

成人データ借用のための治療効果の関連のある項である $\boldsymbol{\beta}_{rt}$  (Ivy et al. (2021)では投与群と時点の交互作用項) については **mixture prior** を用いる。

**Skeptical prior** は以下の多変量正規分布で構成される。

$$N(0, \boldsymbol{\Sigma}_S) = N\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_S & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_S \end{bmatrix}\right)$$

**Skeptical prior** は小児では全く効果がなく、時点間の相関もないことを表している事前分布となる。これは **type I error rate** をコントロールするためのコンポーネントであり、**type I error rate** を妥当なレベルに抑えるために、シミュレーションに基づいて適切な値が設定される。本試験では、 $\sigma_S$ が9から100の範囲で検討され、以下のシミュレーション実施時には $\sigma_S = 9$ と設定されている。

また、**adult prior** は以下の多変量正規分布で構成される。

$$N(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_A) = N\left(\begin{bmatrix} \mu_{A1} \\ \vdots \\ \mu_{AT} \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1T} \\ \vdots & \ddots & \vdots \\ \sigma_{1T} & \cdots & \sigma_{TT} \end{bmatrix}\right)$$

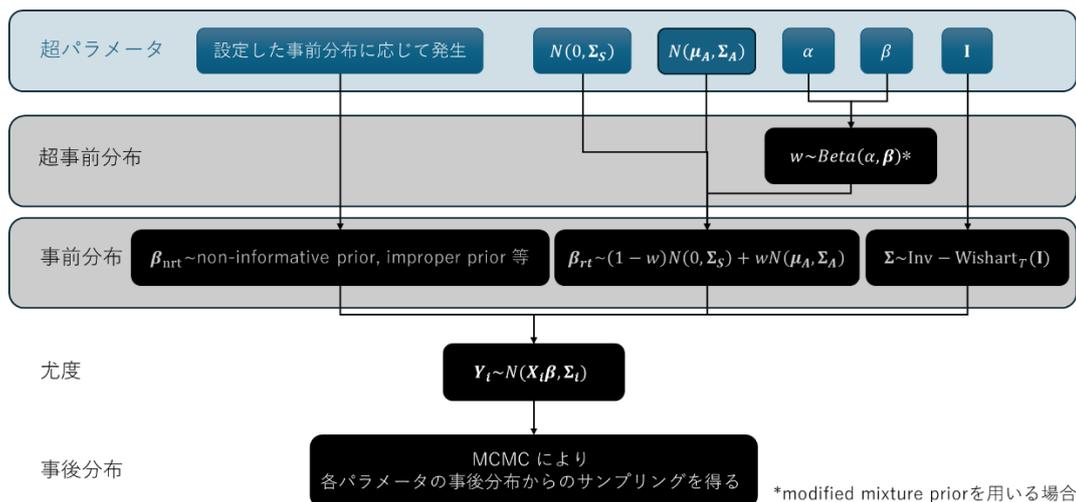
ここで $\boldsymbol{\mu}_A$ 及び $\boldsymbol{\Sigma}_A$ は実施済みの成人臨床試験結果に基づいて形成される。

上記2つのコンポーネントに基づき、 $\boldsymbol{\beta}_{rt}$ の事前分布を以下のように設定する。

$$\boldsymbol{\beta}_{rt} \sim (1 - w)N(0, \boldsymbol{\Sigma}_S) + wN(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_A)$$

混合割合 $w$ は任意の値を用いる。なお Ivy et al. (2021)では複数の医学専門家へアンケートを行い、その結果に基づき重み0.8を用いている。また、本稿では  $w$  を確率変数とみなし、事前分布を設定する場合についても検討する（以下、modified mixture prior）。 $w$  の事前分布はベータ分布を用いることとする。

上記モデルの理解のため、各モデル及びパラメータの関係を下図で示した。



Decision criteria :

Ivy et al. (2021) では主要評価項目である24週時の投与群と時点の交互作用項のパラメータ $\beta_T$ について以下の判定基準を用いている。

$$\Pr(\beta_T > 0) > 0.95$$

### 3.1.3 シミュレーション

#### 3.1.3.1 シミュレーションの目的

本シミュレーションでは、mixture prior を用いた Bayesian MMRM の動作特性を確認するため、複数のシナリオを設定し、条件設定に基づき type I error rate、検出力、バイアス、MSE、ESS を用いた性能評価を行う。

#### 3.1.3.2 シミュレーションの設定

小児の仮想データを作成してシミュレーションを実施した。

小児の治療効果について以下の3つのシナリオを設定した。

1. Null : プラセボと同じ効果であった場合
2. 50% effect of optimistic : 成人の半分程度の効果があった場合
3. Optimistic : 成人と同程度の効果があった場合

各シナリオにおける共分散行列は同一のものを仮定する。それぞれのシナリオの下で、ベースライン時点を含まない4時点分の変化量のデータを多変量正規分布に従う乱数を用いて発生させた。被験者数は各群17例、51例の2通り設定した。この時、実薬群の多変量

正規分布の平均ベクトルと共分散行列は Ivy et al. (2021)の3、5時点目を除いたものを用いている。Optimistic なシナリオは以下の多変量正規分布に基づき実薬群のシミュレーションデータを発生させている。また、50% effect of optimistic のシナリオでは optimistic なシナリオの平均ベクトルの各値を半分に、null のシナリオ及びプラセボ群では全て0に変更している。

$$N\left(\begin{bmatrix} 8 \\ 15 \\ 21 \\ 24 \end{bmatrix}, \begin{bmatrix} 2078 & 1463 & 1075 & 950 \\ 1463 & 2266 & 1543 & 1450 \\ 1075 & 1543 & 2858 & 1950 \\ 950 & 1450 & 1950 & 2858 \end{bmatrix}\right)$$

解析モデルとして時点、投与群と時点の交互作用を固定効果とし、共分散構造を unstructured と仮定するモデルを用いる。誤差ベクトル  $\boldsymbol{\varepsilon}_i$  の共分散行列  $\boldsymbol{\Sigma}$  の事前分布として自由度4、スケールパラメータ1の逆ウィシャート分布  $\text{Inv-Wishart}_4(\mathbf{I})$  を用いる。治療効果に関連のない項（時点）の  $\boldsymbol{\beta}_{nrt}$  の事前分布は improper prior として区間  $[-\infty, \infty]$  の一様分布を用いる。治療効果に関連のある項（投与群と時点の交互作用）の  $\boldsymbol{\beta}_{rt}$  は skeptical prior と adult prior から構成される mixture prior を事前分布として用いる。mixture prior は混合割合  $w$  を0.5、0.8（図3-1参照）とするものに加え、 $w$  に事前分布を設定する modified mixture prior の計3通りを設定した。各事前分布の詳細は以下の通りとする。

Skeptical prior:

$$N(0, \boldsymbol{\Sigma}_S) = N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 81 & 0 & 0 & 0 \\ 0 & 81 & 0 & 0 \\ 0 & 0 & 81 & 0 \\ 0 & 0 & 0 & 81 \end{bmatrix}\right)$$

Adult prior:

$$N(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_A) = N\left(\begin{bmatrix} 8 \\ 15 \\ 21 \\ 24 \end{bmatrix}, \begin{bmatrix} 67 & 69 & 56 & 45 \\ 69 & 78 & 71 & 63 \\ 56 & 71 & 78 & 79 \\ 45 & 63 & 79 & 90 \end{bmatrix}\right)$$

混合割合  $w$  (modified mixture prior の場合のみ):

$$w \sim \text{Beta}(0.5, 0.5)$$

Mixture prior/modified mixture prior:

$$\boldsymbol{\beta}_{rt} \sim (1 - w)N(0, \boldsymbol{\Sigma}_S) + wN(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_A)$$

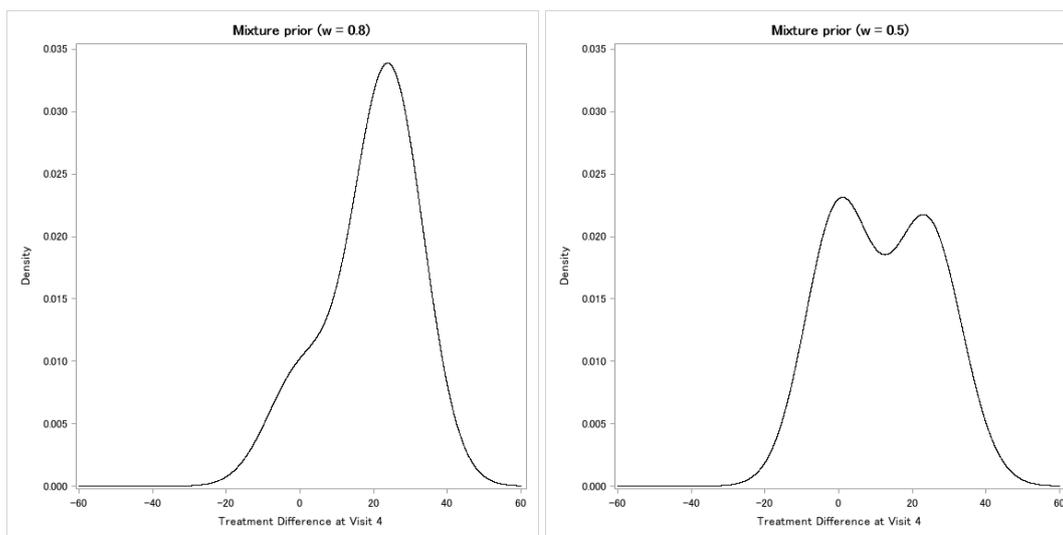


図 3-1 4時点目の治療効果の mixture prior (左図 : w=0.8 右図 : w=0.5)

$\beta_{rt}$ の事前分布として、上述の3通りに加え、成人試験のデータを利用しない想定として improper prior を用いた場合の計4通りの条件を設定した。シナリオ3通り×被験者数2通り×事前分布4通りの合計24通りのシミュレーションをそれぞれ3000回実施した。得られた事後分布から、最終時点である4時点目において実薬群がプラセボ群よりも変化量が大きい確率を求め、decision criteria の0.95を超えた割合を type I error rate (シナリオ1)あるいは検出力(シナリオ2、3)とする。

ESS は以下の式により得られる<sup>13</sup>。

$$ESS = N \left( \frac{V_1(\beta_T)}{V_2(\beta_T)} - 1 \right)$$

ここでNは新規試験の被験者数、 $V_1(\beta_T)$ は improper prior を用いて算出した $\beta_T$ の分散、 $V_2(\beta_T)$ は mixture prior を用いて算出した $\beta_T$ の分散である。

### 3.1.3.3 シミュレーションの結果

3.1.3.2で記載したシミュレーションの結果を以下の表に示す。

・ Type I error rate、検出力

被験者数	シナリオ	improper prior	mixture prior w=0.5	mixture prior w=0.8	modified mixture prior
17例/群	null	5.0%	6.1%	21.2%	6.6%
	50% effect of optimistic	15.3%	17.9%	40.5%	18.1%
	optimistic	37.4%	36.3%	65.1%	39.1%
51例/群	null	4.9%	7.8%	17.4%	8.8%
	50% effect of optimistic	30.1%	37.8%	58.6%	37.6%
	optimistic	72.6%	77.1%	90.6%	78.1%

・ バイアス

被験者数	シナリオ	improper prior	mixture prior w=0.5	mixture prior w=0.8	modified mixture prior
------	------	----------------	---------------------	---------------------	------------------------

17例/群	null	-0.3	7.0	12.1	7.2
	50% effect of optimistic	-0.4	0.3	4.7	0.3
	optimistic	0.2	-6.3	-2.6	-6.3
51例/群	null	0.0	4.5	7.3	4.7
	50% effect of optimistic	0.1	1.1	3.9	1.1
	optimistic	0.0	-2.4	-0.9	-2.4

・ MSE

被験者数	シナリオ	improper prior	mixture prior w=0.5	mixture prior w=0.8	modified mixture prior
17例/群	null	323.2	113.4	210.2	116.1
	50% effect of optimistic	337.3	76.1	81.9	77.8
	optimistic	341.6	111.7	55.4	116.6
51例/群	null	113.2	71.2	108.7	75.6
	50% effect of optimistic	109.0	66.6	64.1	67.3
	optimistic	114.5	58.5	35.4	57.9

・ ESS

被験者数	シナリオ	improper prior	mixture prior w=0.5	mixture prior w=0.8	modified mixture prior
17例/群	null	-	70.4	67.4	71.4
	50% effect of optimistic	-	62.6	75.2	65.0
	optimistic	-	61.2	86.4	65.0
51例/群	null	-	78.2	64.4	80.2
	50% effect of optimistic	-	55.0	76.5	60.2
	optimistic	-	75.7	106.2	88.1

### 3.1.3.4 シミュレーションの結果の考察

Type I error rate について、成人試験のデータを利用しない improper prior を用いた場合は名義水準（5%）通りの値となったが、成人試験のデータを外挿した事前分布を用いた場合は被験者数に依らず名義水準を超える傾向となった。 $w = 0.5$  の mixture prior 及び modified mixture prior を用いた場合は6.1～8.8%となり、 $w = 0.8$  の mixture prior の場合は17.4～21.2%と大きなインフレが見られた。

これは成人試験と小児試験の治療効果に差異が生じたことによるものと考えられ、成人試験の利用の程度が大きいほどその傾向がより顕著に出ている。また、これは被験者数に依らず同様の傾向であった。

検出力について、全てのシナリオ及び被験者数で improper prior が最も低かった。mixture prior を用いた場合はいずれの場合でも improper prior と比較して検出力が増加した。 $w = 0.5$  の mixture prior 及び modified mixture prior はいずれのシナリオ、被験者数でも同程度の検出力であった。 $w = 0.8$  の mixture prior を用いた場合が最も検出力が大きく、特に optimistic のシナリオで検出力が大きく増加していた。このことから、事前分布の成人試験の利用の程度が大きい、または成人試験と小児試験の治療効果の差異が小さい場合に検出力が増加することが示唆された。

バイアスについて、improper prior はほぼ0であったが、mixture prior では improper prior と比べてバイアスが大きくなる傾向が見られた。mixture prior の場合は事前分布の期待値と

小児試験の治療効果の差異が小さい場合にバイアスが小さく、反対に乖離が生じている場合にはバイアスが大きくなる傾向となることが示された。

MSE について、全てのシナリオ及び被験者数で *improper prior* が最も大きかった。*mixture prior* の場合は成人試験と小児試験の治療効果の差異が小さい場合に MSE が小さく、反対に乖離が生じている場合には MSE が大きくなる傾向となることが示された。

ESS について、 $w = 0.8$  の *mixture prior* では成人試験と小児試験の治療効果の差異が小さいほど ESS は大きくなった。一方で  $w = 0.5$  の *mixture prior* 及び *modified mixture prior* では同様の傾向が見られず、類似性が高い 50% effect of optimistic のシナリオでの ESS は他のシナリオと同程度か小さい値となった。 $w = 0.5$  の *mixture prior* 及び *modified mixture prior* を用いた場合の事後分布の分散は他の事後分布の分散より大きく、これにより類似していることによる ESS の増大を打ち消しているためと考えられる。

事前分布に成人試験のデータを外挿したことで検出力が大きくなることが示されたが、一方でその利用の程度が大きい場合 *type I error rate* の顕著な増加が見られ、*mixture prior* の利用に際しては検出力と *type I error rate* のバランスを十分に評価、検討することが必要となる。

*Mixture prior* は外部試験の要約データの情報を事前分布に組み込むベイズ流アプローチであり、個別データを必要としないこともメリットの一つであると考えられる。一方で個別データを用いることができる場合は、3.2章の CMP 法等の異なるベイズ流アプローチも利用可能であるため、本手法の実装を検討する際には他の手法との比較検討も併せて行うことができる。

### 3.1.3.5 R を用いたコードのポイント

*Brms* パッケージ<sup>14,15,16</sup>は *Stan* を用いて Bayesian マルチレベルモデルを実装するための R パッケージであり、幅広い分布やリンク関数に対応している。また、それをさらに MMRM に特化させた *brms.mmrm* パッケージ<sup>17</sup>も存在する。

今回参照した論文のモデルは *brms.mmrm* や *brms* パッケージ内の関数のみでは実装が困難であった。このような場合では、*brms* パッケージ内の関数で MMRM モデルを *Stan* コードで書いたものを出力させ、それを書き換えたものを *rstan* パッケージ<sup>18</sup>でシミュレーションを実施することが対応案の一つとして挙げられる。

今回使用したパッケージは以下の通りである。

```
library(MASS) #多変量正規分布で乱数を発生させるために使用
library(tidyverse) #データ成形用
library(brms)
library(rstan)
library(bayesplot) #結果確認用
```

`brms::bf` で *Stan* コードを出力させたいモデルを読み込ませる。モデルは3.1.3.2章で示した通りである。

#### #モデルの設定

```
formula <- bf(value ~ 1 + time * trt, autocor = ~unstr(time=time, gr=subject_id))
```

次に、`brms::set_prior` を用いて各固定効果パラメータや共分散行列に事前分布を設定する。前述の通り、実際の事前分布の設定は後の Stan コードを書き換えるステップで対応するため、ここでは仮のものを設定している。

#### #事前分布の設定

```
priors <- c(
  set_prior("normal(0,1e10)", class = "b",coef = "time2"),
  set_prior("normal(0,1e10)", class = "b",coef = "time3"),
  set_prior("normal(0,1e10)", class = "b",coef = "time4"),
  set_prior("normal(9,81)", class = "b", coef = "trt"),
  set_prior("normal(9,81)", class = "b", coef = "time2:trt"),
  set_prior("normal(9,81)", class = "b", coef = "time3:trt"),
  set_prior("normal(9,81)", class = "b", coef = "time4:trt"),
  set_prior("lkj(1)", class = "Lcortime")
)
```

以上の `formula`、`priors` と小児の仮想データ(`data1`)を用いて、`brms::stancode` で Stan コードを生成することができる。ここで生成したコードを直接編集し、事前分布形成に使用するパラメータの情報の追加や、事前分布の形の修正を行う。以下では読み込ませた `data` ブロックに混合分布のパラメータの情報を追加し、`transformed parameters` ブロック内の交互作用の固定パラメータと切片の事前分布の形を修正するコード例を示す。書き換えた `stan` コードを `stan_code2` に格納し、さらにそれを `rstan::stan_model` を用いて `stanmodel2` に格納した。

#### #stan\_code:仮のモデルの stan コードを格納

```
stan_code <- stancode(
  formula,
  data = data1,
  family = gaussian,
  prior = priors
)
#stancode2:stan_code を元にモデルを改造
stanmodel2 <- stan_model(model_code = stan_code2)
```

たとえば `data` ブロックでは以下のように適切なデータ型を用いて必要なパラメータを追記している。

```
vector[n unique t] mean1;
vector[n unique t] mean2;
cov matrix[n unique t] var1;
```

```
cov_matrix[n_unique_t] var2;
real w;
```

また、`transformed parameters` ブロックでは、上記の仮で設定した事前分布は以下のように出力される。今回のような実装例であれば、黄色ハイライト箇所を以下のように書き換える。

#### 【書き換え前】

```
transformed parameters {
  real lprior = 0; // prior contributions to the log posterior

  lprior += normal_lpdf(b[1] | 0,1e10); // diffuse prior
  lprior += normal_lpdf(b[2] | 0,1e10); // diffuse prior
  lprior += normal_lpdf(b[3] | 0,1e10); // diffuse prior
  lprior += normal_lpdf(b[4] | 9,81);
  lprior += normal_lpdf(b[5] | 9,81);
  lprior += normal_lpdf(b[6] | 9,81);
  lprior += normal_lpdf(b[7] | 9,81);
  lprior += student_t_lpdf(Intercept | 3, 0.3, 50.3);
  lprior += lkj_corr_cholesky_lpdf(Lcortime | 1);
  lprior += student_t_lpdf(sigma | 3, 0, 52.7)
  - 1 * student_t_lccdf(0 | 3, 0, 52.7);
}
```

#### 【書き換え後例】

```
transformed parameters {
  real lprior = 0; // prior contributions to the log posterior

  lprior += normal_lpdf(b[1] | 0,1e10); // diffuse prior
  lprior += normal_lpdf(b[2] | 0,1e10); // diffuse prior
  lprior += normal_lpdf(b[3] | 0,1e10); // diffuse prior
  lprior += log_mix(w, multi_normal_lpdf(b[4:7] | mean1,var1), multi_normal_lpdf(b[4:7]
  | mean2,var2));
  lprior += normal_lpdf(Intercept | 0,100);
  lprior += lkj_corr_cholesky_lpdf(Lcortime | 1);
  lprior += student_t_lpdf(sigma | 3, 0, 52.7)
  - 1 * student_t_lccdf(0 | 3, 0, 52.7);
}
```

さらに、`rstan` パッケージで MCMC を実行する準備として、`Stan` コードを用いたシミュレーションに対応できるデータとモデルの情報を `brms::standata` を用いて読み込ませた `stan_data` を作成する。元のデータには含まれないが `Stan` コード内で使用したい値はここ

で作成した `stan_data` に追加する必要がある。ここでは混合分布のパラメータの情報と、サンプリングする際にデータを使用するか、事前情報のみでサンプリングするかを選べる `prior_only` の値を指定している。

```
#データを stan で使えるようにする
stan_data<-standata(
  formula,
  data = data1,
  family = gaussian
)
#データ以外で使いたい情報を stan_data に組み込む*stan コードへの追記と併せて必要
stan_data$mean1 <- mean1
stan_data$mean2 <- mean2
stan_data$var1 <- var1
stan_data$var2 <- var2
stan_data$w <- 0.8 #重み
stan_data$prior_only <- 0 #1 :prior only で sampling
```

以上より、`stanmodel2`と `stan_data` を用いて、`rstan::sampling` 関数で MCMC を実行することができる。

```
fit_custom <- rstan::sampling(stanmodel2, data= stan_data,
                             iter=200000, warmup=10000, thin=50, chain =1, cores=1)
```

### 3.1.3.6 SAS を用いたコードのポイント

本章では SAS の MCMC プロシジャによる実装例も併せて示すこととする。Bayesian MMRM の基本的な実装については伊庭ら(2023)<sup>19</sup>を参考されたい。ここでは mixture prior ( $w = 0.8$ ) の実装部分についてのみ抜粋して説明する。

```
/*adult component of the mixture prior*/
array muA[4] muA1-muA4 (8 15 21 24);
array SA[4, 4] SA1-SA16 (67 69 56 45,
                        69 78 71 63,
                        56 71 78 79,
                        45 63 79 90);

/*skeptical component of the mixture prior*/
array muS[4] muS1-muS4 (0 0 0 0);
array SS[4, 4] SS1-SS16 (81 0 0 0,
                        0 81 0 0,
                        0 0 81 0,
                        0 0 0 81);

/*MCMCプロシジャ内でmixture priorの計算用の配列を用意*/
array DIFA[4];
array tDIFA[1, 4];
array invSA[4, 4];
array XMA[1, 4];
array XMXA[1];
```

```

array DIFS[4];
array tDIFS[1, 4];
array invSS[4, 4];
array XMS[1, 4];
array XMXS[1];

/*CALLルーチンを用いて各配列（事前分布の算出に用いる）の初期値を設定*/
begincnst;
  call identity(S);
  call det(SA, dSA);
  call inv(SA, invSA);
  call det(SS, dSS);
  call inv(SS, invSS);
endcnst;

/*規定されていない事前分布を設定した場合は初期を設定する必要あり。beta_tmはmixture
priorを事前分布とするパラメータ群の配列*/
parms beta_tm {8 15 21 24};

/*mixture prior の計算。CALL ルーチンを用いている。*/
beginndata;
  n=4;
  w=0.8;
  call subtractmatrix(beta_tm, muA, DIFA);
  call transpose(DIFA, tDIFA);
  call mult(tDIFA, invSA, XMA);
  call mult(XMA, DIFA, XMXA);

  call subtractmatrix(beta_tm, muS, DIFS);
  call transpose(DIFS, tDIFS);
  call mult(tDIFS, invSS, XMS);
  call mult(XMS, DIFS, XMXS);

  const adlut=1/((CONSTANT('PI')**(n/2))*(dSA**(1/2)));
  adlut_pdf=const_adlut*exp(-0.5*XMXA[1]);

  const skeptical=1/((CONSTANT('PI')**(n/2))*(dSS**(1/2)));
  skeptical_pdf=const_skeptical*exp(-0.5*XMXS[1]);

  lp=log(w*adlut_pdf + (1-w)*skeptical_pdf);
endndata;

/*上記で導出したmixture priorを事前分布として指定*/
prior beta_tm ~ general(lp);

/*他の事前分布を指定*/
/*general(0)は区間[-∞,∞]の一様分布に相当*/
prior beta_m: ~ general(0);
prior R ~ iwish(4, S);

/*解析モデル及び推定値の指定（今回のモデルではlsmの指定は不要だが拡張性も考慮して設
定している）*/
beginndata;
do j=1 to 4;

```

```

lsm1[j] = beta_m[j] + beta_tm[j];
lsm2[j] = beta_m[j];
lsmd[j] = beta_tm[j];
end;
endnodata;

do j=1 to 4;
  if trt=1
    then mu[j] = beta_m[j] + beta_tm[j];
    else mu[j] = beta_m[j];
  end;

  model y ~ mvn(mu, R);

```

## 3.2 外部対照の場合

### 3.2.1 想定される試験の状況

ICH-E10ガイドラインでは、外部対照デザインの使用は、治療効果が劇的であり、疾患の通常の経過が十分に予測可能である場合に限定されている。さらに、外部対照を採用するのは、エンドポイントが客観的であり、評価項目に対するベースラインや治療変数の影響の特徴が十分に分かっている場合に限ることが述べられている。さらに新規試験の対照群の一部として利用するデータを提供可能な既存試験の条件として Pocock (1976)<sup>20</sup>は6つの条件を提唱している。しかし提唱された条件を全て満たす外部対照データが存在することは現実的にはほとんどない。そこで、全ての条件が厳密に満たされていない状況でヒストリカルデータの情報を活用する有効な方法を考える必要がある<sup>21</sup>。単一の外部対照データしかない場合、全ての条件が満たされていない状況下だと、新規試験の対照群と外部対照データの間には異質性が生じる可能性が高く、試験間で結果が変動 (drift) してしまう場合、信頼性をもって治療効果を推定することが難しい。その場合、解析担当者が事前に drift に対する情報を付与する必要がある、これが検出力や第一種の過誤確率に大きな影響を与える。

Hobbs (2011)<sup>22</sup>らは、外部対照データと新規試験の対照群の類似性をパラメータ化し、外部対照データと新規試験の対照群の共通性 (commensurability) に応じて、借用の程度を調整する方法を提案している (commensurate prior 法。以下、CMP 法)。

本項で取り上げる試験は、治療効果が時間の経過に基づき研究間で "drift" してしまうことを想定、つまり外部対照と計画している臨床試験における対照群の疾患の経過が異なることを想定し、その "drift" の存在を考慮しながら外部対照データを利用することを検討した試験である。CMP 法は臨床試験における同時対照患者と外部対照患者の反応が一貫しているか、どの程度異なるかを確認し、その類似性に応じて外部対照データから借用する情報量を決定し、解析に反映させることができる。3.2.2には本シミュレーション実験に参考とした試験デザインとその解析方法を、3.2.3にはシミュレーション実験の計画とその結果を、3.2.4にはシミュレーション実験の実施に用いた R package の紹介とその実行例を記載した。3.2.5には、データ解析に焦点を当てた simple example を与えおり、CMP 法の詳細な解釈のために、3.2.5から先に読み進めることも可能である。

### 3.2.2 試験デザイン案

本項では、Complex Innovative Trial Design (CID) のウェブサイトにて CID pilot meeting program に採択された試験の1つとして公開されている、びまん性大細胞型 B 細胞リンパ腫を対象とした臨床試験の事例を取り上げる。詳細な試験デザインや commensurate prior の数理的背景は、製薬協成果物「Complex Innovative Trial Design の現状と事例の整理 (2023)」<sup>23</sup>も参照のこと。

本試験では、バイオマーカー陽性患者 (N=420) を2:1の割付比で新規併用療法 (N=280)、対照群として R-CHOP (N=140) に組み入れ、さらに外部対照データとして100例分のデータを用意する。本試験では主要な副次的評価項目として全生存期間 (OS) を設定し、OS の解析集団は外部対照群からの患者を加えることで補強される。OS の解析ではワイブル分布を仮定したベイズ流の CMP 法を採用し、外部対照群からの情報を動的に借用する。さらに、傾向スコアマッチングを行い、解析に含める外部対照患者を選定する。治療効果の推定は、ハザード比の事後分布の事後平均と95%信用区間に基づいて行われ、ハザード比が1を下回る事後確率が97.5%を超える場合に治療効果があると判断する。

表 3.1 試験デザインと解析方法

試験標題	びまん性大細胞型 B 細胞リンパ腫の第 III 相試験
目的	びまん性大細胞型 B 細胞リンパ腫のバイオマーカー陽性患者を対象とする1次療法の第3相試験
投与群	新規併用療法群、対照群 (R-CHOP)
割付比	2:1
主要評価項目	医師判定の無増悪生存期間
重要な副次評価項目	OS
解析方法	<p>重要な副次評価項目である OS に対する解析では、ワイブル分布を仮定した commensurate prior を用いたベイズ流手法を利用する。commensurate prior の数理的側面を以下に示す。</p> <p>被験者 <math>i</math> の time to event アウトカムの確率変数を <math>T_i</math> とし、特に外部対照データ (External Control: EC) の場合に <math>T_{EC,i}</math> と表し、新規試験 (Internal Control: IC) の場合に <math>T_{IC,i}</math> と表す。加えて、被験者 <math>i</math> の <math>T_i</math> の打ち切り情報を <math>\delta_i</math> で示し、<math>\delta_i = 1</math> をイベントあり、<math>\delta_i = 0</math> を打ち切りする。確率変数 <math>T_i</math> の従う確率分布を <math>P(T_i, \delta_i, \mathbf{X}_i, \boldsymbol{\theta})</math> と定義する。<math>\mathbf{X}_i</math> は外部対照データか否かを表すダミー変数、試験治療の有無を表すダミー変数や患者背景情報を含む説明変数 (共変量) とし、<math>\boldsymbol{\theta}</math> は確率分布のパラメータを表す。具体的には、<math>\boldsymbol{\theta} = (\beta_0, \beta_{ext}, \beta_{trt}, \boldsymbol{\beta}_{covariate}, \gamma)</math> で構成され、<math>\beta_0</math> はワイブル回帰に対応する切片項であり、<math>\beta_{ext}</math> は外部対象データの有無を表すダミー変数に対する回帰パラメータ、<math>\beta_{trt}</math> は治療の有無を表すダミー変数に対する回帰パラメータ、<math>\boldsymbol{\beta}_{covariate}</math> は患者背景に対する <math>J</math> 次元の回帰パラメータベクトル、</p>

$\gamma$ はワイブル分布の形状パラメータを表す。  
 各被験者のデータ  $(T_i, \delta_i, \mathbf{X}_i) = (T_i, \delta_i, X_{ext,i}, X_{trt,i}, \mathbf{X}_{covariate,i})$ が得られた  
 場合の $\theta$ の事後分布 $P(\theta|T_{EC}, \delta, \mathbf{X})$ 、 $P(\theta|T_{IC}, \delta, \mathbf{X})$ は

- 外部対照データにおける事後分布  

$$P(\theta|T_{EC}, \delta, \mathbf{X})$$

$$= P(\beta_{ext}, \beta_{covariate}, r|T_{EC}, \delta, \mathbf{X})$$

$$\propto L(\beta_{ext}, \beta_{covariate}, r|T_{EC}, \delta, \mathbf{X})P(\beta_{ext})P(\beta_{covariate})P(r)$$

- 新規試験における事後分布  

$$P(\theta|T_{IC}, \delta, \mathbf{X}) =$$

$$P(\beta_0, \beta_{trt}, \beta_{covariate}, r|T_{IC}, \delta, \mathbf{X})$$

$$\propto L(\beta_0, \beta_{trt}, \beta_{covariate}, r|T_{IC}, \delta, \mathbf{X})P(\beta_0)P(\beta_{trt})P(\beta_{covariate})P(r)$$

で表すことができる。対して、commensurate prior  $P(\beta_0|\beta_{ext}, \tau)$ を用いて、  
 外部対照データを利用しつつ、新規試験と外部対照データ間のアウトカム $T_i$ の違いを考慮すると、 $\theta$ の事後分布 $P(\theta|T_{EC}, T_{IC}, \delta, \mathbf{X})$ は

- $$P(\theta|T_{EC}, T_{IC}, \delta, \mathbf{X})$$

$$= P(\beta_0, \beta_{trt}, \beta_{ext}, \beta_{covariate}, r, \tau|T_{EC}, T_{IC}, \delta, \mathbf{X})$$

$$\propto L(\beta_{ext}, r|T_{EC}, \delta, \mathbf{X})L(\beta_0, \beta_{trt}, r|T_{IC}, \delta, \mathbf{X})L(\beta_{covariate}|T_{EC}, T_{IC}, \delta, \mathbf{X})$$

$$P(\beta_0|\beta_{ext}, \tau)P(\beta_{ext})P(\beta_{trt})P(\beta_{covariate})P(r)P(\tau)$$

と表すことができる。

$P(\beta_{ext})$ 、 $P(\beta_{trt})$ 、 $P(\beta_{covariate})$ 、 $P(r)$ 、 $P(\tau)$ は各パラメータに対する事前分布であり、特に $P(\beta_{ext})$ 及び $P(\tau)$ はパラメータ $\beta_0$ に対する事前分布であることから、超事前分布と表現する。ただし、 $\beta_0$ に対する事前分布が新規試験と外部対照データ間のアウトカム $T_i$ の違いを考慮していることから commensurate prior と呼ばれ、新規試験と外部対照データ間のアウトカムの違いは $\tau$ の事前分布 $P(\tau)$ を用いてその程度を調整する。上記の尤度、事前分布、超事前分布及び commensurate prior の関係図を図 3-2に示す。

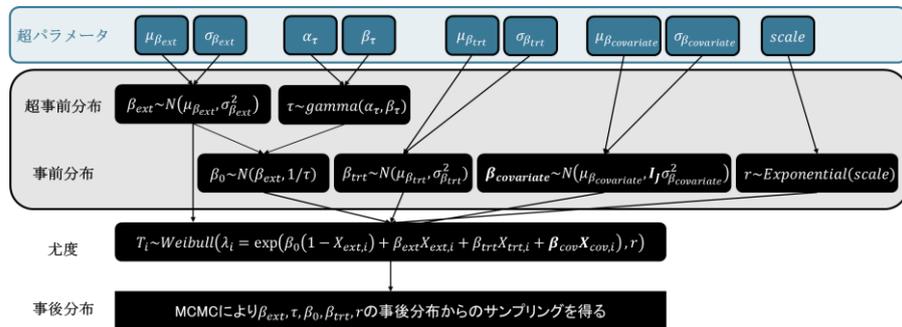


図 3-2 commensurate prior の概念図

### 3.2.3 シミュレーション実験

#### 3.2.3.1 シミュレーションの目的

本シミュレーション実験では、CMP 法を用いて外部対照データを利用した際の第一種の過誤確率、検出力、バイアス、最小二乗平均 (Mean Squared Error : MSE)、effective historical sample size (EHSS) への影響を検討するために、複数の効果の想定のもとでその動作特性を確認する。

参考とする試験では傾向スコアマッチングにて用いる外部対照データを特定するが、本シミュレーション実験では用いる外部対照データは予め決まっているものとする。

#### 3.2.3.2 シミュレーションの設定

シミュレーション実験で検討する試験デザインと生成するデータの条件を3.2.2で紹介した試験デザインに沿って示す。

新規試験に組み入れる被験者数は300名とし、新規併用療法群と対照群への割り付け比率を2 : 1とする (新規併用療法群 : 200名、対照群 : 100名)。外部対照データについて、傾向スコア等に基づいて利用できる被験者数は100名と想定する。新規試験と外部対照データのそれぞれの被験者を登録する期間は24ヶ月と想定し、一定の割合・間隔で登録されると想定する。また、最後の被験者の登録から36ヶ月間を追跡期間とし、当該期間を終えた時点で解析することとする。

外部対照データの利用に際して、新規試験との患者背景の違いが課題になることから、その課題を検討できるように、後述するシミュレーション実行する R パッケージである `psborrow` にて提供されている例を参考に、新規試験と外部対照データの患者背景分布の違いが生じるように設定する。はじめに、評価項目 OS に対する共変量は5つあると想定し、そのうち3つは2値のカテゴリカル変数で、2つは連続変数とする。新規試験と外部対照データのそれぞれの共変量間の相関関係を以下のように考慮する。

- ・ 新規試験の患者背景 (共変量) の分布

共変量が多次元正規分布に従うと仮定し、平均及び共分散構造を以下の通り設定する。本事例では、シミュレーションを実施する R package の `psborrow` を参考に設定した。

$$\begin{bmatrix} \text{covariate1} \\ \text{covariate2} \\ \text{covariate3} \\ \text{covariate4} \\ \text{covariate5} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0.5 \\ 0.5 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0.7 & 0 & 0 \\ 0.5 & 1.2 & 0.9 & 0 & 0 \\ 0.7 & 0.9 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.7 & 0.7 \\ 0 & 0 & 0 & 0.7 & 0.7 \end{bmatrix} \right)$$

ただし、`category1`、`category2`、`category4`を2値のカテゴリカル変数とし、`covariate1`、`covariate 2`、`covariate 4`から以下の通り変換する。

- $\Pr(\text{category1}_i = 1) = \Pr(0.45 \leq \text{covariate1}_i)$
    - $\Pr(\text{category2}_i = 1) = \Pr(0.55 \leq \text{covariate2}_i)$
    - $\Pr(\text{category4}_i = 1) = \Pr(0.50 \leq \text{covariate4}_i)$
  - ・ 外部対照データの患者背景 (共変量) の分布
- 新規試験と同様に、共変量が多次元正規分布に従うと仮定し、平均及び共分散構造

を以下の通り設定する。

$$\begin{bmatrix} \text{covariate1} \\ \text{covariate2} \\ \text{covariate3} \\ \text{covariate4} \\ \text{covariate5} \end{bmatrix} \sim N \left( \begin{bmatrix} 0.7 \\ 0.5 \\ 0.9 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0.7 & 0 & 0 \\ 0.5 & 1.2 & 0.9 & 0 & 0 \\ 0.7 & 0.9 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.7 & 0.7 \\ 0 & 0 & 0 & 0.7 & 0.7 \end{bmatrix} \right)$$

新規試験と同様に、2値のカテゴリカル変数 category1、category2、category4を covariate1、covariate 2、covariate 4を用いて以下の通り変換する。

- $\Pr(\text{category1}_i = 1) = \Pr(0.65 \leq \text{covariate1}_i)$
- $\Pr(\text{category2}_i = 1) = \Pr(0.55 \leq \text{covariate2}_i)$
- $\Pr(\text{category4}_i = 1) = \Pr(0.50 \leq \text{covariate4}_i)$

新規試験と外部対照データにおける患者背景分布について、category1、category2及び covariate3についてはその分布が異なり、category4及び covariate5は同じ分布であると仮定する。

次に、新規試験と外部対照データにおける評価項目 OS のデータ発生条件を検討する。OS はワイブル分布に従うと仮定し、尺度パラメータ $\lambda_i$ と形状パラメータ0.9のワイブル分布に基づいてデータを発生させる。ワイブル分布の尺度パラメータ $\lambda_i$ の設定について、新規試験の治療群間差（ハザード比：HR）と新規試験の対照群と外部対照データの差（drift HR）を用いて以下のように表現する。 $\lambda_c$ は新規試験の対照群における OS の中央値等から算出される。

- Internal control:  $\lambda_i = \lambda_c \exp(\mathbf{X}_i\boldsymbol{\beta}) = \lambda_c \times \exp(0.5\text{category1}_i + 0.5\text{category2}_i + 0.5\text{covariate3}_i + 0.5\text{category4}_i + 0.5\text{covariate5}_i)$
- Internal treatment:  $\lambda_i = \lambda_c \times HR \times \exp(0.5\text{category1}_i + 0.5\text{category2}_i + 0.5\text{covariate3}_i + 0.5\text{category4}_i + 0.5\text{covariate5}_i)$
- External control:  $\lambda_i = \lambda_c \times \text{driftHR} \times \exp(0.5\text{category1}_i + 0.5\text{category2}_i + 0.5\text{covariate3}_i + 0.5\text{category4}_i + 0.5\text{covariate5}_i)$

最後に、転院等により OS が追跡できない状況を想定し、1年あたり0.05人/年でランダムな脱落が生じると仮定する。ただし、利用するパッケージの仕様により上記の設定を適切に反映できていないことを後の章で詳細に明記する。上記のランダムな脱落時点又は追跡期間終了にイベントが起こっていない場合、上記のランダムな脱落時点と追跡期間終了時のいずれか早い時点で打ち切りとする。

検討するシナリオとして、HR 及び drift HR を変更した場合の動作特性を評価する。以下に OS の中央値として新規試験の対照群の治療効果の想定と検討する HR 及び drift HR を示す。

- 新規試験の対照群（R-CHOP）の真の効果（ $\lambda_c$ ）：Log(2)/24（OS の中央値 24ヶ月を参考にした値）
- 新規試験との真の治療群間差（HR）：1, 0.65
- 新規試験の対照群と外部対照データとの真の差（drift HR）：1, 1.2

本シミュレーション実験では、外部対照データを借用しない頻度流に基づく **no borrowing** 法、外部対照データの全てを借用する頻度流に基づく **full borrowing** 法を参考に、CMP 法の動作特性を評価する。**no borrowing** 法と **full borrowing** 法では、ワイブル回帰を用いた最尤推定に基づく推定を行うことから、その解析モデルの詳細を示す。さらに、先述した CMP 法における事前分布の詳細を以下に示す。確率密度関数が異なる背景は、後述する解析プログラム（パッケージ）に応じて推定されるパラメータが異なるため、その点を明確にするために分けて記載した。

- **no borrowing** 法と **full borrowing** 法の解析モデル（ワイブル回帰）
  - 解析プログラム：R survival パッケージ
  - 確率密度関数： $f(t_i) = \frac{1}{\sigma t} \exp\left(\frac{\log(t_i) - \mu_i}{\sigma}\right) \exp\left(-\exp\left(\frac{\log(t_i) - \mu_i}{\sigma}\right)\right)$
  - パラメータ  $\mu_i$  に対する解析モデル： $\mu_i = \beta_0 + \beta_{trt} X_{trt,i} + \beta_{covariate} X_{covariate,i}$
  - 対照群に対する新規併用療法群の HR： $\exp(-\beta_{trt} / \sigma)$
- CMP 法の解析モデル
  - 解析プログラム：R psborrow2 パッケージ
  - 確率密度関数： $f(t_i) = \gamma \lambda_i t_i^{\gamma-1} \exp(-\lambda_i t_i^\gamma)$
  - パラメータ  $\lambda_i$  に対する解析モデル： $\lambda_i = \exp(\beta_0(1 - X_{ext,i}) + \beta_{ext} X_{ext,i} + \beta_{trt} X_{trt,i} + \beta_{covariate} X_{covariate,i})$
  - 対照群に対する新規併用療法群の HR： $\exp(-\beta_{trt})$
  - CMP 法の前分布
    - ◇  $P(\beta_0) \sim Normal(\text{平均} = \beta_{ext}, \text{標準偏差} = \sqrt{1/\tau})$
    - ◇  $P(\beta_{ext}) \sim Normal(\text{平均} = 0, \text{標準偏差} = \sqrt{1000})$
    - ◇  $P(\beta_{trt}) \sim Normal(\text{平均} = 0, \text{標準偏差} = \sqrt{1000})$
    - ◇  $P(\beta_{covariate}) \sim Normal(\text{平均} = 0, \text{標準偏差} = \sqrt{1000})$
    - ◇  $P(r) \sim exponential(\lambda = 1)$
    - ◇  $P(\tau) \sim gamma(\alpha = 1, \beta = 0.001)$

上記は共変量を含めたモデルに基づく評価をすることを目的とするが、共変量を利用しないモデルに基づく評価も合わせて実施する。

CMP 法を用いて HR を推定するために MCMC 法を用いる。MCMC 法にて設定する事後分布推定のための条件は以下の通りとする。

- Burn in 期間：1000
- iteration 数：10000
- chain 数：2
- thinning 数：1

CMP 法の動作特性の評価に際して、上記で設定したシナリオに応じて、第一種の過誤確率、検出力、バイアス、MSE 及び EHSS を以下の通り評価する。ただし、それぞれの評価

に際してシミュレーション実験回数は1000回とする。

- 第一種の過誤確率、検出力
  - no borrowing 法、full borrowing 法：推定された HR（最尤推定値）の95%信頼区間の上限が1を下回った回数に基づき評価
  - CMP 法：ハザード比の事後分布の95%信用区間の上限が1を下回った回数に基づき評価
- バイアス、MSE
  - no borrowing 法、full borrowing 法：推定された HR（最尤推定値）と新規試験の真の治療群間差（HR）との差に基づき評価
  - CMP 法：HR の事後平均値と新規試験の真の治療群間差（HR）との差に基づき評価
- EHSS
  - no borrowing 法：計算しない
  - CMP 法、full borrowing 法：治療の有無を表す回帰パラメータ $\beta_{trt}$ の精度（1/分散）を用いて算出

EHSS は、事後 effective sample size から新規試験の被験者数を差し引く方法であり、以下の式により得られる。本報告書における EHSS の算出では、 $\text{Prec}(\beta_{trt}|\text{Concurrent trial data})$ の計算ではベイズ流手法に基づき計算された値を用いる。詳細は、3.2.5.4に明記する。

$$EHSS = N_{ic} \left( \frac{\text{Prec}(\beta_{trt}|\text{all data})}{\text{Prec}(\beta_{trt}|\text{Concurrent trial data})} - 1 \right)$$

### 3.2.3.3 シミュレーションの結果

3.2.3.2に記載した設定のもとで、それぞれの指標を評価した結果を以下に示す。

- 第一種の過誤確率

解析モデル中の共変量	HR	drift HR	no borrowing 法	CMP 法	full borrowing 法
あり	1	1	0.027	0.025	0.029
あり	1	1.2	0.027	0.086	0.102
なし	1	1	0.039	0.053	0.060
なし	1	1.2	0.039	0.105	0.134

- 検出力

解析モデル中の共変量	HR	drift HR	no borrowing 法	CMP 法	full borrowing 法
あり	0.65	1	0.817	0.929	0.929
あり	0.65	1.2	0.817	0.971	0.984
なし	0.65	1	0.482	0.690	0.719
なし	0.65	1.2	0.482	0.79	0.855

- バイアス

解析モデル中の共変量	HR	drift HR	no borrowing 法	CMP 法	full borrowing 法
あり	0.65	1	0.817	0.929	0.929
あり	0.65	1.2	0.817	0.971	0.984
なし	0.65	1	0.482	0.690	0.719
なし	0.65	1.2	0.482	0.79	0.855

あり	1	1	0.013	0.016	0.008
あり	1	1.2	0.013	-0.050	-0.068
あり	0.65	1	0.003	0.005	0.001
あり	0.65	1.2	0.003	-0.037	-0.048
なし	1	1	0.008	-0.012	-0.023
なし	1	1.2	0.008	-0.057	-0.076
なし	0.65	1	0.112	0.096	0.088
なし	0.65	1.2	0.112	0.062	0.048

• MSE

解析モデル中の 共変量	HR	drift HR	no borrowing 法	CMP 法	full borrowing 法
あり	1	1	0.024	0.034	0.017
あり	1	1.2	0.024	0.033	0.019
あり	0.65	1	0.010	0.014	0.007
あり	0.65	1.2	0.010	0.014	0.008
なし	1	1	0.025	0.031	0.016
なし	1	1.2	0.025	0.033	0.020
なし	0.65	1	0.027	0.027	0.017
なし	0.65	1.2	0.027	0.021	0.010

• EHSS

解析モデル中の 共変量	HR	drift HR	no borrowing 法	CMP 法	full borrowing 法
あり	1	1	-	124.2	134.8
あり	1	1.2	-	117.2	139.8
あり	0.65	1	-	119.4	129.2
あり	0.65	1.2	-	111.8	130.5
なし	1	1	-	114.6	153.5
なし	1	1.2	-	110.0	154.8
なし	0.65	1	-	110.7	146.4
なし	0.65	1.2	-	104.9	147.9

### 3.2.3.4 シミュレーションの結果の考察

第一種の過誤確率について、drift HR = 1でありかつ適切に共変量にて調整した場合を除いて、いずれの手法でも名義水準である2.5%を超える結果であり、no borrowing 法に対する CMP 法や full borrowing 法の増大傾向は顕著であった。特に、Drift HR が1.2の場合には、共変量による調整によらず一貫して名義水準を大きく上回る傾向であった。本傾向は、外部対照データを利用することで、新規併用療法群と外部対照データを組み合わせた対照群 (R-CHOP) の OS に違いが生じたため、治療群間差 HR が1の状況にて第一種の過誤確率が名義水準を上回ったと考える。

次に CMP 法や full borrowing 法の検出力について、no borrowing 法と比較して検出力が大きくなった。CMP 法と full borrowing 法の間では大きな検出力の差はないが、外部対照データの利用に伴い情報量が増加して、no borrowing 法と比較して検出力が増加したと考える。特に、drift HR = 1.2の場合、外部対照データを利用することで治療間差が大きくなるシナリオのため、drift HR = 1の場合と比較して検出力の増加の程度は大きかった。

バイアスについて、共変量による調整の有無によりその傾向の向きが変わるが、CMP 法

のバイアスはおおむね no borrowing 法と full borrowing 法の間位置することが確認できた。特に drift HR = 1 の場合でかつ共変量による調整を行った場合、full borrowing 法のバイアスが最も0に近い結果となるが、drift HR = 1.2 の場合では最も0から離れる傾向が確認できた。外部対照データとの drift が存在する場合にはバイアスの程度が上昇することは明らかであり、その傾向が結果にも示されたと考える。

MSE について、drift HR に関わらず、no borrowing 法と full borrowing 法と比較して CMP 法の MSE が最も大きい傾向であった。MSE についても、バイアスと同じく no borrowing 法と full borrowing 法の間位置することを期待したが、結果として最も MSE が大きかった。本傾向はベイズ流と頻度流に基づく違いと想像され、特に階層的な事前分布を用いることで結果の不安定性が上昇したものと考える。本件については、シミュレーション回数や MCMC 法にて設定する事後分布推定のための条件を変更することで改善が見込めると想像する。no borrowing 法と full borrowing 法を比較すると、drift HR に関わらず full borrowing 法の MSE 小さくなり、外部対照データを用いることで推定精度の上昇は確認された。

EHSS について、CMP 法おおよそ110付近の結果であり、full borrowing 法と比較して小さい値であった。このことから、CMP 法では外部データの全ての情報量を利用していないことが分かる。加えて、共変量の設定に応じて、CMP 法と full borrowing 法の EHSS の差の程度は異なり、共変量を適切に調整することでデータ間の患者背景の違いを含めた drift を考慮した情報借用ができていていることが伺える。

### 3.2.4 R package の紹介 (psborrow, psborrow2)

#### 3.2.4.1 psborrow と psborrow2 の整理

time to event データを対象に CMP 法を使用できる R package として psborrow<sup>26</sup>と psborrow<sup>28</sup>がある。psborrow という名前は、傾向スコアとベイズ動的借用法を組み合わせたものであり、いずれのパッケージも動的借用と傾向スコア調整/重み付け法を組み合わせるために使用することが可能である。psborrow パッケージが先に開発され、その改訂版として psborrow2 が存在する。psborrow2 は現在も開発中のパッケージであり、状況に応じて psborrow も利用可能と考えたため、それぞれのパッケージの特徴を以下に示す。特に、利用可能なサンプリングプログラムが異なるため、利用者の開発環境に応じて適切なパッケージを選択いただきたい。更新状況を考慮すると、基本的には psborrow2 を利用してシミュレーション実験することで問題無いが、Stan が使用できない場合には psborrow を利用することを提案する。

表 3.2 パッケージ間の比較

	psborrow	psborrow2
バージョンと作成年月日	v0.2.1 (2023/03/03)	v0.0.3.4 (2024/5/1)
パッケージのサポート	なし	あり
サンプリングプログラム	JAGS	Stan
利用可能な Outcome	time to event	<ul style="list-style-type: none"> <li>• continuous data</li> <li>• categorical data</li> <li>• time to event data</li> </ul>

傾向スコアを用いた共変量の利用	可	可
傾向スコアモデルに基づくマッチング	不可	可

### 3.2.4.2 psborrow package の使用例

CMP 法を実装するための R package として psborrow が提供されており、DLBCL の臨床試験においてもこの package が使用されているという事から、まずは CRAN にて提供されている psborrow の userguide 等を基に整理した。

```
library(psborrow)
library(dplyr)
library(survival)
library(survminer)
```

なお使用するには別途 JAGS のインストールも必要である。psborrow では以下のフロー<sup>24</sup>に従ってシミュレーションを実装する。

#### Flow 1: Set treatment arm indicators and covariates

治療群150人、対照群150人の同時試験、及び外部対照群100人を設定する。

```
ss <- set_n(ssC=200, ssE=100, ssExt=100)
#ssC=Number of observations in the internal control arm.
#ssE=Number of observations in the internal experiment arm.
#ssExt=Number of observations in the external control arm.
```

共変量の設定を行う。下の code ではカテゴリカル変数2つ、連続変数1つを設定し、それぞれの相関を下記の様に見積もる。

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0.5 \\ 0.5 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0.7 \\ 0.5 & 1.2 & 0.9 \\ 0.7 & 0.9 & 1 \end{bmatrix} \right)$$

```
covset1 <- set_cov(
  n_cat = 2,
  n_cont = 1,
  mu_int = c(0, 0.5, 0.5),
  mu_ext = c(0.7, 0.5, 0.9),
  var = c(1, 1.2, 1),
  cov = c(0.5, 0.7, 0.9),
  prob_int = c(0.45, 0.55), #Probability of binary covariate equalling 1 in the internal trial.
  prob_ext = c(0.65, 0.55) #Probability of binary covariate equalling 1 in the external trial
)
```

ここで prob\_int 及び prob\_ext は多変量正規分布から発生した値に応じて、カテゴリカル変数（バイナリ変数）に変換される。例えば prob\_int = c(0.45, 0.55) は  $Pr(x_1 = 1) = 0.45$ 、 $Pr(x_2 = 1) = 0.45$  であり、 $x_1 \geq 0.45$  のとき  $x_1 = 1$  と変換される。必要に応じて共変量を更に追加する事ができる。ここではカテゴリカル変数2つ ( $x_4, x_5$ )、

連続変数3つ ( $x_6, x_7, x_8$ ) を追加した。なお、今回の設定では同時試験と外部試験で共変量において系統的な偏りが生じている事を許容している。

$$\begin{bmatrix} x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.7 & 0.7 & 0.7 & 0.7 & 0.7 \\ 0.7 & 0.7 & 0.7 & 0.7 & 0.7 \\ 0.7 & 0.7 & 0.7 & 0.7 & 0.7 \\ 0.7 & 0.7 & 0.7 & 0.7 & 0.7 \\ 0.7 & 0.7 & 0.7 & 0.7 & 0.7 \end{bmatrix} \right)$$

```
covset2 <- set_cov(
  n_cat = 2,
  n_cont = 3,
  mu_int = 0,
  mu_ext = 0.2,
  var = 0.7,
  cov = 0.7,
  prob_int = 0.5
)
```

上記までで設定した共変量に従って、共変量データセットを作成する。複数の HR を同時に評価できるので、同時治療群と同時対照群のハザード比 (HR=0.65 or 1)、及び外部対照群と同時対照群のハザード比 (driftHR = 1 or 1.2) を指定する。2章で示した様に、FDA ガイドラインに基づく試験デザインの動作特性評価にはシミュレーション回数は 100000 回が推奨されているが、本 package を用いた simulation は1回の繰り返しあたりの実行に多くの時間を要するため、現実的に100000回を行うのは困難なため、今回はシミュレーションを1000回とした。

```
cov_list <- c(covset1, covset2)

sample_cov <- simu_cov(
  ssObj = ss,
  covObj = cov_list,
  HR = c(0.65, 1), #a list of hazard ratio of treatment and control arms
  driftHR = c(1, 1.2), #hazard ratio of external control and internal control arms
  nsim = 1000, #number of simulation
  seed = 47
)
```

下記のようなデータセット (matrix) を得る。

```
head(sample_cov[[1]], 5)

##   driftHR HR ext trt cov1 cov2   cov3 cov4 cov5   cov6
## [1,]  1.0  1  0  0  0  1.426442353  1  1  0.06551554
## [2,]  1.0  1  0  1  1  0.413408109  1  1 -0.86978706
## [3,]  1.0  1  0  1  1 -0.667792987  1  1 -0.52705948
## [4,]  1.0  1  0  0  1 -0.008703362  1  1 -0.66796308
## [5,]  1.0  1  0  1  1 -0.941826735  0  0  0.90304808
##           cov7   cov8
## [1,] 0.06551554 0.06551554
```

```
## [2,] -0.86978706 -0.86978706
## [3,] -0.52705948 -0.52705948
## [4,] -0.66796308 -0.66796308
## [5,] 0.90304808 0.90304808
```

## Flow 2: Set trial operational parameters and Simulate survival times

次のフローでは生存時間・打ち切りの情報と治験の登録・脱落に関する情報を設定する。両試験とも、死亡までの期間はワイブル分布に従うと仮定した。ここでは形状パラメータを0.9と設定する。

$$t_i \sim Weibull(0.9, \lambda_i)$$

尺度パラメータは下記のように表現でき、 $I_{trt}, I_{ext}$ はそれぞれ同時治療群、外部対照群の時1となる指示変数である。なお、この時 $\beta_{trt} = \log(HR)$ 及び $\beta_{ext} = \log(driftHR)$ が成り立つ。

$$\lambda_i = \lambda_c * \exp(\beta_0 + \beta_{trt} * I_{trt} + \beta_{ext} * I_{ext} + X_i\beta)$$

$$\beta = (\beta_{x_1}, \beta_{x_2}, \dots, \beta_{x_g})$$

このとき、 $\beta_0$ は同時対照群のリスクを表し、共変量の影響を受けない同時対照群の尺度パラメータは $\lambda_i = \lambda_c * \exp(\beta_0)$ と表される。同時対照群の患者の全生存期間の中央値は51ヵ月と設定され、イベント発現率は1ヵ月あたり0.0135人の患者の割合に変換される。さらに、各患者の生存時間は、ベースラインの共変量値とその係数を指定し、尺度パラメータに対する共変量の影響度を設定する。

```
evt <- set_event(
  event = "weibull",
  shape = 0.9,
  lambdaC = 0.0135, #同時対照群のベースラインハザード関数
  beta = 0.5 #尺度パラメータλに対する共変量の影響
)
```

治験の登録・脱落情報について、下記を仮定する。

- 新規試験では 300 例が登録期間 24 ヶ月内に均一に登録される (12.5 例/月)
- 外部対象データでは 100 例が登録期間 24 ヶ月内に均一に登録される (4.17 例/月)
- また解析は最後の患者の登録から 36 ヶ月後に開始されるように設定 (CCOD="fixed-last") する。

```
c_int <- set_clin(
  gamma = 12.5, #A vector of rate of enrollment per unit of time
  CCOD = "fixed-last", #Type of analysis start time
  CCOD_t = 45, #Type of analysis start time
  etaC = 0.05, #A vector for dropout rate per unit time for control arm
  etaE = 0.05, #A vector for dropout rate per unit time for experimental arm
```

```

d_itv = 36#A vector of duration of time periods for dropping out the study with rates specified
in etaC and etaE
)
c_ext = set_clin(
  gamma = 4.17,
  CCOD = "fixed-last",
  CCOD_t = 36,
  etaC = 0.05
)

```

以上の情報をもとに、`simu_time()`関数を用いて、登録パターン、イベント発生までの時間、ドロップアウトまでの時間をシミュレートする。シード値もここで指定できる。なお、この設定では45か月後に登録された被験者は解析から除外する事を設定しているため、実行後に該当する症例が除外されたことを知らせる **Warning** が発出されるが、以降の解析に支障はない。

```

sample_time <- simu_time(
  dt = sample_cov,
  eventObj = evt,
  clinInt = c_int,
  clinExt = c_ext,
  seed = 47
)

head(sample_time[[1]], 5)

##   driftHR  HR ext trt cov1 cov2   cov3 cov4 cov5   cov6   cov7
## [1,]    1.0 0  1  0  0 1.69867515  0  0 1.4069318 1.4069318
## [2,]    1.0 0  1  1  1 -0.42546841  1  1 -0.3578853 -0.3578853
## [3,]    1.0 0  1  1  1 -0.04927021  0  0 0.1370222 0.1370222
## [4,]    1.0 0  1  1  1 -0.42705058  0  0 0.3709728 0.3709728
## [5,]    1.0 0  1  0  1 -0.10903808  0  0 0.2500303 0.2500303
##          cov8   time cnsr
## [1,] 1.4069318 5.952542  0
## [2,] -0.3578853 12.561759  1
## [3,] 0.1370222  8.891383  1
## [4,] 0.3709728 26.564068  1
## [5,] 0.2500303 16.332164  1

```

### Flow 3: Choose borrowing method and run Bayesian analysis

シミュレーションデータの作成が完了したら、ベイズモデルを使ってさまざまな借用法の分析に進む。今回は、**time-to-event** が Weibull 分布でモデルされることを仮定しており、ユーザーは尺度パラメータの予測に共変量を使用するかどうかを決めることができる。この例では、`psborrow` で利用可能な借用法を用いて以下の6つのシナリオにおいて、事後分布を生成する。

- ① 共変量による調整あり、外部データを全て借用する。

- ② 共変量による調整あり、外部データを借用しない。
- ③ 共変量による調整あり、commensurate prior による動的借用。
- ④ 共変量による調整なし、外部データを全て借用する。
- ⑤ 共変量による調整なし、外部データを借用しない。
- ⑥ 共変量による調整なし、commensurate prior による動的借用。

commensurate prior アプローチを用いた動的借用法では、同時対照と外部対照の間の精度を表す超パラメータ $\tau$ は、ガンマ分布と半コーシー分布に従う事を指定できる。この例ではガンマ分布に従う場合を仮定している。他の2つの設定では、外部試験からすべての情報を借用する場合と、外部情報を一切使用しない場合の影響を評価するために、従来のベイズモデルを使用する。

- ・ ①や④のように外部データを全て借用する場合

$$\beta_0 = \beta_{\text{ext}} = \beta \sim N(0, 0.0001)$$

- ・ ②や⑤のように外部データを借用しない場合

$$\beta_0 = \beta \sim N(0, 0.0001)$$

それぞれのパラメータは上記の無情報事前分布が設定される。なお上記の  $N(0, 0.0001)$  の0.0001は分散の逆数である精度を表す。超パラメータ $\tau$ の prior の設定は下記のように set\_prior 関数の prior 引数で指定する。

```
##### 共変量による調整あり#####
##Full borrowing##
pr1 <- set_prior(
  pred = "all", #Weibull 分布に含める説明変数
  prior = "full_ext",
  r0 = 1, #Weibull の形状パラメータの初期値
  alpha = 0 #ベースラインハザードの対数値の初期値
)

##Non borrowing##
pr2 <- set_prior(
  pred = "all",
  prior = "no_ext",
  r0 = 1,
  alpha = 0
)

##Commesurate prior##
pr3 <- set_prior(
  pred = "all",
  prior = "gamma", #借用の度合いを制御する精度パラメータ  $\tau$  の prior
  r0 = 1,
  alpha = c(0, 0) #外部対照と同時対照のベースラインハザードの対数値の初期値
)

##### 共変量による調整なし#####
##Full borrowing##
pr4 <- set_prior(
```

```

pred = "none",
prior = "full_ext",
r0 = 1,
alpha = 0
)

##Non borrowing##
pr5 <- set_prior(
  pred = "none",
  prior = "no_ext",
  r0 = 1,
  alpha = 0
)

##Commesurate prior##
pr6 <- set_prior(
  pred = "none",
  prior = "gamma",
  r0 = 1,
  alpha = c(0, 0)
)

```

最後に6つのシナリオを組み合わせ、MCMCを実行して事後分布からサンプルを生成する。モデルの並列チェーン数、適応のための反復数、バーンインとして破棄される反復数を指定でき、例えば下記の設定に基づくとモニターするサンプリングはそれぞれ2、100、100、200である。ユーザーはシード値を指定することもできる。JAGSでは効率的なMCMCサンプリングに係る数値設定を適用させるフェーズが設けられており、この工程で効率的なステップ幅を設定する事ができる。この探索に必要な反復回数を `n.adapt` で設定、その後初期値の影響を除くために除外する反復数を `n.burn` で設定、サンプリングデータとして取得する反復数を `n.iter` で設定する。

なお、`run_mcmc_p()`を呼び出すことで、並列処理を利用することもできる。

```

pr_list <- c(pr1, pr2, pr3, pr4, pr5, pr6)

res <- run_mcmc(
  dt = sample_time,
  pr_list,
  n.chain = 2,
  n.adapt = 100,
  n.burn = 100,
  n.iter = 200,
  seed = 47,
)

```

#### Flow 4: Summarize results to assess impact from borrowing

`get_summary()`関数により MCMC 結果と要約統計量を表示できる。

```
summ <- get_summary(res)
```

シミュレーション結果から試験デザインの動作特性（第一種の過誤確率、検出力、バイ

アス、MSE)を確認する。

```

plot_type1error(
  summ,
  driftHR = 1.2,
  pred = "all"
)

plot_power(
  summ,
  HR = 0.65,
  driftHR = 1,
  pred = "all"
)

plot_hr(
  summ,
  HR = 0.65,
  driftHR = 1.2,
  pred = "all"
)

plot_bias(
  summ,
  HR = 1,
  driftHR = 1.2,
  pred = "all"
)

plot_mse(
  summ,
  HR = 0.65,
  driftHR = 1,
  pred = "all"
)

```

### 3.2.4.3 psborrow2 package の使用例

psborrow の改訂版として psborrow2が提供されており、psborrow に沿った利用方法をまとめた。psborrow2を用いたシミュレーションの詳細は CRAN の Articles 7.Data Simulation を参照のこと。

psborrow2の実施に際して、psborrow2パッケージのインストールに加えて、Rstan のインストールが必要である。また、psborrow2パッケージでは、cmdstanr ライブラリを用いるため、事前の読み込みが必要である。

```

library(psborrow2)
library(cmdstanr)
library(dplyr)

```

#### Flow 1: Set treatment arm indicators and covariates

治療群、対照群、外部対照群の N 数と共変量の設定を同時に行う（治療群200人、対照

群100人の同時試験、及び外部対照群100人)。ここで covariate1、covariate2、covariate4はカテゴリカル変数（バイナリ変数）に変換する必要があるため、transformations の部分にて変換を行う。

```
baseline <- create_baseline_object(
  n_trt_int = 200,
  n_ctrl_int = 100,
  n_ctrl_ext = 100,
  covariates = baseline_covariates(
    name = c("covariate1", "covariate2", "covariate3", "covariate4", "covariate5"),
    means_int = c( 0, 0.5, 0.5, 0, 0),
    means_ext = c(0.7, 0.5, 0.9, 0, 0),
    covariance_int = covariance_matrix(
      diag = c(1, 1.2, 1, 0.7, 0.7),
      upper_tri = c(0.5, 0.7, 0.9, 0, 0, 0, 0, 0, 0, 0.7)
    ),
    covariance_ext = covariance_matrix(
      diag = c(1, 1.2, 1, 0.7, 0.7),
      upper_tri = c(0.5, 0.7, 0.9, 0, 0, 0, 0, 0, 0, 0.7)
    )
  ),
  transformations = list(
    category1 = binary_cutoff("covariate1", int_cutoff = 0.45, ext_cutoff = 0.65),
    category2 = binary_cutoff("covariate2", int_cutoff = 0.55, ext_cutoff = 0.55),
    category4 = binary_cutoff("covariate4", int_cutoff = 0.50, ext_cutoff = 0.50)
  )
)
```

psborrow ではこの時点で共変量の分布に加えて評価するハザード比を設定するが、psborrow2では後のデータ生成の手順で、共変量データの生成及び評価するハザード比を設定する。

## Flow 2: Set trial operational parameters and Simulate survival times

次のフローでは生存時間の従う関数、治験の登録・脱落に関する情報と生存時間・打ち切りの情報を設定する。Flow 1:で作成した患者背景が生存時間に与える程度として各共変量に対する係数を指定し、その後に生存時間の従う関数を指定する。

治験の登録期間、観察期間、脱落情報として下記を仮定し、パイプ演算子を用いて1つのオブジェクト指向の変数内（data\_sim）にまとめる。

- 新規試験、外部対照データでは登録期間24ヶ月内に均一に登録される
- Clinical cut off：追跡期間は36ヶ月とし最後の被験者の登録から36ヶ月でデータカットオフ
- 脱落における想定：1年あたり0.05人の脱落が発生すると想定

```
#Generating Survival Data setting
data_sim <- create_data_simulation(
  baseline = baseline,
  coefficients = c(category1 = 0.5, category2 = 0.5, covariate3 = 0.5, category4 = 0.5, covariate5 = 0.5),
  event_dist = create_event_dist(dist = "weibull", lambdas = log(2) / (24), gammas = 0.9)
)
# Enrollment setting
```

```

) %>% set_enrollment(
  internal = custom_enrollment(fun = function(n) runif(n, min = 0, max = 24), label = "Uniform
enrollment distribution"),
  external = custom_enrollment(fun = function(n) runif(n, min = 0, max = 24), label = "Uniform
enrollment distribution")
# Clinical cut-off setting
) %>% set_cut_off(
  internal = cut_off_after_last(time = 36),
  external = cut_off_after_last(time = 36)
) %>%
# Drop out setting
set_dropout(
  internal_treated = create_event_dist(dist = "exponential", lambdas = 1 - 0.05/12),
  internal_control = create_event_dist(dist = "exponential", lambdas = 1 - 0.05/12),
  external_control = create_event_dist(dist = "exponential", lambdas = 1 - 0.05/12)
)

```

ここで、ランダムな脱落の設定における注意点を明記する。psborrow2パッケージでは、イベント時間と打ち切り時間を比較した上で、イベントと打ち切りの判定を含めた採用する時間を決定している。しかし、psborrow2のバージョン v0.0.3.4では、イベント時間と打ち切り時間の大小関係に応じたイベントと打ち切りの判定及びその時間の取り扱いが逆転している。具体的なデータの取り扱いに関するプログラムの箇所を以下に示す。そのため、本パッケージでのランダムな脱落の設定は困難であり、今回のシミュレーション実験では代案として設定するハザードから1を引いた値を用いた。psborrow2パッケージの当該プログラムの詳細は simulate\_data.R を参照されたい。

```

# Generate drop out times
if (length(dropout@params)) {
  drop_df <- tryCatch(
    do.call(simsurv::simsurv, args = c(dropout@params, list(x = data))),
    error = function(e) {
      cat("Error caught when generating drop out times. Check parameters:¥n")
      print(dropout@params)
      stop(e)
    }
  )
  drop_flag <- data$eventtime < drop_df$eventtime
  data$eventtime <- ifelse(drop_flag, drop_df$eventtime, data$eventtime)
  data$status <- ifelse(drop_flag, 0, data$status)
}

```

以上の情報をもとに、generate 関数を用いて、各被験者の生存時間データをシミュレートする。このとき、評価する治療群間差 (HR) とデータ間の差 (Drift HR) も設定する。

```
data list <- generate(data sim, n = 1000, treatment hr = c(1, 0.65), drift hr = c(1, 1.2))
```

### Flow 3: Choose borrowing method and run Bayesian analysis

評価する解析モデルを設定する。構成として、ベースラインモデルの指定、治療効果モデルの指定、共変量モデルの指定、borrowing の方法の指定を行う。

ベースラインモデルの指定では、アウトカムの従う分布に応じた関数を用いて、アウトカムデータの特定とアウトカムの従う分布のパラメータの事前分布を指定する。本シミュレーション実験ではワイブル分布に従うと仮定し、形状パラメータの事前分布に指数分布

を設定し、尺度パラメータの事前分布に正規分布を設定した。なお、正規分布に渡す引数は平均と標準偏差であることに注意する。

```
outcome <- outcome_surv_weibull_ph(
  time_var = "eventtime",
  cens_var = "cens",
  baseline_prior = prior_normal(0, sqrt(1000)),
  shape_prior = prior_exponential(1)
)
```

次に治療効果モデルの指定を行う。具体的には、治療群を表すダミー変数を特定し、その係数の事前分布を指定する。

```
treatment <- treatment_details(
  trt_flag_col = "trt",
  trt_prior = prior_normal(0, sqrt(1000))
)
```

次に共変量モデルの指定を行うが、基本的には治療効果モデルの指定と同じで、係数の事前分布の指定が主な作業である。

```
covariate <- sim_covariate_list(
  list(
    "None" = NULL,
    "All" = add_covariates(c("category1", "category2", "covariate3", "category4", "covariate5"),
      prior_normal(0, sqrt(1000)))
  )
)
```

最後に `borrowing` の方法を指定する。シミュレーション実験では、CMP 法の評価に際して、外部対象データを利用しない `no borrowing` 法と外部対照データの全てを利用する `full borrowing` 法との比較を実施する。この3つのモデルについて以下の手順にてそれぞれ指定する。CMP 法の設定では、 $\tau$  の事前分布の指定もこの時点で行う。

```
borrowing_list <- sim_borrowing_list(
  list(
    "No borrowing" = borrowing_none("ext"),
    "CMP" = borrowing_hierarchical_commensurate(ext_flag_col = "ext", prior_gamma(alpha = 1, beta = 0.001)),
    "Full borrowing" = borrowing_full("ext")
  )
)
```

Flow 2: で作成したデータと指定した解析モデルを用いて、Stan で解析できる状態にする。

```
sim_obj <- create_simulation_obj(
  data_matrix_list = data_list,
  outcome = outcome,
  covariate = covariate,
  borrowing = my_borrowing_list,
  treatment = treatment
)
```

作成されたシミュレーション実験用のオブジェクト指向の変数 (`sim_obj`) を `mcmc_sample()` に渡して、ベイズ推測を実施する。ここで、`parallel_chains` は `rstan` にて用意される `chain` 別に並行で `mcmc` を実行するための宣言であり、若干の計算の高速化が期待される。

```
results <- mcmc_sample(
  sim_obj,
  posterior_quantiles = c(0, 0.975),
  iter_warmup = 1000,
  iter_sampling = 10000,
  chains = 2,
  parallel_chains = 2,
  seed = 2024
)
```

mcmc\_sample()にて推測された結果 results を要約するために、get\_results(results)に渡す。その結果は以下の通り表示される。

```
simulation_res <- get_results(results)
simulation_res
```

sim_id	treatment_hr	drift_hr	n_datasets_per_param	outcome_scenario	borrowing_scenario
1	1	1	1000	default	No borrowing
2	1	1	1000	default	CMP
3	1	1	1000	default	No borrowing
4	1	1	1000	default	CMP

covariate_scenario	treatment_scenario	trt_var	mse_mean	bias_mean	null_coverage	true_coverage
No adjustment	default	0.0421208	0.1089941	0.0563908	0.943	0.943
No adjustment	default	0.0288338	0.0615265	0.0155753	0.947	0.947
All	default	0.0434742	0.1148511	0.0591642	0.94	0.94
All	default	0.0306729	0.0692940	0.0339503	0.95	0.95

バイアス及び MSE は出力の通りの結果を用いる。null\_coverage は posterior\_quantiles で指定する区間が1を含む回数を示しており、true\_coverage は posterior\_quantiles で指定する区間が真の治療群間差 (HR) を含む回数を示している。本報告書に示す結果では、第一種の過誤確率及び検出力は1- null\_coverage より算出した。

最後に EHSS の算出において、1回あたりの結果ごとに EHSS を算出し平均化する必要がある。具体的には、simulation\_res に格納される trt\_var はシミュレーション回数で平均した要約値であるため、要約される前の trt\_var の値を用いて EHSS の算出を行う。算出の詳細は、CRAN の Articles 4. Conduct a simulation study を確認いただきたい。シミュレーションシナリオによらず汎用的な利用を期待したプログラム例を以下に示す。

```
var_mat <- do.call(rbind, results@results$trt_var)
N_internal <- baseline@n_trt_int+baseline@n_ctrl_int
No_borrowing_num
  <- which(sim_obj@guide[, "borrowing_scenario"] == No_borrowing_name)
scenario_num <- nrow(data_list@guide)
No_borrowing_vec
```

```

<- as.vector( unlist(lapply(split(No_borrowing_num,
                             gl(length(No_borrowing_num)/scenario_num,scenario_num)),
                             function(x) rep(x, nrow(sim_obj@borrowing@guide))))
)
simulation_res$EHSS <- rowMeans(var_mat[No_borrowing_vec,]/var_mat - 1)*N_internal
    
```

### 3.2.5 Case simple example results

#### 3.2.5.1 本項の位置づけ

3.2.3項では、ベイズ流ワイブル回帰モデルに基づく commensurate prior (CMP) 法を用いた Bayesian dynamic borrowing (BDB) の頻度論的動作特性の評価方法を説明した。シミュレーションは既存の R package である `psborrow` 及び `psborrow2` を用いて<sup>26,28</sup>、内部対照群と外部対照群に重要な予後因子に偏りがある場合並びに両集団間でアウトカム（ハザード比）が異なる場合で、第一種の過誤確率、検出力及び推定値のバイアスをそれぞれ評価した。これらの package では検討に必要な多くの機能が利用可能であるものの、CMP 法に基づく BDB アプローチを検証的試験で適用する場合には規制当局との合意が必要不可欠であり、検討には多くのシミュレーションの実施が必要となる。詳細は **Complex Innovative Design** に関する FDA ガイダンスを参照されたい<sup>25</sup>。このため、既存の package では適用できない設定下で試験デザインの動作特性の評価が求められる可能性も否定できない。そこで、本項では既存の package を使用せず、ベイズ流ワイブル回帰モデルに基づく CMP 法を適用するための基本的な考え方の説明に焦点を当てる。なお、シミュレーションの評価指標とその解釈は package の利用の有無によらず、3.2.3項で示したとおりである。このため、本項ではデータ解析に焦点を当て、**simple example** を与えることとする。具体的には、最も単純な状況を想定したベイズ流ワイブル回帰モデルに基づく CMP 法の数理を説明し、R (RJags、RStan) 及び SAS の双方で解析のためのプログラムを解説する。考え方の説明を目的とするため、共変量は治療効果のみを指定するが、その他の重要な予後因子を共変量に含めて解析することも容易に拡張可能である。また、シミュレーションによる評価指標は、3.2.3項の内容を参考に繰り返し処理を適用することで算出可能である。なお、本項の目的に鑑み、特に断りが無い限り全スクリプトを本文中に記載する。

本項の記載と併せて、数理的な詳細は Hobbs et al. (2011) や野村ら (2022) を、適用事例については Lewis et al. (2019) を参照されたい<sup>8,22,27</sup>。本項の内容が CMP 法の理解のための自学習にも役立てば幸いである。

#### 3.2.5.2 本項で使用する仮想データ

本項では、抗がん剤における time to event 型のアウトカムに対する仮想データを用いる (ファイル名: `data_CMP.csv`)。具体的な構造を表 3.3 に示す。ただし、表 3.3 はデータセットの一部を示す。このデータは、被験者番号 (`patient`)、試験治療の有無 (`trt`: 試験治療1、

標準治療0)、外部対照か否か (ext: 外部データ1、臨床試験データ0)、生存時間 (time)、打切りの有無 (cnsr: 打切り1、イベント0) の5つの変数を有し、内部対照群及び試験治療群を含む臨床試験データとして300例、外部対照群として100例の計400例からなるデータセットである。なお、下線部は参照水準を示す。また、CSV形式のデータの読み込み並びに survfit 関数及び survminer 関数の適用のために必要となるデータのフォーマットを指定する R スクリプトを以下に示す。また、対照群データ (内部対照群+外部対照群) 及び臨床試験データ (内部対照群+試験治療群) のみを抽出したデータセットも用意しておく。

表 3.3 仮想データ

patient	trt	ext	time	cnsr
1	1	0	27.86176	0
2	0	0	20.15954	0
3	1	0	39.81461	0
4	1	0	16.83621	0
5	0	0	36.20025	1

```
data <- readr::read_csv(file = "data_CMP.csv",
  col_types = cols(trt = readr::col_factor(levels = c("0", "1")),
  ext = readr::col_factor(levels = c("0", "1")))
)
data.nb <- data %>% filter(ext == 0) # No borrowing dataset
data.C <- data %>% filter(trt == 0) # Control arm dataset
```

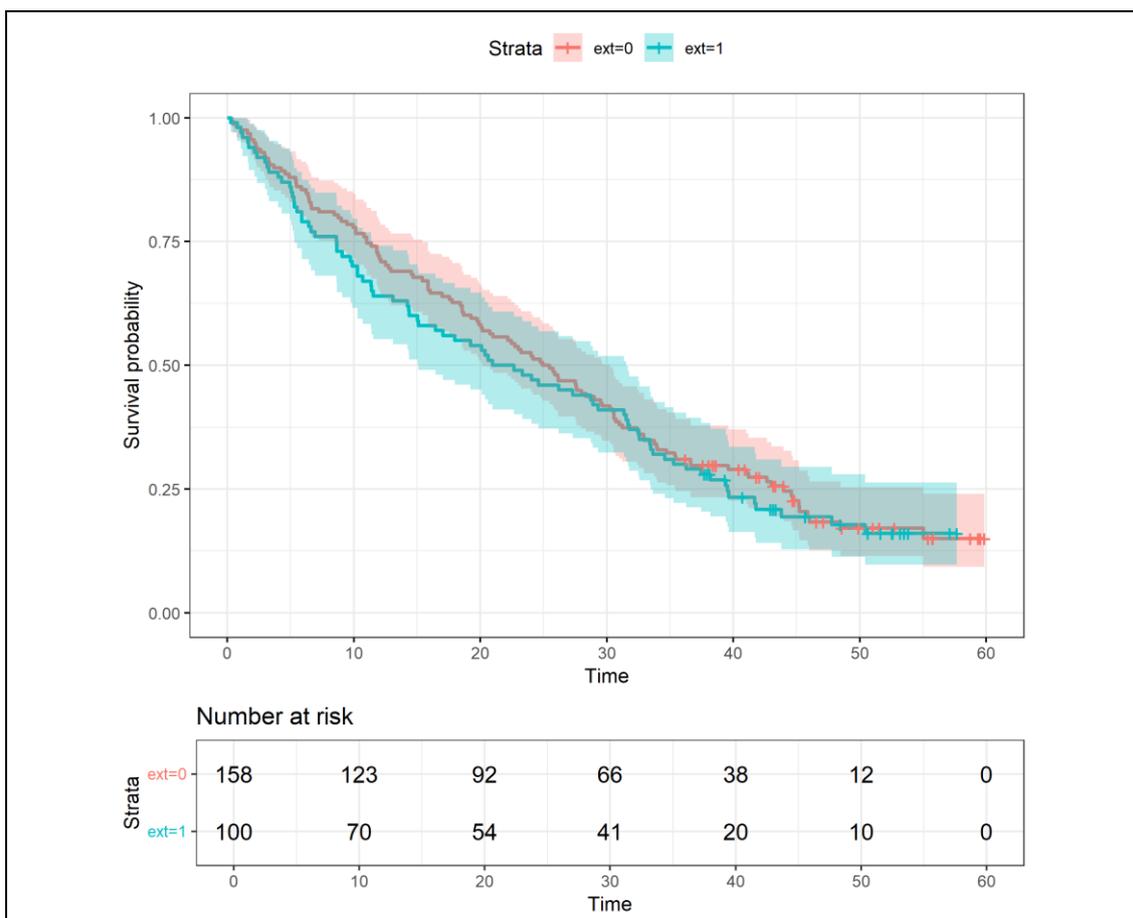
仮想データに対して、外部対照群 (EC、External Control) 及び内部対照群 (IC、Internal Control) の生存時間分布を確認する。本項では、治療効果を示すアウトカムにハザード比 (HR) を用いる。生存時間の確率分布にはワイブル分布を仮定し、パラメトリックに EC 群に対する IC 群のハザード比を推定するための R スクリプト並びに出力結果を以下に示す。

Kaplan Meier 推定

```
est_km.C <- survfit(Surv(time, 1 - cnsr) ~ ext, data = data.C)
kmpplot.C <- ggsurvplot(data = data.C,
  fit = est_km.C,
  size = 1,
  conf.int = TRUE,
  censor = TRUE,
  risk.table = TRUE,
  ggtheme = theme_bw()
)
```

出力結果

```
Call: survfit(formula = Surv(time, 1 - cnsr) ~ ext, data = data.C)
n events median 0.95LCL 0.95UCL
ext=0 158 125 25.2 20.2 30.3
ext=1 100 81 21.8 15.1 31.5
```



### 加速モデル

```
res.C <- survreg(Surv(time, 1 - cnsr) ~ ext, dist = "weibull", data = data.C)
summary(res.C)
```

### 出力結果

```
Call:
survreg(formula = Surv(time, 1 - cnsr) ~ ext, data = data.C,
  dist = "weibull")
      Value Std. Error      z      p
(Intercept)  3.4802    0.0816 42.65 <2e-16
ext1         -0.0825    0.1302 -0.63  0.53
Log(scale)  -0.0918    0.0601 -1.53  0.13
Scale= 0.912

Weibull distribution
Loglik(model)= -914.4  Loglik(intercept only)= -914.6
  Chisq= 0.4 on 1 degrees of freedom, p= 0.53
Number of Newton-Raphson Iterations: 5
n= 258
```

### ハザードの推定

```
res.ph.C <- exp(- res.aft.C$coefficients / res.aft.C$scale)[2]
```

注：パラメトリック生存時間解析を実施するための `survreg` package では、加速モデルによる `parameterization` を前提としている。ワイブル分布は加速性と比例ハザード性を同時に有する唯一の分布であり、比例ハザードモデルの下で回帰係数を解釈するため、変換の処理を追加する。

### 出力結果

```
ext1
1.094619
```

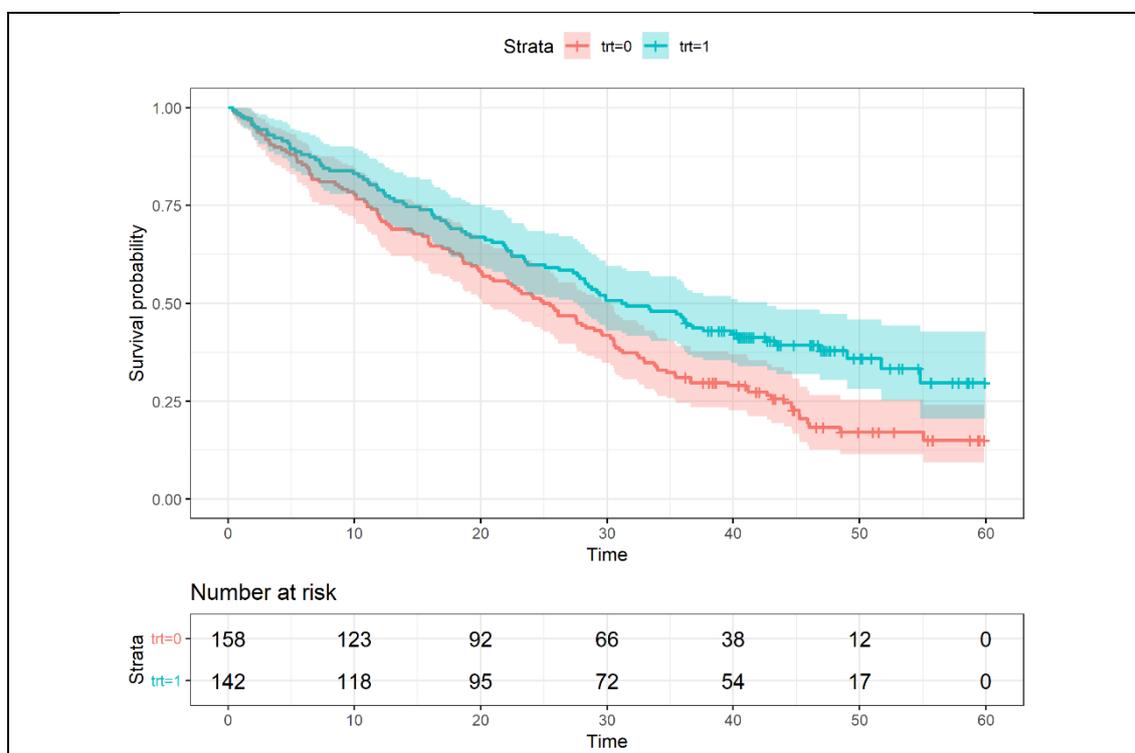
また、CMP 法の利用に先立ち、外部対照のデータを借用しない場合 (No borrowing) と外部対照のデータをすべて借用した場合 (Full borrowing) のそれぞれに対する Kaplan Meier 推定及びパラメトリックなベイズ流ワイブル回帰モデルを適用した結果を以下に示す。なお、出力に必要な R code は上記のコードに対して参照先のデータセット及び群情報の変更のみに留まるため、R スクリプトは省略する。出力結果より、No borrowing 及び Full borrowing のそれぞれにおいて、標準治療に対する試験治療の HR は0.671及び0.649と推定された。

### No borrowing

#### Kaplan Meier 推定

```
Call: survfit(formula = Surv(time, 1 - cnsr) ~ trt, data = data.nb)
```

	n	events	median	0.95LCL	0.95UCL
trt=0	158	125	25.2	20.2	30.3
trt=1	142	89	31.3	27.7	42.7



#### 加速モデル

```
Call: survreg(formula = Surv(time, 1 - cnsr) ~ trt, data = data.nb,
  dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	3.4802	0.0817	42.58	<2e-16
trt1	0.3651	0.1276	2.86	0.0042
Log(scale)	-0.0904	0.0600	-1.51	0.1323

Scale= 0.914

Weibull distribution

Loglik(model)= -992 Loglik(intercept only)= -996.2  
 Chisq= 8.41 on 1 degrees of freedom, p= 0.0037  
 Number of Newton-Raphson Iterations: 5  
 n= 300

ハザード比の推定

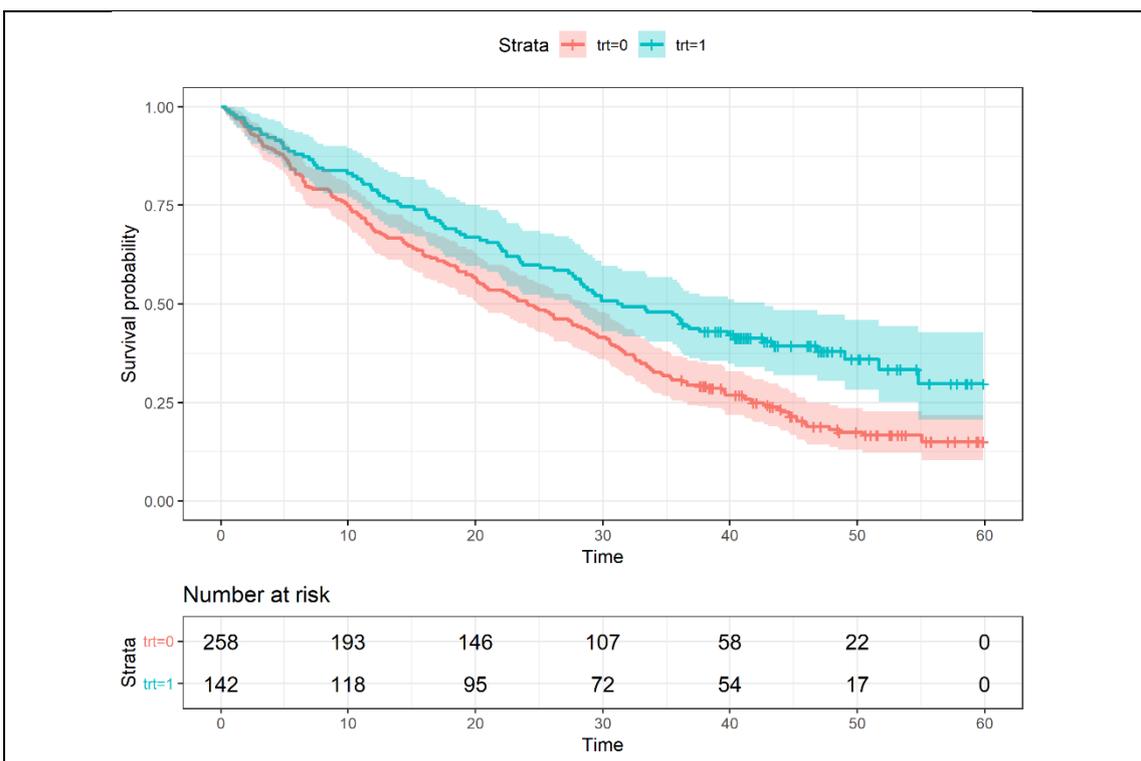
```
trt1
0.6705592
```

**Full borrowing**

Kaplan-Meier 推定

```
Call: survfit(formula = Surv(time, 1 - cnsr) ~ trt, data = data)

      n events median 0.95LCL 0.95UCL
trt=0 258   206   24.1   20.2   28.7
trt=1 142    89   31.3   27.7   42.7
```



加速モデル

```
Call:
survreg(formula = surv(time, 1 - cnsr) ~ trt, data = data, dist = "weibull")
      Value Std. Error      z      p
(Intercept)  3.4478    0.0648  53.21 < 2e-16
trt1         0.4015    0.1187   3.38 0.00072
Log(scale)  -0.0732    0.0508  -1.44 0.14969

Scale= 0.929

Weibull distribution
Loglik(model)= -1347.9  Loglik(intercept only)= -1354
  Chisq= 12.16 on 1 degrees of freedom, p= 0.00049
Number of Newton-Raphson Iterations: 5
n= 400
```

ハザード比の推定

```
trt1
0.6492361
```

### 3.2.5.3 ベイズ流ワイブル回帰モデルに基づく CMP 法とその実装

本章では、EC と IC のアウトカムに対する乖離度に応じて借用程度を調整することを目的として、ベイズ流ワイブル回帰モデルに基づく CMP 法の数理及び解析方法を説明する。いま、 $T_i$  を被験者  $i$  の time to event 型アウトカムの確率変数とする。また、 $X_{1i}$  を外部対照データか否か、 $X_{2i}$  を試験治療の有無を表すダミー変数とする。比例ハザードモデルでのワイブル分布に基づくパラメトリック推定として、以下の統計モデルを用いる。

$$T_i \sim \text{Weibull}(\lambda_i, r)$$

ただし、

$$\lambda_i = \exp\{\beta_0(1 - X_{1i}) + \beta_{\text{ext}}X_{1i} + \beta_{\text{trt}}X_{2i}\}.$$

また、パラメータの事前分布を以下のように設定する。

$$\begin{aligned} \beta_{\text{ext}} &\sim N(0, 1000), \\ \beta_{\text{trt}} &\sim N(0, 1000), \\ \tau &\sim \text{gamma}(1, 0.001), \\ \beta_0 &\sim N(\beta_{\text{ext}}, \tau), \\ r &\sim \exp(1). \end{aligned}$$

$\beta_{\text{ext}}$  は EC に対する切片項であり、 $\beta_0$  は IC に対する切片項である。ここで、超事前分布、事前分布及び  $n$  個の観測された生存時間  $t_1, \dots, t_n$  の尤度を **Error! Reference source not found.** 図 3-3**Error! Reference source not found.** のように整理する。なお、 $\tau$  に設定する分布に応じて借用の程度が変化するため、設定には慎重な議論が必要であるが、ここではガンマ分布を想定する。他にも EC と IC のアウトカムの違いの程度が大きいと仮定する場合には半コーシー分布等も設定し得る。

以下に、ベイズ流ワイブル回帰モデルにおける CMP 法を適用するための R (RJags, RStan) 及び SAS で実施する code を示す。**Error! Reference source not found.** 図 3-3**Error! Reference source not found.** の構造を直接的に記述することで、事後分布の MCMC サンプルングを容易に得ることができる。いずれの場合でも、必要な工程として model 及び対数尤度の指定等共通する点が多く、以降では RJags を中心に解説する。ただし、ワイブル分布の確率密度関数の parameterization はソフトウェアごとに異なる点には注意が必要である。なお、MCMC のサンプルングアルゴリズムはそれぞれ default の指定に従うこととする。

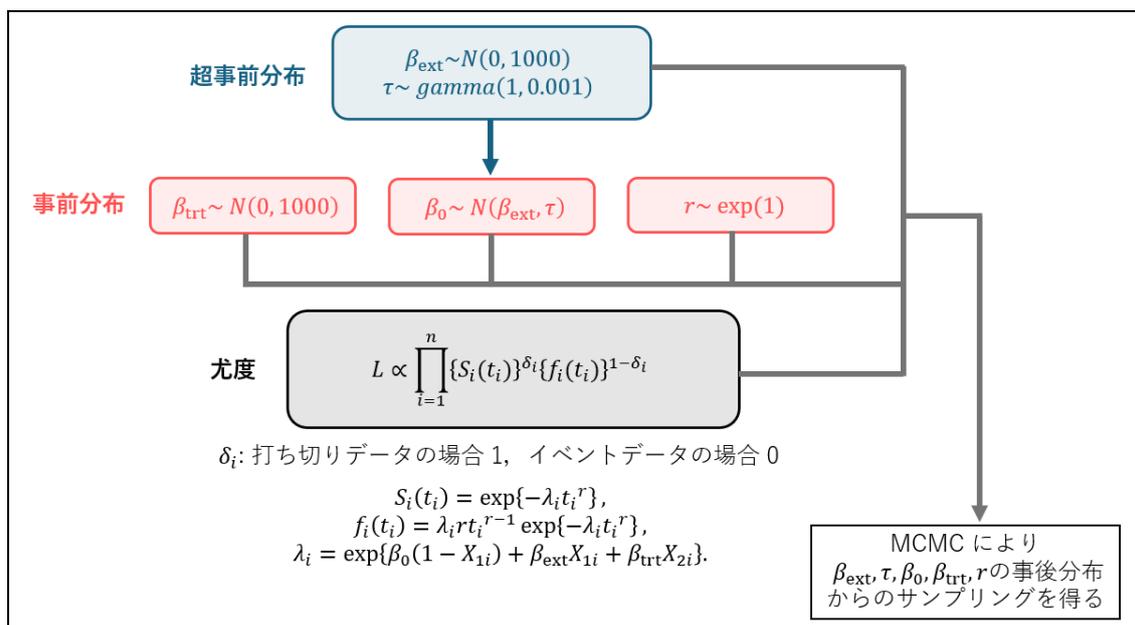


図 3-3 Error! Reference source not found.ベイズ流階層モデルの構造

RJags: 11 weibull cmp.R

```
# 1.Pre-processing
data <- read.csv("data_CMP.csv") %>%
  as.data.frame() %>%
  arrange(cnsr)

# 2.Definition of the model (Jags code)
model_code <- "
model{
  for (i in 1:n.obs){
    time[i] ~ dweib(r, mu.obs[i])
    mu.obs[i] <- exp(beta0 * (1 - ext[i]) + beta.ext * ext[i] + beta.trt * trt[i])
  }

  for (i in (n.obs + 1):n){
    cnsr[i] ~ dbern(S[i])
    S[i] <- 1 - pweib(time[i], r, mu.cens[i])
    mu.cens[i] <- exp(beta0 * (1 - ext[i]) + beta.ext * ext[i] + beta.trt * trt[i])
  }

  #Priors
  beta.ext ~ dnorm(0.0, 0.001)
  beta0 ~ dnorm(beta.ext, tau)
  beta.trt ~ dnorm(0.0, 0.001)
  tau ~ dgamma(1, 0.001)
  r ~ dexp(1)
}
"

# 3.Set up data for MCMC
model_data <- list(
  n.obs = sum(data[, "cnsr"] == 0), n = nrow(data), time = data[, "time"],
  trt = data[, "trt"], cnsr = data[, "cnsr"], ext = data[, "ext"]
)

# 4.Choose parameter to monitor in MCMC
model_parameters <- c("beta0",
  "beta.trt",
  "beta.ext",
  "tau",
  "r")

# 5.Run the model
set.seed(2024)
model_run_cmp <- jags(
  data = model_data,
```

```

parameters.to.save = model_parameters,
model.file         = textConnection(model_code),
n.chains           = 4,
n.iter            = 100000,
n.burnin          = 10000,
n.thin            = 10
)

# 6.Diagnosis of convergence
png("./output/11_traceplot_cmp.png", width = 1500, height = 800)
R2jags::traceplot(model_run_cmp, varname = c("beta.trt"), ask = FALSE)
dev.off()

# 7.Summary of results
post <- print(model_run_cmp)
PostBetaTrt <- data.frame(beta.trt = post$sims.matrix[, "beta.trt"])
beta.trt_plot <- ggplot(PostBetaTrt, aes(x = beta.trt)) + geom_density(adjust = 1)

```

RJags を用いて MCMC サンプリングに基づく事後分布を計算するプロセスを以下に示す。

1. Pre-processing
  - 2.で記述する尤度を記述するため、打切りの有無 (csnr) でソートしておく。
2. Definition of the model (Jags code)
 

Jags code を記述する。ダブルクォーテーションで囲み、文字変数として用意する。具体的には、model 文に **Error! Reference source not found.図 3-3Error! Reference source not found.**で示した尤度と事前及び超事前分布を指定する。

注1：打切例に対する生存関数の記法は、ベルヌーイ分布を想定することで $S_i(t_i)$ を構成している。ベルヌーイ分布の確率密度関数は、 $S_i(t_i)^{\delta_i}(1 - S_i(t_i))^{1-\delta_i}$ とかける。ここに、iteration は打切例に対してのみ記述することとしているため、本質的には $\delta_i = 1$ のみを寄与させている。

注2：Jags では、パラつきを示すパラメータの指定にあたって分散ではなく精度を指定する必要がある。このため、例えば dnorm(0.0, 0.001)は平均0、分散1000の正規分布に対応する。
3. Set up data for MCMC
  - 2.で記述した Jags code に対して、参照するデータ (変数名) を指定する。
4. Choose parameter to monitor in MCMC
 

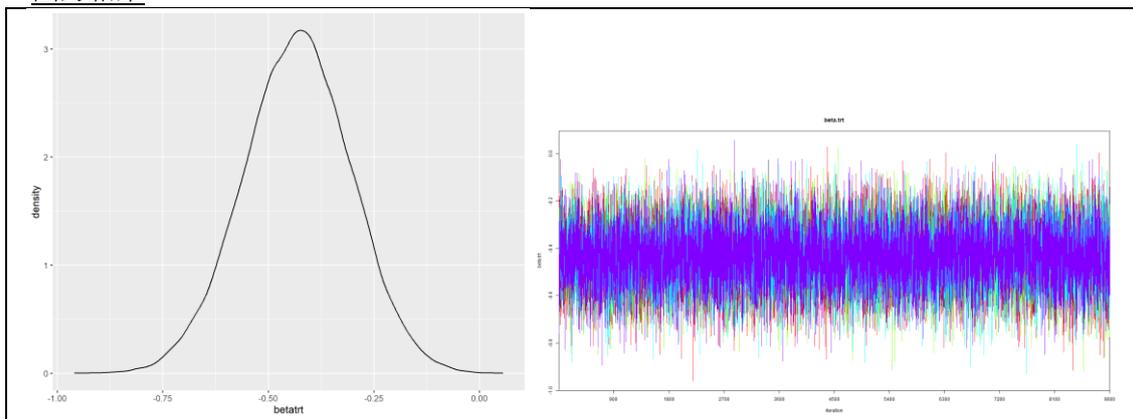
事後分布からの MCMC サンプリングを得たいパラメータを指定する。
5. Run the model
 

MCMC の Option を指定する。textConnection 関数の引数に2で記述した Jags code を指定し、Jags に引き渡す。
6. Diagnosis of convergence
 

MCMC の収束診断のためのトレースプロットを描画する。
7. Summary of results
 

事後分布からの MCMC サンプリングの要約統計量を出力する。また、関心のあるパラメータの事後分布を描画する。

出力結果



Inference for Bugs model at "3", fit using jags,  
 4 chains, each with 1e+05 iterations (first 10000 discarded), n.thin = 10  
 n.sims = 36000 iterations saved

	mu.vect	sd.vect	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
beta.ext	-3.708	0.206	-4.120	-3.844	-3.702	-3.568	-3.311	1.002	4000
beta.trt	-0.432	0.127	-0.685	-0.517	-0.430	-0.346	-0.184	1.001	19000
beta0	-3.718	0.205	-4.129	-3.854	-3.713	-3.578	-3.328	1.002	3800
r	1.076	0.055	0.972	1.039	1.075	1.113	1.185	1.001	4800
tau	1033.955	999.154	41.410	321.845	733.631	1425.851	3724.537	1.001	36000
deviance	2698.914	2.457	2695.995	2697.120	2698.309	2700.066	2705.267	1.001	16000

For each parameter, n.eff is a crude measure of effective sample size,  
 and Rhat is the potential scale reduction factor (at convergence, Rhat=1).  
 DIC info (using the rule,  $pD = \text{var}(\text{deviance})/2$ )  
 $pD = 3.0$  and  $\text{DIC} = 2701.9$   
 DIC is an estimate of expected predictive error (lower deviance is better).

RStan

Stan ファイル: 21\_weibull\_cmp.stan

```

data {
  int <lower = 0> N;
  array[N] real time;
  array[N] int cnsr;
  array[N] int trt;
  array[N] int ext;
}

parameters {
  real beta_ext;
  real <lower = 0> tau;
  real beta_trt;
  real beta0;
  real <lower = 0> r;
}

transformed parameters {
  array[N] real lambda;
  for (n in 1:N){
    lambda[n] = exp(beta0*(1 - ext[n]) + beta_ext * ext[n] + beta_trt * trt[n]);
  }
}

model {
  beta_ext ~ normal(0, 1000);
  tau ~ gamma(1, 0.001);
  beta0 ~ normal(beta_ext, (1/tau));
  beta_trt ~ normal(0, 1000);
  r ~ exponential(1);

  for(n in 1:N){
    if(cnsr[n] == 1){
      target += weibull_lccdf(time[n]| r, (1/lambda[n])^(1/r));
    }else{
      target += weibull_lpdf(time[n]| r, (1/lambda[n])^(1/r));
    }
  }
}
    
```

```
}

```

R ファイル: 22 weibull cmp stan.R

```
data_CMP <- read.csv("data_CMP.csv") %>% as.data.frame()

data <- list(
  N = nrow(data_CMP),
  time = data_CMP$time,
  cnsr = data_CMP$cnsr,
  trt = data_CMP$trt,
  ext = data_CMP$ext
)

fit1 <- stan(
  file = "./R/21_weibull_cmp.stan",
  data = data, # A named list of data
  chains = 4, # The number of Markov chains
  warmup = 10000, # The number of warmup iterations per chain
  iter = 100000, # Total number of iterations per chain
  thin = 10, # The period for saving samples
  cores = 1, # The number of cores
  refresh = 0, # No progress shown
  seed = 2024 # The seed for random number generation
)
```

注：上記は rstan package を用いて事後分布からの MCMC サンプリングを行っているが、cmdstanr package を適用することも可能である。

出力結果 (一部抜粋)

```
Inference for Stan model: anon_model.
4 chains, each with iter=1e+05; warmup=10000; thin=10;
post-warmup draws per chain=9000, total post-warmup draws=36000.
```

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%
n_eff								
Rhat								
beta_ext	-3.69	0.01	0.20	-4.11	-3.82	-3.69	-3.55	-3.32
384	1.01							
tau	476.35	109.65	392.55	20.54	174.51	380.99	653.96	1479.19
13	1.11							
beta_trt	-0.45	0.01	0.13	-0.70	-0.54	-0.45	-0.36	-0.20
314	1.01							
beta0	-3.69	0.01	0.20	-4.11	-3.82	-3.69	-3.55	-3.32
382	1.01							
r	1.07	0.00	0.05	0.97	1.04	1.07	1.11	1.18
424	1.01							

SAS: 31 weibull cmp.sas

```
/* ****には data_CMP.csv を格納したフォルダのパスを指定する */
proc import out = data_CMP
  datafile = "****/data_CMP.csv"
  dbms= csv replace;
run;

title 'weibull Survival Model (CMP)';
proc mcmc data = data_CMP outpost = weisurvout nbi = 10000 nmc = 110000 thin = 10 seed
= 1234
  monitor = (_parms_) stats = (summary intervals);

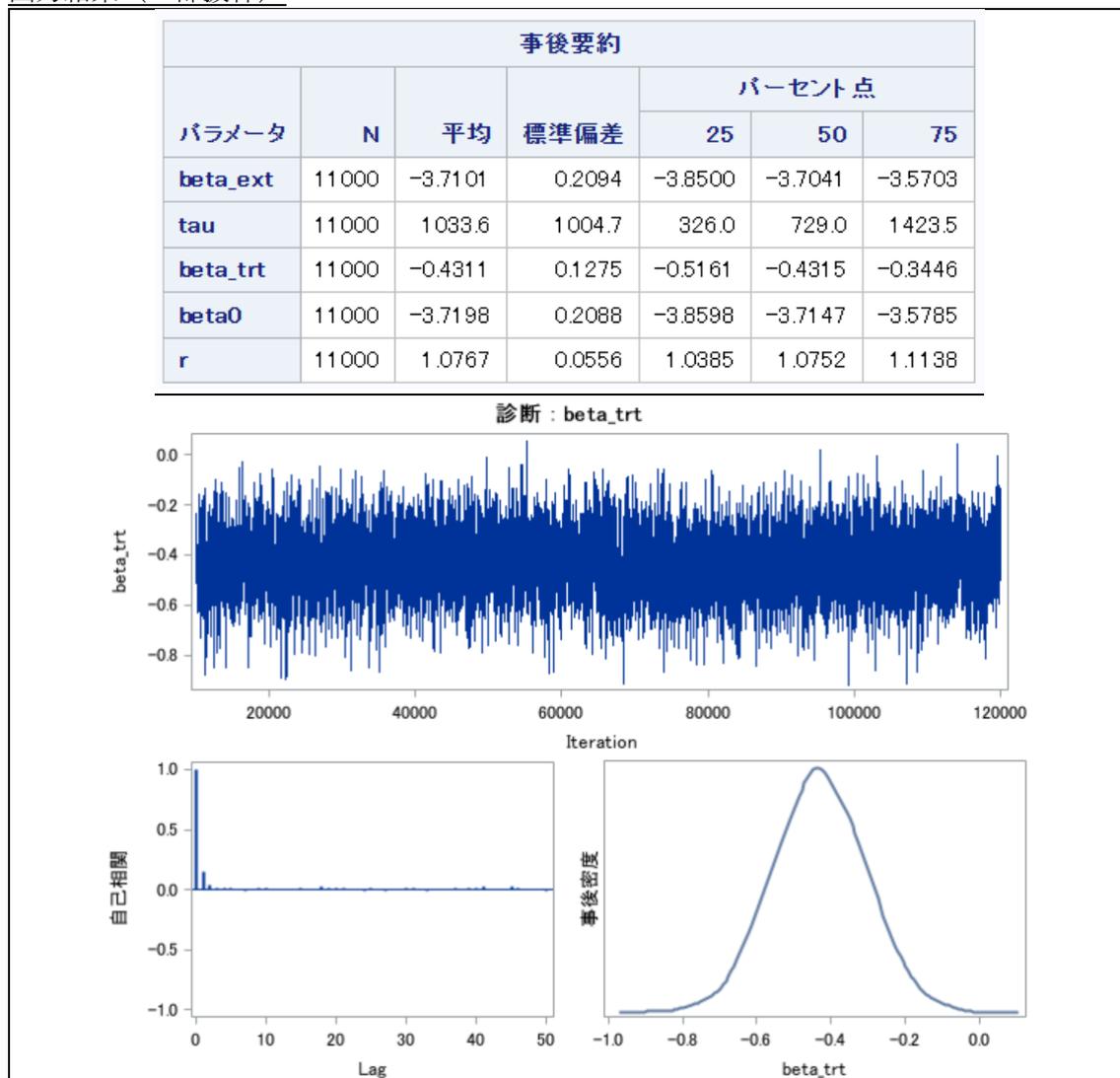
  parms beta_ext 0;
  parms tau 1;
  parms beta_trt 0;
  parms beta0 0;
  parms r 1;

  hyperprior beta_ext: ~ normal(0, prec = 0.001);
  hyperprior tau: ~ gamma(1, is = 0.001);
  prior beta_trt: ~ normal(0, prec = 0.001);
  prior beta0: ~ normal(beta_ext, prec = tau);
  prior r: ~ expon(is = 1);

  lambda = beta0*(1-ext) + beta_ext*ext + beta_trt*trt;
  mu = (1 / exp(lambda) )**(1/r);

  llike = (1-cnsr)*logpdf('weibull', time, r, mu) + cnsr*logsd('weibull', time, r,
mu);
  model general(llike);
run;
```

出力結果（一部抜粋）



### 3.2.5.4 Effective Historical Sample Size (EHSS) とその実装

本項では、事後分布の推測に事前分布がどの程度影響を与えるかを評価することを目的として、Effective historical sample size (EHSS) を評価する。EHSS は、事後 Effective sample size (ESS) からデータの対象者数を差し引く方法である。ここでは対数ハザード比のスケールで EHSS を評価することとし、分散比に基づく事後 ESS を用いて EHSS を算出する。具体的な EHSS の算出式を以下に示す。なお、対照群のベースラインハザードに対して事後 ESS を算出することもしばしば行われる。EHSS はベイズ流の枠組みで評価され、外部データの実際の利用程度を被験者数のスケールで反映する。詳細は野村ら (2022)等を参照されたい<sup>8</sup>。

$$N_{CT} \left( \frac{\text{Prec}(\beta_{trt}|\text{all data})}{\text{Prec}(\beta_{trt}|\text{clinical trial data})} - 1 \right).$$

ただし、 $N_{CT}$ は臨床試験データの被験者数であり、 $\text{Prec}(\cdot)$ は推定精度を示す ( $1/\text{sd}^2$ )。EHSS の分子に寄与する $\text{Prec}(\beta_{\text{trt}}|\text{all data})$ は、CMP 法により得られた $\beta_{\text{trt}}$ の事後分布の推定精度であり、分母に寄与する $\text{Prec}(\beta_{\text{trt}}|\text{clinical trial data})$ は外部対照データを用いずに臨床試験データのみを用いた場合におけるベイズ流ワイブル回帰モデルの $\beta_{\text{trt}}$ の推定精度である。以降、 $\text{Prec}(\beta_{\text{trt}}|\text{clinical trial data})$ を計算するための数理について補足する。ここでは、比例ハザードモデルでのワイブル分布に基づくパラメトリック推定として、以下の統計モデルを用いる。

$$T_i \sim \text{Weibull}(\lambda_i, r).$$

ただし、 $\lambda_i = \exp\{\beta_0 + \beta_{\text{trt}}X_{2i}\}$ とし、パラメータの事前分布を以下のように設定する。

$$\beta_0 \sim N(0, 1000),$$

$$\beta_{\text{trt}} \sim N(0, 1000),$$

$$r \sim \exp(1).$$

被験者数に基づく EHSS に焦点を当てたが、time to event 型アウトカムにおいては ESS ではなく、イベント数に基づく Effective number of events (ENE) にも関心がある。ENE に関する詳細は、Han et al. (2017)、Neuenschwander et al. (2020) 及び Roychoudhury and Neuenschwander (2020) 等を参照されたい<sup>29-31</sup>。他にも、expected-local-information-ratio 法 (Neuenschwander et al., 2020) 等による評価も有用である<sup>30</sup>。これらの ESS に関する詳細は、野村ら (2022) を参照されたい<sup>8</sup>。なお、本項における ESS は、MCMC 及び inverse probability of treatment weighting 法の文脈で用いられる Effective sample size とは異なる点に注意されたい。

ベイズ流ワイブル回帰モデルにおける各パラメータ ( $\beta_0$ 、 $\beta_{\text{trt}}$ 、 $r$ ) の事後分布からの MCMC サンプリングは、上記の CMP 法のスクリプトを基に容易に導出可能であるため、ここではスクリプトの説明は省略し、RJags による事後分布の要約統計量のみを出力する。

#### RJags の出力結果

Inference for Bugs model at "4", fit using jags,								
4 chains, each with 1e+05 iterations (first 10000 discarded), n.thin = 10								
n.sims = 36000 iterations saved								
	mu.vect	sd.vect	2.5%	25%	50%	75%	97.5%	Rhat
n.eff								
beta.trt	-0.401	0.140	-0.675	-0.494	-0.400	-0.307	-0.126	1.001
36000								
beta0	-3.811	0.247	-4.303	-3.976	-3.809	-3.642	-3.337	1.001
28000								
r	1.094	0.065	0.968	1.049	1.093	1.138	1.223	1.001
23000								
deviance	1987.097	2.451	1984.302	1985.309	1986.468	1988.205	1993.460	1.001
36000								

For each parameter, n.eff is a crude measure of effective sample size, and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule,  $\text{pD} = \text{var}(\text{deviance})/2$ )  
 $\text{pD} = 3.0$  and  $\text{DIC} = 1990.1$   
 DIC is an estimate of expected predictive error (lower deviance is better).

#### EHSS の計算と出力結果

```
sd_ct <- model_run_ehss$BUGSoutput$summary %>% as.data.frame()
```

```
sd_ct %<>%
  filter(rownames(sd_ct) == "beta.trt") %>%
  select(sd) %>%
  as.numeric()

sd_all <- model_run_cmp$BUGSoutput$summary %>% as.data.frame()

sd_all %<>%
  filter(rownames(sd_all) == "beta.trt") %>%
  select(sd) %>%
  as.numeric()

ehss <- num_ct * ((1/sd_all^2) / (1/sd_ct^2) - 1)

[1] 64.32353
```

### 3.2.5.5 結果の解釈と課題

仮想事例のデータは、EC と IC でアウトカムは比較的類似した傾向を示しており、No borrowing と Full borrowing におけるワイブル分布を仮定したパラメトリックな生存時間解析として、標準治療に対する試験治療の HR はそれぞれ0.671及び0.649と推定された。また、3.2.5.3項で示した階層ベイズ法に基づく CMP 法を適用した結果、標準治療に対する試験治療のハザード比は RJags より0.649 ( $=\exp(-0.432)$ ) と推定された。トレースプロットにより、MCMC サンプリングに対して自己相関も高くなく安定したサンプリングが得られていることが確認できる。CMP 法による推定結果は、No borrowing と Full borrowing の間に収まり、EHSS より外部対照データを多く借用したことが確認できる。これは Commensurate prior として EC と IC の違いの程度が小さいと仮定した場合に設定されるガンマ分布を指定したこと、及び EC と IC にアウトカムで生存時間分布に大きな差異が見られないこと等によるものと考えられ、Full borrowing と同様の結果が得られた。なお、RStan 及び SAS の MCMC procedure の結果とわずかに異なるが、結果の解釈に影響を与えるものではない。

本項で示したとおり、CMP 法は既存の package を使用せずとも容易に実装が可能であるが、シミュレーションの実施にあたっては実行時間に伴う実施上の課題がある。事後分布の導出にあたって MCMC を必要とするため、動作特性の評価に際して並列化計算を施す等の実施上の工夫は必至であろう。

### 3.2.5.6 プログラムの利用方法

本項で用意したプログラム及び実行結果は、すべて JPMA\_TF32\_ext ファイルに格納している。また、R の開発環境は以下の操作により復元可能である。ただし、R2jags 及び rstan package の利用にあたっては、それぞれ JAGS 及び RTools を事前にインストールしておく必要がある。

1. R コンソール上で `install.packages("renv", type = "win.binary")` を実行し、renv package をインストールする。
2. ディレクトリパスに2バイト文字を含まない環境で zip ファイルを展開する。
3. Rproj ファイルをダブルクリックし、RStudio を起動する。

4. R コンソール上で”`renv::restore()`”を実行する。

### 3.2.5.7 謝辞

本項の解析プログラムの作成にあたりご助言及びご協力いただいた中外製薬株式会社の澤本 涼氏、東京理科大学の大東智洋先生に感謝申し上げます。

## 3.3 参考文献

1. 厚生労働省. (2021). 承認申請等におけるレジストリの活用に関する基本的考え方. [https://www.mhlw.go.jp/web/t\\_doc?dataId=00tc5763&dataType=1&pageNo=1](https://www.mhlw.go.jp/web/t_doc?dataId=00tc5763&dataType=1&pageNo=1)
2. 厚生労働省. (2022). 「臨床試験の一般指針」の改正について. 医薬審第 1223第5号. <https://www.pmda.go.jp/files/000250244.pdf>
3. 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. "Rare disease の治療効果の推測法." (2022)
4. 厚生労働省. (2000). 小児集団における医薬品の臨床試験に関するガイダンスについて. 医薬審第 1334 号. <https://www.pmda.go.jp/files/000156072.pdf>
5. 厚生労働省. (2017). 小児集団における医薬品開発の臨床試験に関するガイダンスの補遺について. 医薬審第 1227 第 5 号. <https://www.pmda.go.jp/files/000222107.pdf>
6. ICH HARMONISED GUIDELINE. (2024). INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE. <https://www.pmda.go.jp/files/000270245.pdf>
7. 厚生労働省. (2001). 「臨床試験における対照群の選択とそれに関連する諸問題」について. 医薬審第 136 号. <https://www.pmda.go.jp/files/000156634.pdf>
8. 野村尚吾, 大東智洋, 澤本涼. (2022). Hybrid control アプローチを用いるランダム化比較試験の計画と解析：外部データが要約統計量の場合. 計量生物学 43(1), 63-96.
9. Dunbar Ivy, Damien Bonnet, Rolf MF Berger, Gisela M.B. Meyer, Simin Baygani, and Baohui Li. (2021). Efficacy and safety of tadalafil in a pediatric population with pulmonary arterial hypertension: phase 3 randomized, double-blind placebo-controlled study. *Pulmonary Circulation*, 11(3) 1–8.
10. Statistical Analysis Plan H6D-MC-LVHV (v4). [https://cdn.clinicaltrials.gov/large-docs/90/NCT01824290/SAP\\_001.pdf](https://cdn.clinicaltrials.gov/large-docs/90/NCT01824290/SAP_001.pdf)
11. ClinicalTrials.gov A Study of Tadalafil in Pediatric Participants With Pulmonary Arterial Hypertension (PAH). <https://www.clinicaltrials.gov/study/NCT01824290>
12. Ye J, Travis J. (2017). A Bayesian approach to incorporating adult clinical data into pediatric clinical trials. Food and Drug Administration, Office Biostatistics (DB V and II). <https://www.fda.gov/media/107649/download>
13. Food and Drug Administration. (2010). Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-use-bayesian-statistics-medical-device-clinical-trials>
14. Paul-Christian Bürkner (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1), 1-28. doi:10.18637/jss.v080.i01
15. Paul-Christian Bürkner (2018). Advanced Bayesian Multilevel Modeling with the R Package

- brms. The R Journal, 10(1), 395-411. doi:10.32614/RJ-2018-017
16. Paul-Christian Bürkner (2021). Bayesian Item Response Modeling in R with brms and Stan. Journal of Statistical Software, 100(5), 1-54. doi:10.18637/jss.v100.i05
  17. Landau WM, Kunzmann K, Sidi Y, Stock C (2024). `_brms.mmrm`: Bayesian MMRMs using 'brms'. R package version 1.1.1, <<https://CRAN.R-project.org/package=brms.mmrm>>.
  18. Stan Development Team (2025). RStan: the R interface to Stan. R package version 2.32.7. <https://mc-stan.org/>.
  19. 伊庭 克拓, 浅野 豊, 松嶋 優貴, 毛利 誠. (2023). BGLIMM プロシジャおよび MCMC プロシジャによるベイズ流経時測定データ解析. SAS ユーザー総会2023. <https://sas-user2023.ywstat.jp/download.html?n=10&key=paper>
  20. Pocock, S. J. (1976). The combination of randomized and historical controls in clinical trials. Journal of Chronic Diseases 29, 175-188.
  21. 武田健太郎, 大庭真梨, 橋爪智行, 坂巻頭太郎, 田栗正隆, 森田智視 (2015) 臨床試験におけるヒストリカルコントロールデータの利用. 計量生物学会36(1), 25-50
  22. Hobbs, B. P., Carlin, B. P., Mandrekar, S. J. and Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. Biometrics, 67(3), 1047-1056.
  23. 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. "Complex Innovative Trial Design の現状と事例の整理." (2023)
  24. Lu, Y., Lin, A., Pang, H. and Zhu, J. (2021). Bayesian Dynamic Borrowing Tool for Complex Innovative Trial Designs. ASA Biopharmaceutical Report, Summer 2021, Volume 28, Issue 3, 11-19
  25. Food and Drug Administration. (2023). Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products. <https://www.fda.gov/media/164960/download>
  26. Gower-Page, C., Lu, Y., Lin, A. (2023). `_psborrow`: Bayesian Dynamic Borrowing with Propensity Score. R package version 0.2.1.
  27. Lewis, C. J., Sarkar, S., Zhu, J. and Carlin, B. P. (2019). Borrowing From Historical Control Data in Cancer Drug Development: A Cautionary Tale and Practical Guidelines. Statistics in Biopharmaceutical Research. 11(1), 67-78. <https://doi.org/10.1080/19466315.2018.1497533>
  28. Secrest M and Gravestock I (2024). `_psborrow2`: Bayesian Dynamic Borrowing Analysis and Simulation. <https://github.com/Genentech/psborrow2>, <https://genentech.github.io/psborrow2/index.html>.
  29. Han, B., Zhan, J., John, Zhong Z., Liu, D. and Lindborg, S. (2017). Covariate-adjusted borrowing of historical control data in randomized clinical trials. Pharmaceutical Statistics, 16, 296-308.
  30. Neuenschwander, B., Weber, S., Schmidli, H. and O'Hagan, A. (2020). Predictively consistent prior effective sample sizes. Biometrics, 76(2), 578-587.
  31. Roychoudhury, S. and Neuenschwander, B. (2020). Bayesian leveraging of historical control data for a clinical trial with time-to-event endpoint. Statistics in Medicine, 39(7), 984-995.

## 4. 用量反応試験

本章では、用量反応試験におけるベイズ流試験デザインの活用事例として Smith et al. (2006) の試験設定を参考に、ベイズ流試験デザインのシミュレーションによる動作特性の検討を行った<sup>1</sup>。

Smith et al. (2006) は、帯状疱疹後神経痛 (postherpetic neuralgia : PHN) に対する新たな治療薬候補について、用量反応関係及びプラセボと比較した効果の大きさを評価するための proof of concept (POC) 試験である。当該試験では、試験計画時に用量反応関係を仮定をする情報が十分に集積されていなかった。そのため、比較的多い実薬用量である7用量を設定し、用量反応曲線の形状に対する強い仮定を必要としない normal dynamic linear model (NDLM) を用いて用量反応曲線の推定を行うこととした<sup>2</sup>。また、無益性の中間解析を予定し、無効な治療群への組み入れを停止することとした<sup>1</sup>。

用量反応試験における他のベイズ流試験デザインの活用例として、ベイズ流階層モデル (Bayesian hierarchical model : BHM) を用いたデザインが提案されている<sup>3</sup>。BHM では、用量反応に関する事前の仮定を置くことなく、観察された用量群間の治療効果の類似性に基づいて、用量群間で動的に情報を借用することにより、用量群間で反応が類似している場合は推定精度を上げることができると知られている。検討の参考とする Smith et al. (2006) においては NDLM を用いたベイズ流試験デザインが計画実施されたが、本章では Smith et al. (2006) の試験設定において、NDLM とは異なる特性を持つモデルとして BHM を適用する検討も実施することとした。なお、BHM が用いられた用量反応試験の事例として、全身性エリテマトーデスの第 IIb 相試験がある<sup>4</sup>。事例の詳細については、データサイエンス部会から2024年4月に報告されている「Complex Innovative Trial Design の現状と事例の整理 - FDA ガイダンス邦訳付き」の4.6節で取り上げられているので参照されたい<sup>3</sup>。

本章では、Smith et al. (2006) における試験計画に基づいて用量反応曲線の推定と POC の確認を主目的とする試験を想定し、NDLM 及び BHM を用いる場合の動作特性をシミュレーションで評価した。加えて、比較対象として、ベイズ流の非階層モデル (non-hierarchical model : NHM) を用いる場合の動作特性も評価した。なお、簡明な検討を行うため、シミュレーションの設定では、実際の試験とは一部異なる試験デザインを仮定した。本試験の実際のデザインは、Smith et al. (2006) を参照されたい。

以降では、データサイエンス部会から2024年4月に報告されている「アダプティブデザインを計画するためのシミュレーションの検討」で例示されているシミュレーション報告書の形式を参考にして、シミュレーションの計画及び結果を示す<sup>5</sup>。加えて、1回分のシミュレーションデータに基づいて、想定する試験の主要な解析結果を case example result として例示する。

### 4.1 緒言

#### 4.1.1 本試験のデザイン

試験標題	PHN 患者を対象とした新規治療の二重盲検プラセボ対照ランダム化比
------	-----------------------------------

	較第 II 相試験
目的	PHN に対する新規治療の用量反応関係の評価及び POC の確認
試験期間	6週間 (ベースライン 1週間、二重盲検 4週間、フォローアップ 1週間)
投与群	プラセボ群、試験薬群 7用量 (50, 100, 150, 200, 300, 450, 600 mg)
割付比	均等割付
主要評価項目	Pain score の週平均値 (average daily pain score : ADPS) のベースラインからの変化量 (投与4週後)

#### 4.1.2 解析方法

主要評価項目について、NDLM、BHM、NHMにより、プラセボ群に対する群間差を推定する。

##### 4.1.2.1 NDLM のモデリング

NDLM は時系列データの分析のために開発された統計的方法である<sup>2</sup>。Smith et al (2006) では、NDLM を用量反応関係の推定に用いることで単調性を仮定せずに柔軟なモデリングを行った<sup>1</sup>。ここでは、被験者のアウトカムについて NDLM によるモデル化を行うが、NDLM のより詳細については West et al. (1997)を参照されたい<sup>2</sup>。

用量 $j$  ( $j = 0$  (Placebo),  $1, \dots, J$ )における $k$  ( $k = 1, \dots, n_j$ )番目の被験者のアウトカムを $Y_{jk}$ とし、 $\theta_j$ を用量 $Z_j$ におけるアウトカムの平均とし、以下のようにモデル化する。

$$Y_{jk} = \theta_j + v_{jk}, \quad v_{jk} \sim N(0, \sigma^2)$$

ここで、 $\theta_j$  ( $j = 1, \dots, J$ )は以下のようにモデル化する。

$$\begin{aligned} \theta_j &= \theta_{j-1} + \delta_{j-1} + \omega_j & \omega_j &\sim N(0, W\sigma^2) \\ \delta_j &= \delta_{j-1} + \varepsilon_j & \varepsilon_j &\sim N(0, W\sigma^2) \end{aligned}$$

ただし、 $j = 0$  (Placebo)の場合は以下の通りとする。

$$\theta_0, \delta_0 \sim N(0, 1000)$$

事前分布 :

$$\begin{aligned} \sigma^2 &\sim \text{InvGamma}(0.001, 0.001) \\ W &\sim \text{uniform}(0.001, 100) \end{aligned}$$

理解の簡単化のために、上記の NDLM のパラメータ間の関係を図 4-1に示した。

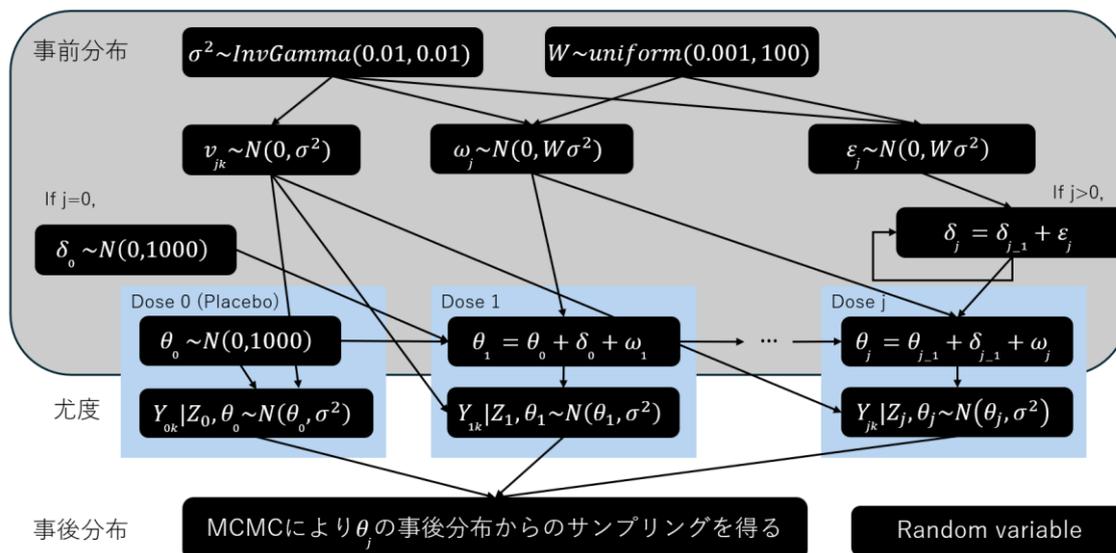


図 4-1 NDLM の概念図

#### 4.1.2.2 BHM のモデリング

BHM は、複数のレベルで記述された階層形式の統計モデルのことである。ここで用いる BHM では、実薬群の各用量の平均を表すパラメータが、特定の分布から生成されるという仮定を置く。用量  $j (j = 0 (Placebo), 1, \dots, J)$  の  $k (k = 1, \dots, n_j)$  番目の被験者のアウトカムを  $Y_{jk}$  としたとき、 $Y_{jk}$  は平均  $\theta_j$ 、分散  $\sigma^2$  の正規分布から生成され、更に実薬群の各用量の平均  $\theta_j (j = 1, \dots, J)$  が平均  $\mu$ 、分散  $\tau^2$  の正規分布から生成されるという BHM によってモデル化する。これを式で表すと  $Y_{jk} \sim N(\theta_j, \sigma^2), \theta_j \sim N(\mu, \tau^2) (j = 1, \dots, J)$  となる。また、プラセボ群の平均  $\theta_0$ 、アウトカムの分散  $\sigma^2$ 、ハイパーパラメータである  $\mu$  及び  $\tau^2$  の事前分布には無情報のフラット事前分布を用いる。また、理解の簡単化のために、上記の BHM のパラメータ間の関係を図 4-2 に示した。

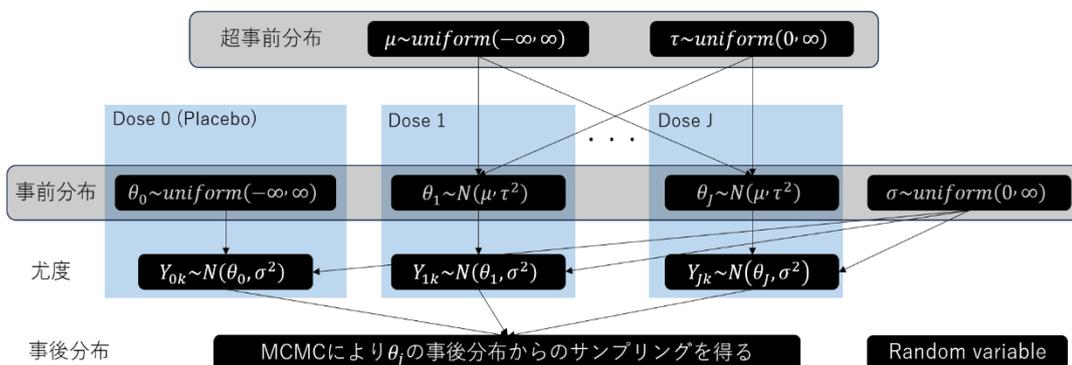


図 4-2 BHM の概念図

上記図 4-2 において  $Y_{jk}$  は用量群ごとのデータ、 $\theta_j$  は用量群ごとの平均を表すパラメータ、 $\sigma^2$  はデータの分散を表すパラメータ、 $\mu, \tau^2$  は  $\theta_j (j = 1, \dots, J)$  が従う超事前分布のハイパーパ

ラメータである。

#### 4.1.2.3 NHM のモデリング

NHM は、上記の BHM の実薬群の各用量の平均 $\theta_j (j = 1, \dots, J)$ の事前分布である $N(\mu, \tau^2)$ を、プラセボ群の平均 $\theta_0$ の事前分布と同じ無情報のフラット事前分布に変更したモデルである。

#### 4.1.2.4 効果判定

臨床的に意味のある有効性（プラセボ群に対して ADPS 変化量1.5の改善）が得られる事後確率に基づき、当該用量の効果有無を判断する。なお、下記 $\theta_j$ は NDLM 又は BHM 並びに NHM に基づく。

中間解析時点で以下の無効の基準に該当した用量 $j$ への新規被験者の組み入れを停止し、全用量が該当する場合は試験を中止する。

- 無効 :  $P(\theta_j - \theta_0 < 1.5 | Data) > 0.8$

最終解析時点で以下の有効の基準に該当した用量 $j$ について有効な用量と判定する。

- 有効 :  $P(\theta_j - \theta_0 > 1.5 | Data) > 0.8$

中間解析時点で無効と判定された用量に割り付けられた被験者のデータは、最終解析においてモデルの推定には用いられるが、当該用量は有効性の判定は行われぬ。

#### 4.1.3 シミュレーションの目的

試験では、幅広い用量の有効性と安全性を評価し、候補薬剤が十分な有効性を示すか判断する。シミュレーションでは、中間解析の時期と回数、arm-dropping のための無益性及び有効性の判断基準に関連した動作特性の評価を行う。

### 4.2 シミュレーション設定

#### 4.2.1 試験デザイン案

本シミュレーションでは、NDLM 及び BHM を用いて用量反応関係を推定する。NHM をベンチマークとして設定した。

#### 4.2.2 試験の設定

パラメータ	設定
最大被験者数	280例（各群35例） <設定根拠> プラセボ群との群間差を統計的に検出するための適切な検出力と用量反応曲線の十分な推定を可能にするための適切な検出力の間でバランスを

	とりながら選択した <sup>1</sup> 。
中間解析	回数：1回 タイミング：最終解析時に計画している被験者数のうち50%のアウトカムが利用可能な時点（情報量50%時）
	中間解析時に $P(\theta_j - \theta_0 < 1.5   Data) > 0.8$ を満たした場合、用量 $j$ への新規被験者の組み入れを停止し、全用量が該当する場合は試験を中止する。新規被験者の組み入れを停止した用量のデータも最終解析のパラメータ推定には用いるが、組み入れを停止した用量については最終解析で再度効果の有無は判定せず、効果なしと扱うこととした。なお、 $\theta_j$ は NDLM 又は BHM 並びに NHM に基づく。

### 4.2.3 データ発生及びシナリオ

Smith et al. (2006)を参考に、用量 $Z_j$ における $k$ 番目の被験者のアウトカムを $Y_{jk}$ とし、平均 $\theta_j$ 及び分散 $2.3^2$ の正規分布に従う疑似乱数を発生させる<sup>1</sup>。各シナリオの用量 $j$ の平均 $\theta_j$ は表4.1に示した。シナリオ1 (Null) は、実薬の効果が全くなく、プラセボと同等という仮定であり、シナリオ2a~2d、3a~3d、4a~4d は、最大の効果をそれぞれ、1.45、2.20、2.9とし、用量反応モデルとして、linear(a)、Emax(b)、betaMod(c)、logistic(d)を仮定している（図4-3）。なお、用量反応モデルのパラメータは、Bornkamp et al. (2009)を参考に設定した<sup>6</sup>。また、それぞれのシナリオにおいて中間解析あり、なしの場合を仮定したシナリオでデータを発生させた。

表 4.1 用量反応関係のシナリオ

Scenario	PLB	Dose1	Dose 2	Dose 3	Dose 4	Dose 5	Dose 6	Dose 7
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2a	0.00	0.12	0.24	0.36	0.48	0.73	1.09	1.45
2b	0.00	0.95	1.17	1.27	1.32	1.38	1.43	1.45
2c	0.00	0.92	1.21	1.36	1.43	1.42	1.12	0.57
2d	0.00	0.00	0.00	0.00	0.00	0.05	1.02	1.45
3a	0.00	0.18	0.37	0.55	0.73	1.10	1.65	2.20
3b	0.00	1.44	1.78	1.93	2.01	2.10	2.17	2.20
3c	0.00	1.40	1.83	2.07	2.18	2.15	1.69	0.87
3d	0.00	0.00	0.00	0.00	0.00	0.08	1.54	2.20
4a	0.00	0.24	0.48	0.73	0.97	1.45	2.18	2.90
4b	0.00	1.90	2.34	2.54	2.65	2.77	2.85	2.90
4c	0.00	1.85	2.42	2.72	2.87	2.84	2.23	1.15
4d	0.00	0.00	0.00	0.00	0.01	0.10	2.04	2.90

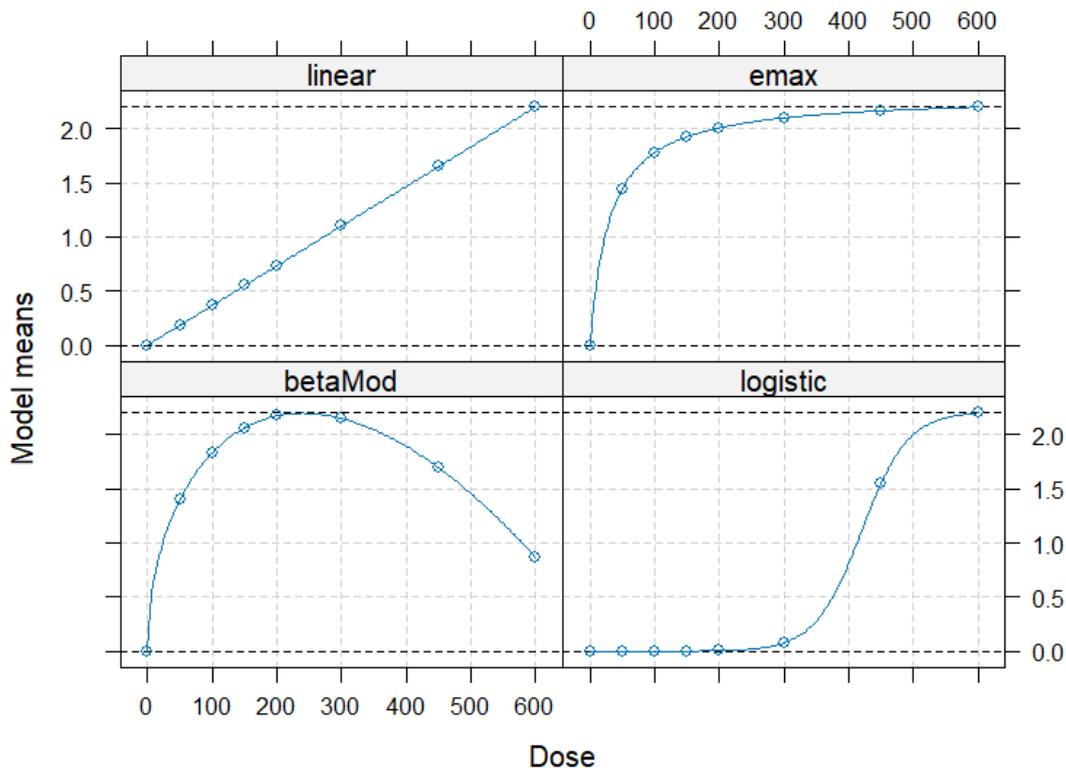


図 4-3 シナリオ3a~3d の用反応曲線

#### 4.2.4 評価指標

動作特性評価のため、下記の指標について検討した。

指標	内容
検出力	少なくとも1用量で効果ありと判定される確率を検出力とする。 全用量においてプラセボ群との差がない仮定（帰無仮説、シナリオ1）のもと少なくとも1つの用量で臨床的に意味のある有効性（プラセボ群との群間差1.5）を示すことを第一種の過誤とする。
被験者数	試験全体及び各用量群の平均被験者数を確認する。
最大効果を持つ用量を正しく特定できる確率	最大効果を持つ用量である確率 $P(\theta_j - \theta_0 = \max_j(\theta_j - \theta_0)   Data)(j = 1, \dots, J)$ に基づいて、真の最大効果を持つ用量を正しく特定できる確率を確認する
治療効果の推定値、バイアス、RMSE	治療効果の推定値（プラセボ群との群間差）のバイアス及び root mean squared error (RMSE) を確認する。

#### 4.2.5 シミュレーションの繰り返し数と乱数シード

各シナリオについて、繰り返し数は1,000回とした。

## 4.2.6 Stan コードのポイント

シミュレーションを実行するための R コードのうち、NDLM 及び BHM を実装する Stan コードのポイントを説明する。シミュレーションを実行するための R コード全体は、GitHub に掲載する。

### 4.2.6.1 NDLM

試験の目的における解析方法に記載した NDLM を実装するコード (NDLM\_simulation\_with\_IA.r) の主となる部分を示す。以下は Stan を用いて NDLM を markov chain monte carlo methods (MCMC) で推定する部分である。

```
stanmodelcode <- "
data {
  int<lower=0> N;
  int<lower=0> K;
  real y[N];
  int<lower=0> arm[N];
}
```

N を被験者アウトカムの観測数、K を投与群数、y を被験者アウトカム、arm を治療群として宣言する。

```
parameters {
  real theta[K];
  real delta[K];
  real<lower=0> sigma2;
  real<lower=0> W;
}
```

ここでは、NDLM のパラメータを宣言する。試験の目的における解析方法に記載した notation で表すと  $\theta_j = \theta_j$ 、 $\delta_j = \delta_j$ 、 $\sigma^2 = \sigma^2$ 、 $W = W$  である。ただし、 $\theta[1]$  と  $\delta[1]$  はプラセボ群の  $\theta_0$  と  $\delta_0$  を表すパラメータとして宣言している。

```
model {
  for(k in 2:K){
    theta[k] ~ normal(theta[k-1]+delta[k-1], sqrt(W*sigma2));
    delta[k] ~ normal(delta[k-1], sqrt(W*sigma2));
  }
}
```

NDLM のシステム方程式を指定している。

$$\text{NDLM のシステム方程式}(j>0) \quad \theta_j = \theta_{j-1} + \delta_{j-1} + \omega_j \quad \omega_j \sim N(0, W\sigma^2)$$

$$\delta_j = \delta_{j-1} + \varepsilon_j \quad \varepsilon_j \sim N(0, W\sigma^2)$$

```
theta[1] ~ normal(0, 1000);
delta[1] ~ normal(0, 1000);
```

$\theta[1]$  と  $\delta[1]$  はプラセボ群の平均を表すパラメータとして宣言しているため、以下

を表している。

$$\theta_0, \delta_0 \sim N(0, 1000)$$

```
for(j in 1:N){
  y[j] ~ normal(theta[arm[j]], sqrt(sigma2));
}
```

以下の NDLM の観測方程式をモデル化している。

$$Y_{jk} = \theta_j + v_{jk}, \quad v_{jk} \sim N(0, \sigma^2)$$

```
W~uniform(0.001,100);
sigma2~inv_gamma(0.001, 0.001);
}
"
```

事前分布の指定を行っている。W には一様分布、 $\sigma^2$  にはパラメータ 0.001、0.001 の逆ガンマ分布を設定している。

#### 4.2.6.2 BHM

試験の目的における解析方法に記載した BHM を実装するコード (BHM\_simulation.r) の主となる部分を示す。以下は Stan を用いて BHM を MCMC で推定する部分である。

```
stanmodelcode <- "
data {
  int<lower=0> N[8];
  vector[N[1]] y0;
  vector[N[2]] y1;
  vector[N[3]] y2;
  vector[N[4]] y3;
  vector[N[5]] y4;
  vector[N[6]] y5;
  vector[N[7]] y6;
  vector[N[8]] y7;
}
```

解析時の各群の被験者数を配列 N として宣言し、各群の被験者のアウトカムデータを y0~y7 のベクトルとして宣言している。シミュレーションでは、中間解析及び最終解析の対象となる被験者を配列 N の値で制御している。

```
parameters {
  real theta0;
  real mu;
  vector[7] theta;
  real<lower=0> sigma;
  real<lower=0> tau;
}
```

theta0はプラセボ群の平均を表すパラメータとして宣言している。mu と tau は用量群ごとの平均と標準偏差を表すハイパーパラメータとして宣言している。sigma は各群のデータの標準偏差を表すパラメータとして宣言している。theta は、用量群ごとの平均を表すパラメータ（ベクトル）として宣言している。

```

model {
  theta ~ normal(mu, tau);
  y0 ~ normal(theta0, sigma);
  y1 ~ normal(theta[1], sigma);
  y2 ~ normal(theta[2], sigma);
  y3 ~ normal(theta[3], sigma);
  y4 ~ normal(theta[4], sigma);
  y5 ~ normal(theta[5], sigma);
  y6 ~ normal(theta[6], sigma);
  y7 ~ normal(theta[7], sigma);
}
"
    
```

試験の目的の解析方法に記載されている BHM の式に対応する。theta ~ normal(mu, tau) の部分で、用量群ごとの平均を表すパラメータ theta に共通の階層事前分布を指定している。

### 4.3 シミュレーション結果

シミュレーション結果のうち、シナリオ1、3a~3d を抜粋して示す。すべてのシナリオのシミュレーション結果は Appendix (5.7.1) に示す。

#### 4.3.1 検出力、被験者数、最大効果を持つ用量を正しく特定できる確率

シナリオ1、3a~3d において、各モデルの検出力、平均被験者数、最大効果を持つ用量を正しく特定できる確率を表 4.2に示す。

表 4.2 各モデルの検出力、平均被験者数、最大効果を持つ用量を正しく特定できる確率

中間解析	シナリオ	検出力（少なくとも1用量で効果ありと判定される確率）			平均被験者数			最大効果を持つ用量を正しく特定できる確率		
		NDLM	BHM	NHM	NDLM	BHM	NHM	NDLM	BHM	NHM
なし	1	0.00	0.00	0.00	280	280	280	NA	NA	NA
	3a	0.70	0.50	0.71	280	280	280	0.96	0.82	0.82
	3b	0.79	0.77	0.89	280	280	280	0.45	0.27	0.27
	3c	0.70	0.71	0.88	280	280	280	0.40	0.32	0.32
	3d	0.68	0.51	0.68	280	280	280	0.96	0.89	0.89
あり	1	0.00	0.00	0.00	151.0	152.1	166.3	NA	NA	NA
	3a	0.74	0.50	0.69	215.2	220.9	222.6	0.95	0.80	0.81

	3b	0.79	0.76	0.90	261.3	272.9	269.7	0.49	0.28	0.28
	3c	0.73	0.71	0.86	253.4	265.3	261.1	0.41	0.35	0.35
	3d	0.68	0.52	0.67	194.7	197.4	202.0	0.94	0.87	0.87

#### 4.3.2 治療効果の推定値、バイアス、RMSE

シナリオ1、3a～3d（中間解析あり）において、各モデルから推定された用量反応曲線を図 4-4に示す。同様に、バイアスと RMSE を図 4-5に示す。

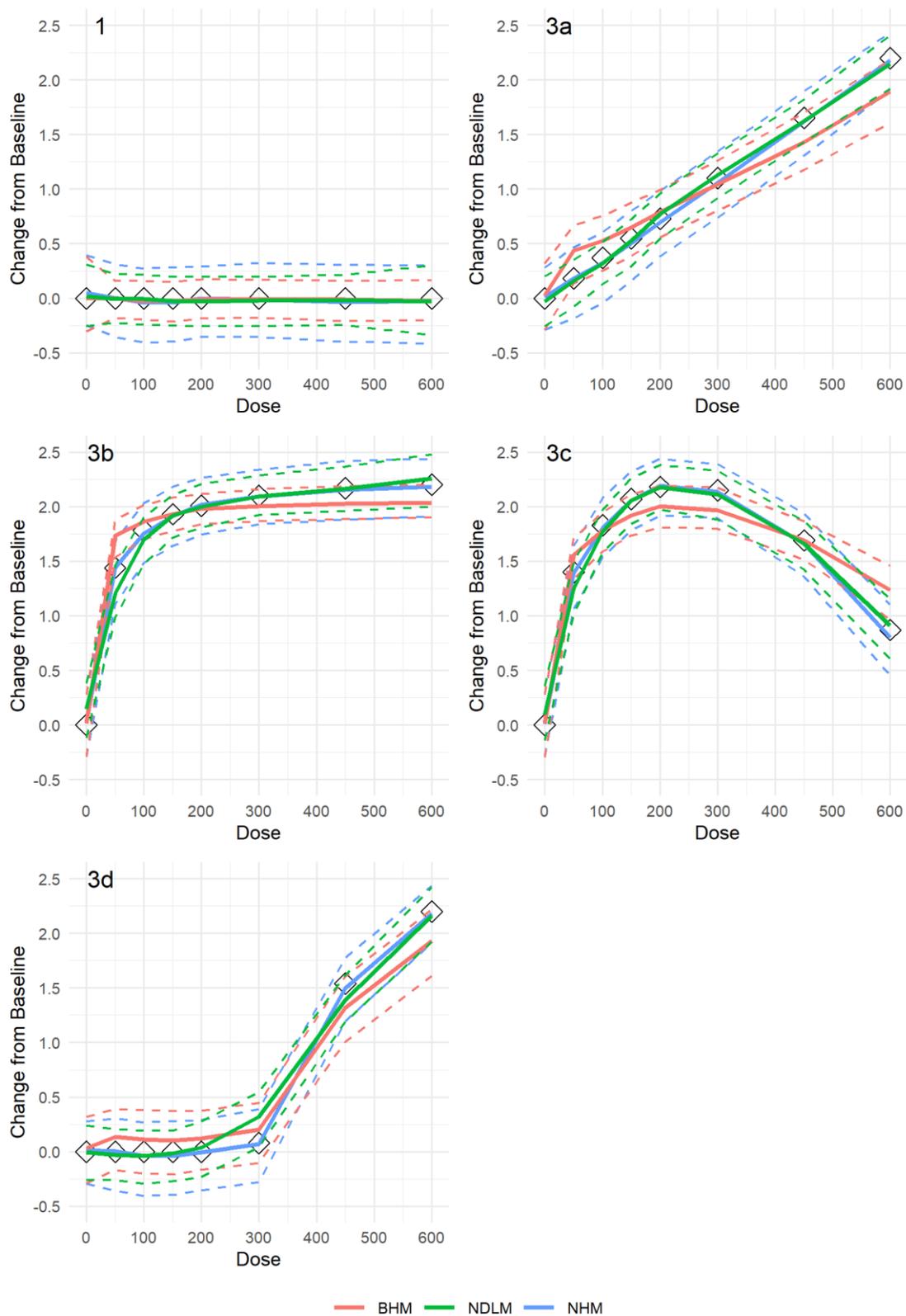


図 4-4-4 シナリオ1、3a~3dの推定された用量反応曲線 (◇: 真値、実線: 事後平均の中央値、破線: 事後平均の25%点及び75%点)

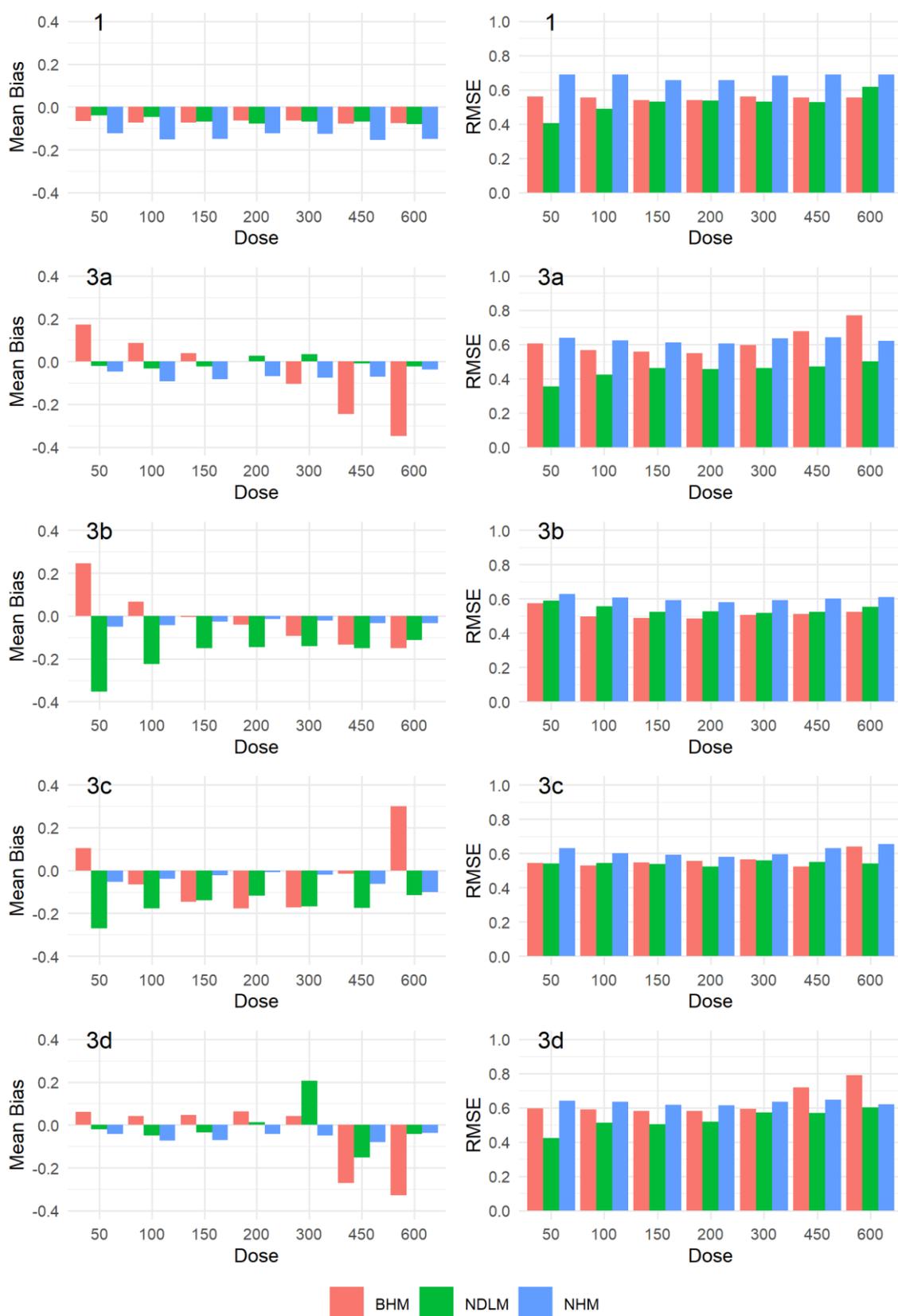


図 4-5 シナリオ1、3a~3dのバイアスと RMSE

## 4.4 要約及び推奨事項

### 4.4.1 考察

シナリオ1、3a~3d（中間解析あり）以外も含むすべてのシミュレーション結果から考察を行う。

#### 4.4.1.1 検出力

本設定では、第一種の過誤確率（シナリオ1における検出力）は、いずれのモデルでも小さい値であり、モデル間の違いは見られなかった。BHMの検出力はNHMより低く、NDLMはNHMと同程度か低い場合があった。

#### 4.4.1.2 平均被験者数

有効性がない又は低いシナリオ（シナリオ1、2a~2d）では、BHMはNHMよりも期待例数が小さいが、有効性のあるシナリオでは違いはなかった。NDLMは他の2つのモデルと比較して、いずれのシナリオでも期待例数が一様に小さかった。

#### 4.4.1.3 最大効果を持つ用量を正しく特定できる確率

最大効果を持つ用量を正しく特定できる確率は、BHMとNHMで大きな違いはなかった。NDLMは他の2つのモデルと比較して、最大効果の用量を選択する確率が高かった。

#### 4.4.1.4 バイアス

無情報事前分布を設定していたため、NHMではバイアスが生じなかった。BHMでは、用量反応がある場合（シナリオ1以外）に縮小により低用量では正のバイアス、高用量では負のバイアスが生じる傾向があった。NDLMでは、プラセボを含め1つ前の用量に推定値が影響を受けるため、一概ではないが、負のバイアスが生じる傾向があった。NDLMとBHMの比較では、相対的にNDLMのバイアスが小さい傾向があった。

#### 4.4.1.5 RMSE

BHM（高用量を除いて）及びNDLMでは、NHMと比べて全体的にRMSEが小さかった。BHMよりもNDLMの方がよりRMSEが小さい傾向があった。

#### 4.4.1.6 推奨される試験デザイン

本試験は用量反応曲線の推定とPOCの確認を行い、次相で有効性を検証する用量を選択することを目的としていることから、各評価指標の結果から総合的に判断すると、解析モデルとしてNDLMの使用が推奨されると考えられる。また、本試験のシミュレーション結果では中間解析の有無で検出力は大きく異ならず、期待例数は中間解析ありの方が小さいことから、中間解析を実施してarm-droppingを行うことが推奨される。

## 4.5 Case Example Result

シナリオ3aのシミュレーションから、想定している試験のNDLMによる解析結果の例を示す。

Case example resultにおける中間解析結果及び最終解析結果をそれぞれ表4.3及び表4.4に示した。

中間解析結果では、dose 1、dose 2及びdose 3は組み入れ停止基準を満たしたため、これらの用量への組み入れを停止して試験を継続することとなる。

最終解析結果では、dose 6及びdose 7が効果ありの基準を満たし、新規治療のPOCが確認されたと解釈できる。ベースラインからの変化量の平均の推定値は、用量が増加するとともに単調に増加し、最大効果を持つ用量である確率はdose 7が最も大きく、dose 7が最も有効な用量であると解釈できる。

表 4.3 Case Example Resultにおける中間解析結果

Group	N	Change from Baseline <sup>a</sup>	Difference (95%CI) <sup>b</sup>	P (Futility)
PLB	17	-0.469	-	-
Dose 1	11	-0.125	0.344 (-0.777, 1.517)	0.973
Dose 2	16	0.117	0.586 (-0.671, 1.875)	0.924
Dose 3	18	0.268	0.737 (-0.596, 2.022)	0.873
Dose 4	20	0.519	0.989 (-0.349, 2.253)	0.774
Dose 5	20	1.063	1.532 (0.219, 2.852)	0.489
Dose 6	19	1.413	1.882 (0.526, 3.186)	0.268
Dose 7	19	2.372	2.842 (1.381, 4.260)	0.033

a: Posterior Mean b: Posterior Mean (95% Credible Interval)

表 4.4 Case Example Resultにおける最終解析結果

Group	N	Change from Baseline <sup>a</sup>	Difference (95%CI) <sup>b</sup>	P (Efficacy)	P (Max)
PLB	35	-0.342	-	-	-
Dose 1	11	-0.076	0.266 (-0.611, 1.132)	0.006	0.000
Dose 2	16	0.164	0.507 (-0.491, 1.446)	0.020	0.000
Dose 3	18	0.390	0.732 (-0.330, 1.663)	0.069	0.000
Dose 4	35	0.800	1.142 (0.183, 2.048)	0.220	0.000
Dose 5	35	1.347	1.689 (0.797, 2.572)	0.657	0.019
Dose 6	35	1.602	1.944 (0.956, 2.914)	0.832	0.041
Dose 7	35	2.201	2.544 (1.547, 3.579)	0.980	0.940

a: Posterior Mean b: Posterior Mean (95% Credible Interval)

## 4.6 参考文献

1. Smith, M. K., Jones, I., Morris, M. F., Grieve, A. P., & Tan, K. (2006). Implementation of a Bayesian adaptive design in a proof of concept study. *Pharmaceutical statistics*, 5(1), 39–50.
2. West, M., & Harrison, J. (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.
3. 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. (2024). *Complex Innovative Trial Design の現状と事例の整理*. [https://www.jpma.or.jp/information/evaluation/results/allotment/DS\\_202404\\_CID.html](https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202404_CID.html)

4. Garces, S., Karis, E., Merrill, J. T., Askanase, A. D., Kalunian, K., Mo, M., & Milmont, C. E. (2023). Improving resource utilisation in SLE drug development through innovative trial design. *Lupus science & medicine*, 10(2), e000890.
5. 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. (2024). アダプティブデザインを計画するためのシミュレーションの検討.  
[https://www.jpma.or.jp/information/evaluation/results/allotment/DS\\_202404\\_AD.html](https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202404_AD.html)
6. Bornkamp, B., Pinheiro, J., & Bretz, F. (2009). MCPMod: An R Package for the Design and Analysis of Dose-Finding Studies. *Journal of Statistical Software*, 29(7), 1–23.

#### 4.7 Appendix

すべてのシナリオのシミュレーション結果を示す。



図 4-6 すべてのシナリオのバイアス

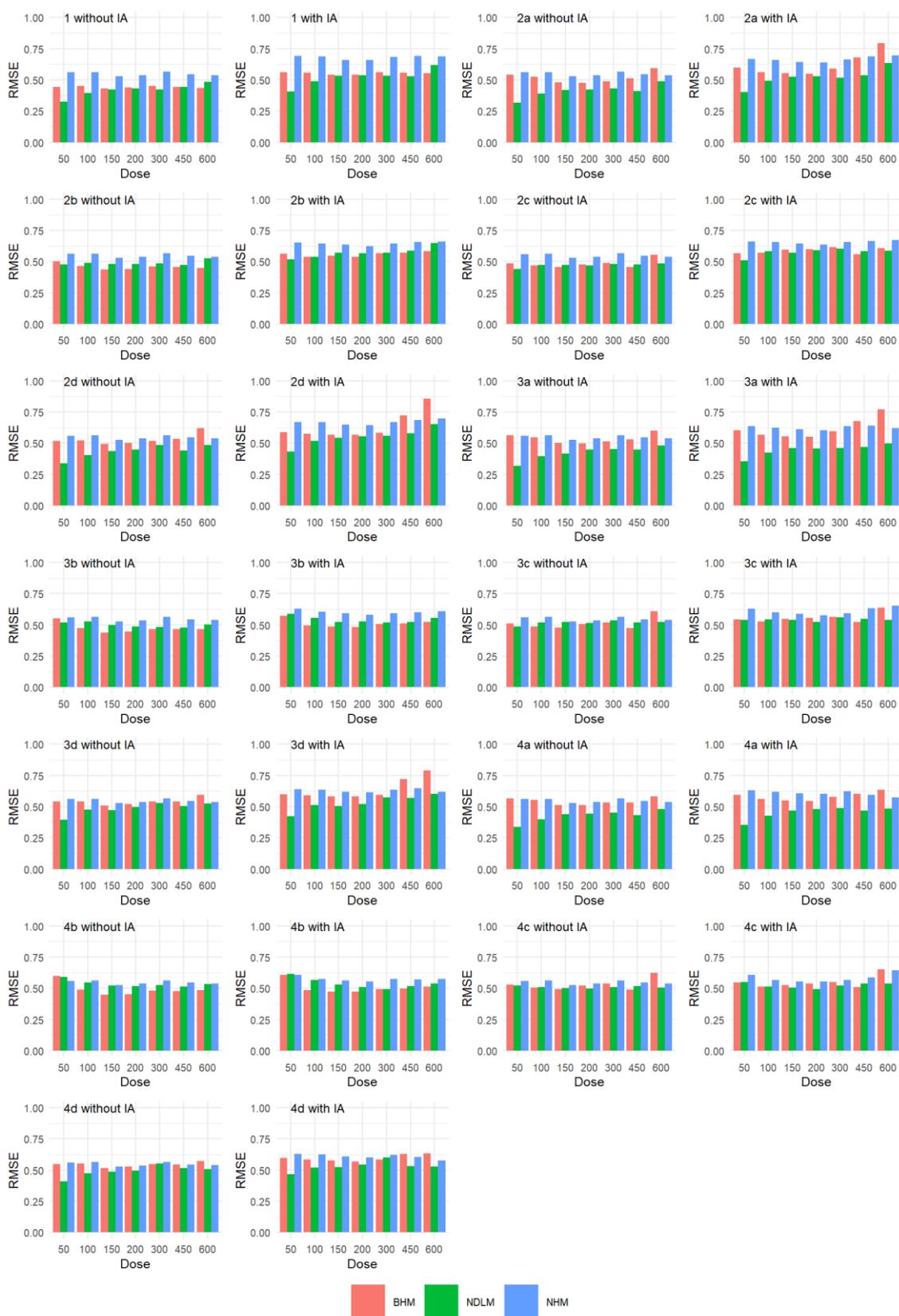


図 4-7 すべてのシナリオの RMSE

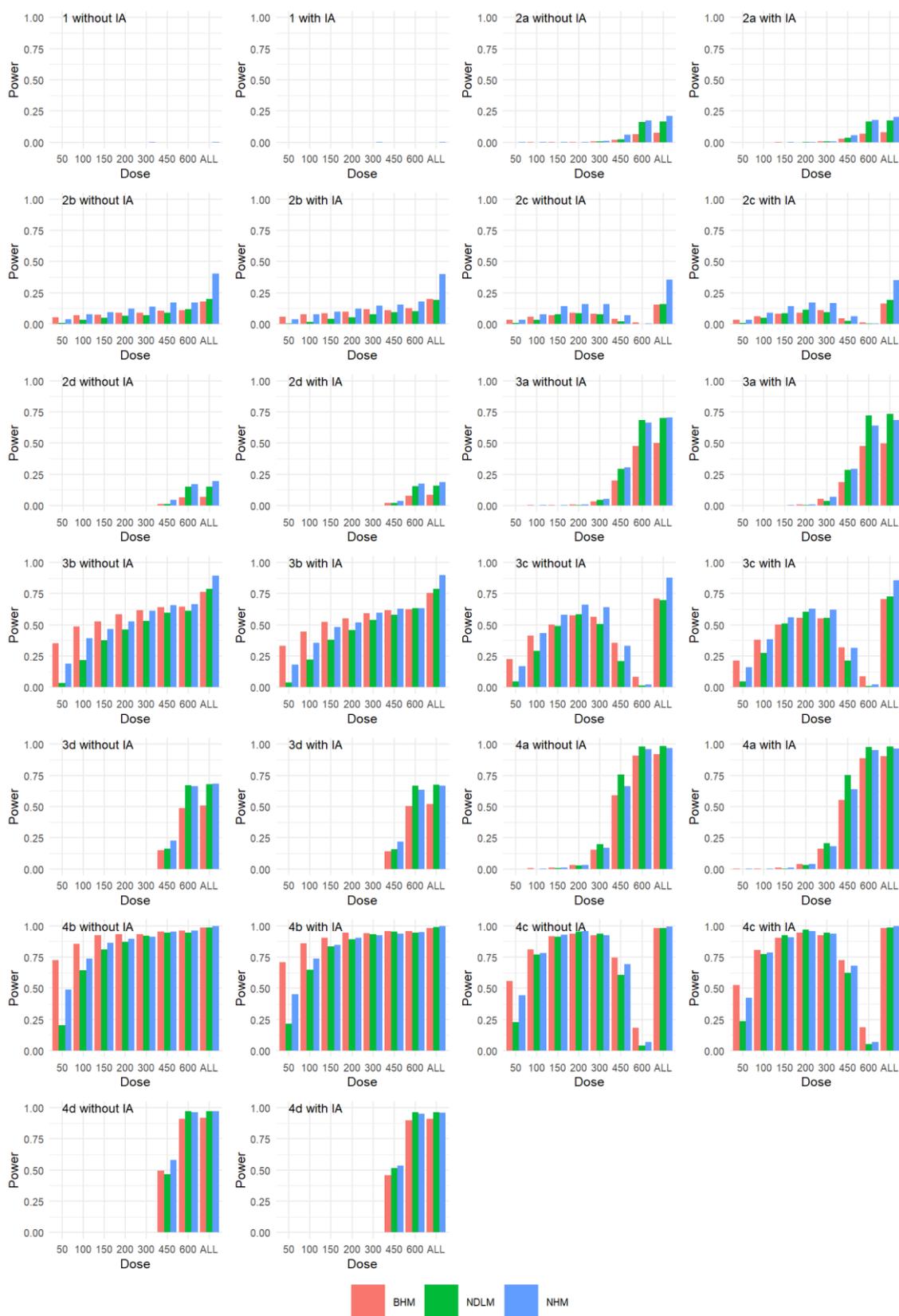


図 4-8 すべてのシナリオの検出力 (ALL は少なくとも1用量で効果ありと判定される確率)



図 4-9 すべてのシナリオの平均被験者数 (ALL は試験全体)

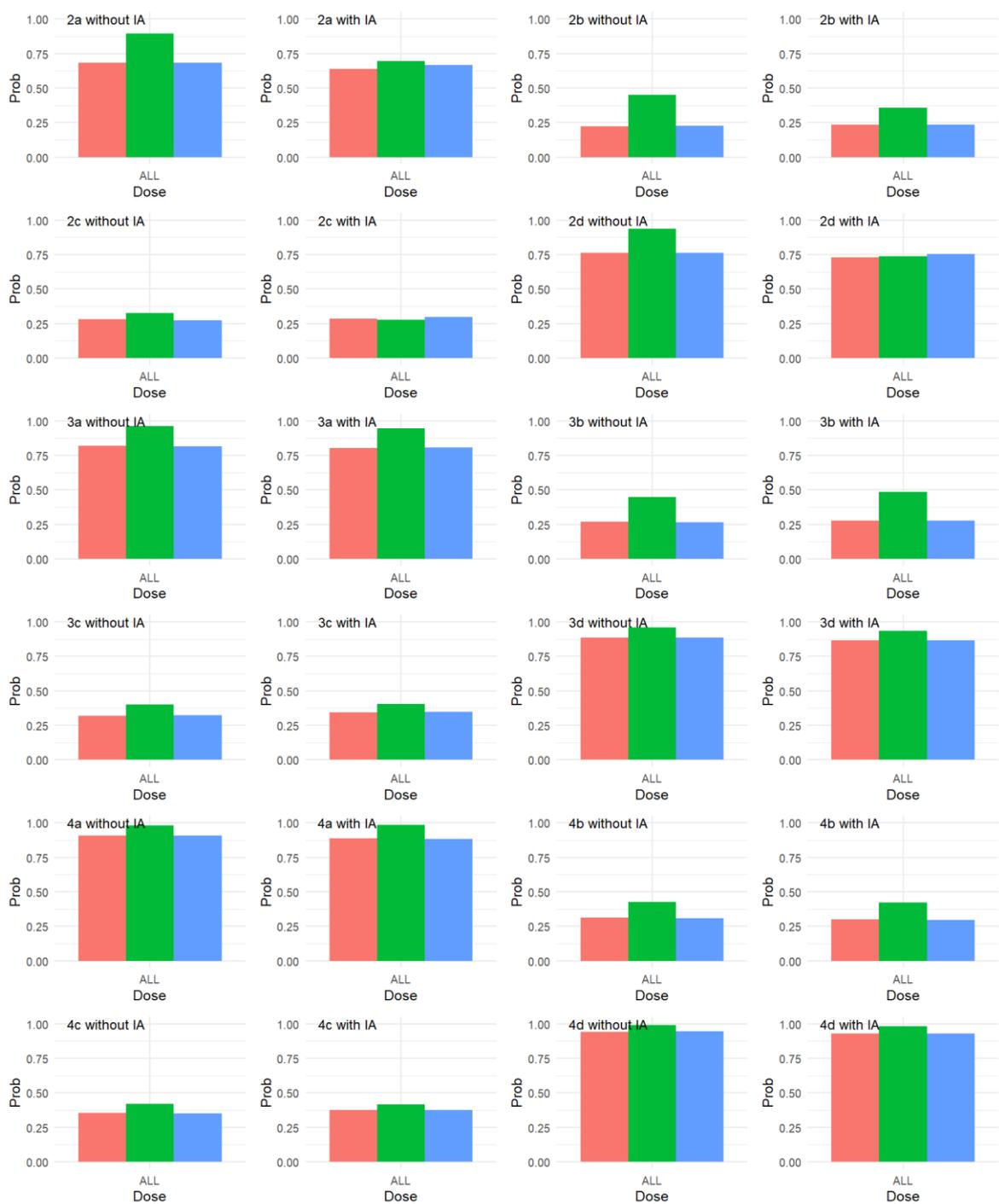


図 4-10 すべてのシナリオの最大効果を持つ用量を正しく特定できる確率

## 5. バスケット試験

Pant et al. (2023)<sup>1</sup>では、FGFR 異常のある複数の固形がんに対する Erdafitinib の有効性と安全性について評価した。本試験では、各がん種で組み入れられる被験者数が限られるため、中間解析及び最終解析時の各がん種での有効性評価に通常の頻度論流の手法を用いた場合、その判断が難しくなる可能性がある。一方で共通の遺伝子異常である FGFR 異常を持っていることがいずれのがん種でも試験参加の選択基準となっていることから、各がん種における奏効率が類似している可能性を考慮し、ベイズ流階層モデルを使用し、中間解析及び最終解析で各がん種における無益性及び有効性を評価した。各がん種で組み入れられる被験者数が限られるものの、ベイズ流階層モデルを用いることでがん種間で情報借用を行い、より精度の高い推定を行うことがベイズ流の手法を用いる理由である。

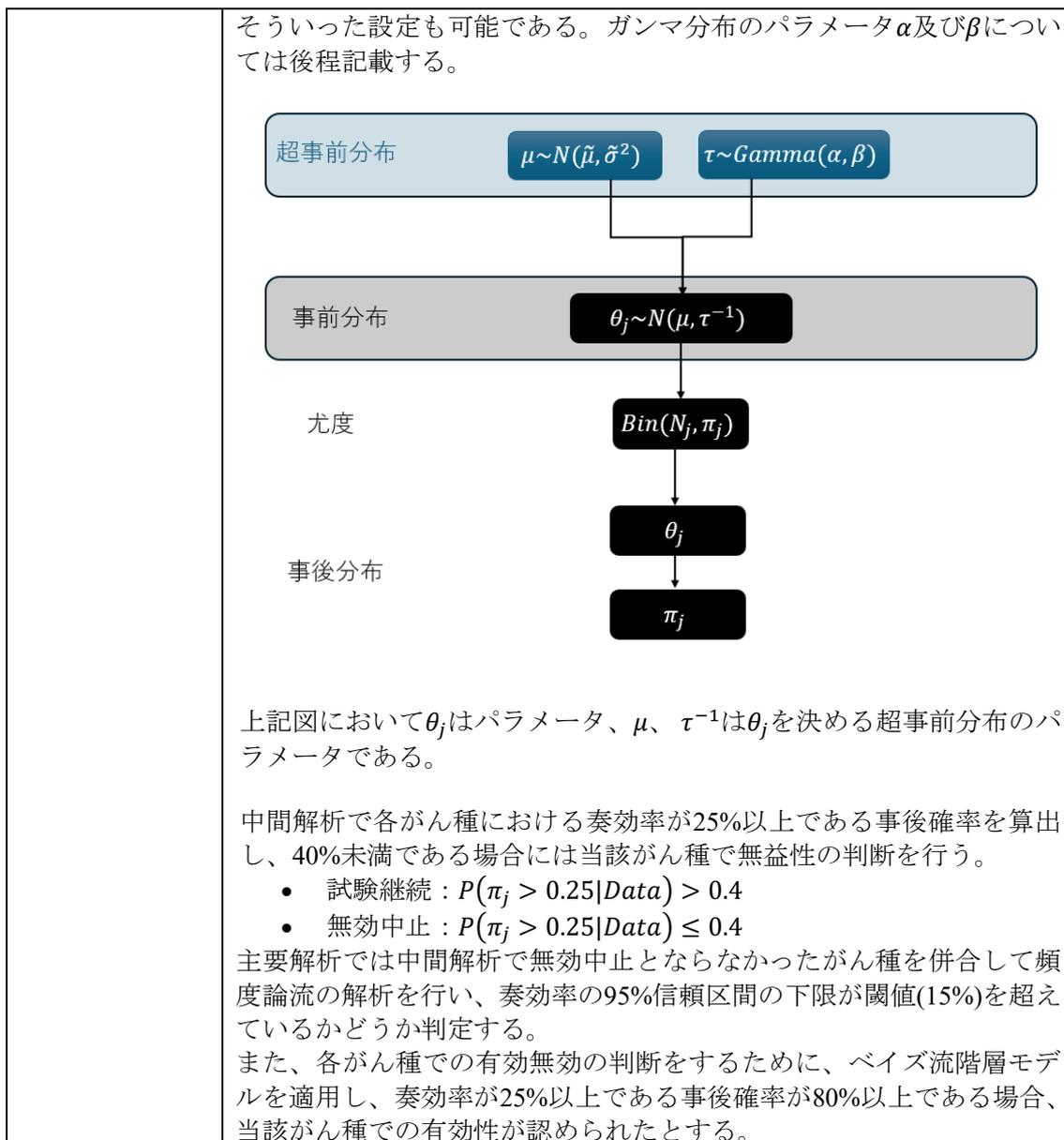
本タスクフォースは、Pant et al. (2023)<sup>1</sup>を計画する状況を想定し、シミュレーションによりベイズ流階層モデルに基づく中間解析の動作特性を評価した。一部、シミュレーションを簡便に実行するために、実際とは異なる試験デザインを仮定している※。本試験の実際のデザインは、Pant et al. (2023)<sup>1</sup>を参照されたい。

### ※実際の試験と異なる仮定

- がん種の種類を 15 種類ではなく 5 種類と設定した。
- 被験者数を 200 例ではなく 100 例とした。
- 中間解析を 3 回ではなく 1 回で設定した。
- 各がん種における組み入れ被験者数の上限及び中間解析での無益性判定に必要な最小被験者数を設定しない。
- R でシミュレーションを行った（原著は FACT version 6.2）。

### 5.1 本試験の目的

試験標題	特定のバイオマーカーで選択された進行がん患者を対象とした新規治療の単群第 II 相試験
目的	複数がん種の進行がんに対する新規治療の有効性の評価
試験期間	治療期（疾患進行がみられるまで）及びフォローアップ期
投与群	新規治療群のみ（単群試験）
割付比	該当なし（単群試験）
主要評価項目	奏効率
解析方法	<p>主要評価項目について、ベイズ流階層モデルにより、各がん種での奏効率を推定する。</p> <p>ベイズ流階層モデルにより各がん種における被験者数が限られている中でも、奏効率の交換可能性を仮定して解析をすることができる。がん種 <math>j</math> における奏効率を <math>\pi_j</math> とし、<math>\theta_j = \text{logit}(\pi_j)</math> とした際に、<math>\theta_j</math> は、平均 <math>\mu</math>、分散 <math>\tau^{-1}</math> の正規分布に従っているとし、超事前分布として <math>\mu \sim N(-2.19, 2^2)</math> 及び <math>\tau \sim \text{Gamma}(\alpha, \beta)</math> を仮定した。<math>\mu</math> の事前平均 -2.19 は奏効率 10% の <math>\text{logit}</math> 変換後の値に対応し、<math>\mu</math> の事前分布の範囲は奏効率 0% から 99.7% に相当する。Thall et al. (2003) は超事前分布の超パラメータに閾値奏効率と期待奏効率の中間値を用いることを提案しており、</p>



## 5.2 シミュレーションの目的

本試験では、幅広いがん種の有効性と安全性を評価し、候補薬剤が十分な有効性を示すか判断する。無益性及び有効性の判断基準に関する動作特性を理解するためにシミュレーションを実施する。

## 5.3 シミュレーション入力設定

### 5.3.1 試験デザイン案

本シミュレーションでは、ベイズ流階層モデルを用いて各がん種における奏効率を推定する。通常の臨床試験で使用される頻度論流における第一種の過誤確率0.05をベンチマークとして設定した。

### 5.3.2 中間解析及び判定基準

パラメータ	設定
最大被験者数	すべてのがん種を合わせて100例 <設定根拠> 頻度論流の解析計画を立てた際に、有意水準片側0.025、新規薬剤での奏効率を35%とした際に、単群試験で閾値15%に対する優越性を99.7%の確率で検出できる。一般的な臨床試験の検出力より高い値であるが、今回は各がん種での活性の有無も検討するため、100例と設定した。
中間解析の回数及び目的	中間解析の回数: 1回 中間解析の目的: 無効中止の判定 中間解析の時期: 50例の患者で奏効データが確定した時点
無効中止の基準	$P(\pi_j > 0.25   Data) \leq 0.4$ ならがん種 $j (j = 1, \dots, 5)$ を無効中止する。

### 5.3.3 仮想データの生成モデル

主要評価項目に二項分布を適用した。欠測値や脱落例はノンレスポnderとして、データを生成した。

### 5.3.4 各がん種での発現割合及び治療効果の仮定

シミュレーションにおける各がん種における登録例数、奏効率のシナリオを表 5.1のように設定した。各がん種での登録割合について、どのがん種も同程度の場合（シナリオ1-1~3）、特定のがん種に偏っている場合（シナリオ2-1~3）、いくつかの希少ながん種がある場合（シナリオ3-1~3）を仮定した。また奏効率については各がん種で活性がある場合、一部のがん種でのみ活性がある場合（シナリオ2-2では最も登録例数が多いがん種でのみ活性がある場合）、いずれのがん種でも活性がない場合を検討した。活性がないがん種の奏効率に関しては灰色で記載した。

表 5.1 各シナリオでの各がん種における登録例数及び奏効率

シナリオ	がん1		がん2		がん3		がん4		がん5	
	登録例数	奏効率								
1-1	20	0.35	20	0.35	20	0.35	20	0.35	20	0.35
1-2		0.35		0.35		0.35		0.15		0.15
1-3		0.15		0.15		0.15		0.15		0.15
2-1	60	0.35	10	0.35	10	0.35	10	0.35	10	0.35
2-2		0.35		0.15		0.15		0.15		0.15
2-3		0.15		0.15		0.15		0.15		0.15
3-1	30	0.35	30	0.35	30	0.35	5	0.35	5	0.35
3-2		0.35		0.35		0.35		0.15		0.15
3-3		0.15		0.15		0.15		0.15		0.15

### 5.3.5 事前分布の設定

事前分布が動作特性に及ぼす影響を評価するために以下の二つの超事前分布で比較し

た。Prior 2の方ががん種間の奏効率に関する情報借用の程度が大きくなる。

Prior 1:  $\tau \sim \text{Gamma}(2, 20)$

Prior 2:  $\tau \sim \text{Gamma}(2, 2)$

### 5.3.6 評価指標

動作特性評価のため、下記の指標について検討した。

指標	内容
Selection Power	活性があるがん種のうち80%以上が最終解析に含まれる確率
Proportion of Success	ベイズ流階層モデルによって一つ以上のがん種で成功と判定される確率 (成功判定基準: $\Pr(\pi_j > 25\%) > 80\%$ )
Family-wise Error Rate (FWER)	活性がないがん種で一つ以上偽陽性により有効性を主張してしまう確率。Proportion of Successと同様のベイズ流階層モデル判定基準を用いる
Power	主要解析の併合解析で有効と判定される確率
Mixing Nonactive	一つ以上活性がないがん種が含まれる中で、主要解析の併合解析で有効と判定される確率

### 5.3.7 シミュレーションの繰り返し数と乱数シード

各シナリオについて、繰り返し数は100,000回とした。これにより、第一種の過誤確率の95%信頼区間の幅が約±0.1%になることが保証される。シミュレーションのシナリオごとに異なる乱数シードを用いた。

### 5.3.8 コードのポイント

本シミュレーションはbasket\_trial.Rのプログラムを用いて実施しており、以下のRのパッケージを用いた。trialrパッケージに含まれるstan\_hierarchical\_response\_thall関数を用いることで二項分布の下でのベイズ流階層モデルの適用が可能となる。

```
library(mvtnorm)
library(dplyr)
library(rstan)
library(trialr)
library(DoseFinding)
```

```
#Prior 1の場合
fith <- stan_hierarchical_response_thall(
  group_responses = c(10, 9, 12, 1, 0), #各がん種での奏効例数
  group_sizes = c(30, 30, 30, 5, 5), #各がん種での被験者数
  mu_mean = -2.19, #μの超事前分布の平均
  mu_sd = sqrt(1 / 4), #μの超事前分布の標準偏差
  tau_alpha = 2, #τの超事前分布のパラメータ
  tau_beta = 20)
```

具体的にベイズ流階層モデルを適用した場合の結果を示す。事後平均1は Prior 1を、事後平均2は Prior 2を用いた場合の結果である。がん種間の奏効率に関する情報借用の程度

がより大きくなる Prior 2を用いた場合の方が、他のがん種の奏効率により引っ張られていること、95%信用区間の区間幅が狭く、各がん種間での奏効率の情報を借用し、より効率の良い推定が行うことができていることがわかる。

	被験者数	奏効数	奏効率	事後平均1 (95%信用区間)	事後平均2 (95%信用区間)
がん種1	30	10	33.3%	33% (18%, 50%)	31% (17%, 47%)
がん種2	30	9	30%	29% (15%, 47%)	28% (15%, 44%)
がん種3	30	12	40%	39% (23%, 57%)	36% (22%, 53%)
がん種4	5	1	20%	19% (1%, 56%)	19% (3%, 49%)
がん種5	5	0	0%	6% (0%, 27%)	12% (1%, 34%)

## 5.4 シミュレーション結果

### 5.4.1 シミュレーション結果の概要

シミュレーション結果を表 5.2に示す。

#### がん種間の登録割合が同じ状況 (シナリオ1-1~1-3)

いずれのがん種でも活性がある状況では (シナリオ1-1)、Prior の設定によらず、いずれの指標でも性能はほぼ同じであった。

一部のがん種で活性がない状況 (シナリオ1-2) では、より情報借用する Prior 2で Proportion of Success が低くなった。これは、シナリオ1-1と比較して奏効率が低いがん種が多くなり、ベイズ流階層モデルを用いることで活性があるがん種でも奏効率が低く推定されるためだと考えられた。

すべてのがん種で活性がない場合では(1-3)、Power が0.05を超えており、erroneous conclusion rate が適切に制御できていなかった。これは中間解析で活性がないにも関わらず、偶然奏効率が高くなったがん種のみを残して試験が続行され、最終解析に各がん種で併合した解析を行うことで、治療効果が過大推定されるためだと考えられる。一方、最終解析でベイズ流階層モデルを用いた評価を行った Proportion of Success は0.05以下となっていた。

#### がん種間の登録割合が異なる状況 (シナリオ2-1~2-3)

いずれのがん種でも活性がある状況では (シナリオ2-1)、Prior の設定によらず、いずれの指標でも性能はほぼ同じであった。しかし Selection Power はシナリオ1-1よりも低くなった。これはがん種2-5での被験者数が少ないため、偶然いずれかのがん種で治療効果が低くなり、無効中止されてしまったものがあったためと考えられる。

一部のがん種で活性がない状況 (シナリオ2-2) では、活性があるがん種は、最も発現割合が高いがん種1のみのシナリオなので、最終解析に含められる確率は高くなり、シナリオ1と比較して Selection Power が高くなった。その一方で、がん種間の情報借用をすることで、活性がないがん種も含めて活性があると判断してしまう、FWER 及び mixing nonactive が高くなった。

すべてのがん種で活性がない状況（シナリオ2-3）では、Prior 2を使い、よりがん種間の情報借用をすることで、偶然奏効率が高かったがん種があった場合でも、他の奏効率が低いがん種の結果に引っ張られてしまうため、Prior 1を用いた場合と比較して Proportion of Success 及び FWER が低くなった。

**一部のがん種の登録割合が非常に低い状況（シナリオ3-1~3-3）**

いずれのがん種でも活性がある状況では（シナリオ3-1）、Prior の設定によらず、いずれの指標でも性能はほぼ同じであった。

希少ながん種で活性がない状況（シナリオ3-2）では、Prior 2を使い、よりがん種間の情報借用をすることで、活性がないがん種の結果に引っ張られてしまうため、Prior 1を用いた場合と比較して Selection power が低くなった。

すべてのがん種で活性がない状況（シナリオ3-3）での動作特性は、シナリオ1-3の場合とほぼ同様であった。

表 5.2 各シナリオでのシミュレーション結果

シナリオ /Prior	Selection Power		Proportion of Success		Family-wise Error Rate (FWER)		Power		Mixing Nonactive	
	1	2	1	2	1	2	1	2	1	2
1-1	0.61	0.61	0.92	0.87	-	-	0.99	0.98	-	-
1-2	0.40	0.40	0.82	0.74	0.06	0.04	0.96	0.90	0.32	0.31
1-3	-	-	0.05	0.03	0.05	0.03	0.16	0.12	0.16	0.12
2-1	0.39	0.39	0.91	0.84	-	-	0.98	0.98	-	-
2-2	0.88	0.84	0.77	0.70	0.13	0.06	0.88	0.84	0.46	0.46
2-3	-	-	0.04	0.02	0.04	0.02	0.13	0.13	0.13	0.13
3-1	0.72	0.62	0.91	0.85	-	-	0.99	0.97	-	-
3-2	0.56	0.49	0.87	0.81	0.11	0.06	0.98	0.96	0.62	0.40
3-3	-	-	0.05	0.01	0.05	0.01	0.14	0.10	0.14	0.10

5.5 要約及び推奨事項

5.5.1 考察

- ベイズ流階層モデルを適用することで、各がん種間での奏効率の情報を借用し、より効率の良い推定が行うことができる。
- ベイズ流階層モデルを用い、無益性中止しなかったがん種のみを最終解析で頻度論流での併合解析に含めるため、治療効果の過大推定が生じ、erroneous conclusion rate が増える場合がある。各がん種での活性の有無についてベイズ流階層モデルを適用して評価することで、erroneous conclusion rate を減らすことが可能である。
- 事前分布の設定でより情報を借用するものを利用した場合、一部のがん種で奏効率が低い場合には、その結果に引っ張られてしまう可能性がある。
- がん種横断的な適応を狙いバスケット試験を行う場合には、一部のがん種で活性が認められなかった場合、その妥当性は議論になる可能性はあると考えられる。

## 5.6 参考文献

1. Pant, S., Schuler, M., Iyer, G., et al. RAGNAR Investigators (2023). Erdafitinib in patients with advanced solid tumours with FGFR alterations (RAGNAR): an international, single-arm, phase 2 study. *The Lancet. Oncology*, 24(8), 925–935.

## 6. がん第 I 相試験のデザイン

本章では、がん第 I 相試験のデザインにおけるベイズ流シミュレーションについて述べる。7.1章では、CTN 時における照会事項を想定しシミュレーションの留意点について記述する。7.2章以降の章では、様々な用量漸増デザインの中で特にベイズ流のデザインである BLRM について取り上げる。BLRM は用量反応関係について、ロジスティック回帰モデルを仮定することで、柔軟な用量反応関係を表現することができる。7.2章では、BLRM の方法論を含めた詳細について簡単に説明する。さらに異質性(レジメン、癌腫、剤型、地域など)を考慮するため BLRM を拡張したモデルである BHLRM、BHLRM+EXNEX についても合わせて説明する。7.3章では、仮想的な第 I 相用量漸増試験を設定し、BHLRM+EXNEX を用いた際の動作特性についてシミュレーションを実施して評価を行う。さらに、シミュレーションプログラムの留意点についても述べる。

### 6.1 がん第 I 相試験のシミュレーションの留意点

#### 6.1.1 デザインの概要

がん第 I 相試験における用量制限毒性(DLT) の発現率を指標に最大耐量(MTD)又は第 II 相推奨用量 (RP2D) を決定するための伝統的な3+3デザインなどの Rule-based デザインを始め、近年では連続再評価法(CRM)やベイズ流ロジスティック回帰モデル(以下 BLRM)などの Model-based デザインや、BOIN や mTPI デザインなどの Model-assisted デザインが数多く利用されている。それぞれの手法において利点と欠点が存在し、試験薬の情報や過去類似試験の実施の有無等によって使い分けることが求められる。また、新しいがん治療に適応した柔軟なデザインの必要性や、安全性と有効性のバランスを考慮した用量設定の重要性も指摘されている。各デザインの詳細については、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 2022年度タスクフォース3 成果物「近年のがん第 I 相試験デザインとその選択<sup>1)</sup>」を参照されたい。

デザインによっては、規定用量で何例の DLT 発現で増量されるのか、MTD の決定までに必要な被験者数、ある真の DLT 発現確率のシナリオ下での真の MTD を特定する確率等を事前に評価するためにシミュレーションが必要となる。医薬品医療機器総合機構 (PMDA)が公開している「30日調査照会事項チェックリスト (抗悪性腫瘍剤分野)<sup>2)</sup>」においても、「統計学的な検討に基づく用量デザインを用いる場合には、用量漸増デザインの動作特性がシミュレーション等により検討され、当該検討結果に基づき、患者の安全性を確保できることが説明すること」とある。

#### 6.1.2 がん第 I 相試験のシミュレーションにおける留意点

先に述べた通り、抗がん剤に関する30日調査については、PMDA より「30日調査照会事項チェックリスト (抗悪性腫瘍剤分野)<sup>2)</sup>」が作成されている(2024年12月10日掲載)。本チェックリストは、PMDA による照会事項作成及び届出者による照会事項に対する回答作成の負担を軽減する目的で作成されており、本チェックリストに記載の内容を事前に治験

実施計画書に記載するか、補足資料として別途提出することが推奨される。

本チェックリスト上の統計学的な観点について、“7. 忍容性の評価方法について”の項には以下の記載があり、本記載の通り、がん第I相試験において統計学的な検討に基づく用量漸増デザインを適用する際に、シミュレーションの結果を提示することが求められている。

統計学的な検討に基づく用量漸増デザインを用いる場合には、用量漸増デザインの動作特性がシミュレーション等により検討され、当該検討結果に基づき、患者の安全性を確保できることが説明されている。なお、説明内容に以下の点が含まれることを確認すること。

- 用量漸増のルール、最大耐量（以下、「MTD」）の定義、忍容性評価終了条件（全体及び各用量レベルにおける最大被験者数）等。
- 最も低い用量レベルで真のDLT発現確率が過毒性の基準をわずかに上回る場合に、各用量レベルにおけるMTD選択割合、平均被験者数及び平均DLT発現例数並びに試験中止確率について、シミュレーションに基づき検討した結果。”

また、2024年12月にPMDAより、「抗悪性腫瘍剤の第I相試験計画時の統計学的留意事項-安全性の観点から（Early Consideration）<sup>3</sup>」が公表された。その中で動作特性の評価に関して、最低限、説明が求められる内容が記載されている。さらに2025年3月に本 Early Consideration の内容の理解の促進や実例等の情報共有を目的とした説明会が開催された。説明会当日の資料についても公開されているため参考にされたい<sup>4</sup>。

治験実施計画書や提出資料に含まれていない場合は、照会事項で説明や結果の提出が求められる可能性がある。したがって、用量漸増試験におけるプロトコルを検討する際、事前に本資料について十分に理解しておくことが重要である。

抗悪性腫瘍剤の第I相用量漸増試験のシミュレーションに関する留意点について、先述した資料<sup>2,3,4</sup>にある内容と合わせ、タスクフォース参加メンバーの各社で過去に受領した照会を参考とし、整理した。

対応方針として考えられる①治験実施計画書に明記する。②動作特性への影響を事前に評価する。の2分類でまとめた。①、②については事前に対応しておくことで、照会事項対応の負担を軽減することができる。また、シミュレーションでの反映が実施可能でない場合は、定性的に動作特性への影響を評価しておくことが推奨される。30日調査は、保健衛生上の危害の発生を防止するために必要な調査とされるため、被験者の安全性の観点の上で定性的な動作特性への影響の評価が必要である。なお、この分類はあくまで参考であり、各試験における特徴に応じて検討範囲や対応方針について検討されたい。

① 治験実施計画書に明記する。(治験実施計画書以外の提出資料に記載することも可)

- ・ 用量漸増のルール
- ・ 忍容性評価の終了条件
- ・ MTDの定義
- ・ 動作特性の評価指標の提示（各用量のMTD選択確率・DLT発現例数、各用量の平

均被験者数、中止割合)

- ・ 各シナリオで仮定した各用量の DLT 発現確率についての説明（過毒性を仮定したシナリオを含む）
- ・ 過毒性シナリオで中止割合よりも最低用量が選択される確率が高い場合の対応

② 動作特性への影響を事前に評価する。（実施条件が明確な場合はプログラミング、そうでない場合は定性的な評価を行う。）

- ・ Dose skipping ルールが適用される場合
- ・ 実際に計画されているコホートサイズが OC のシミュレーションの設定と異なる場合(コホートサイズ・最小評価例数の変更が計画されている場合を含む)
- ・ 用量漸増コホートとは別の追加コホート(backfill コホートなど)の DLT 発現状況をベイズ更新に利用する場合
- ・ 用量追加(最低用量、中間用量、最高用量)を許容する場合
- ・ 先行パート(単剤, QD)の結果に応じた、後続パート(併用, BID)を実施する場合
- ・ 投与間隔・スケジュールの変更が生じる場合の説明

上記以外にも問われる可能性がある照会事項として、下記のようなものがある。

- ・ 投与スケジュールや新たなレジメンに変更する場合(Q1W/Q2W, QD/BID, priming 有無等)での OC の提示
- ・ 併用の片方の毒性だけが全用量で33%を超える場合の OC の提示
- ・ 異なるがん腫が1つのコホートに含まれる際、がん種ごとの登録速度の差を反映した OC の提示
- ・ 同一コホートに真の DLT 発現率が異なる可能性のある症例が含まれる際(摂食の有無等)、一方がすべての用量で33%を超える場合

ここまで、がん第 I 相試験のデザインにおけるシミュレーションの留意点について述べた。がん第 I 相試験において、異なるレジメンや複数のがん腫、複数の地域など異質性を考慮すべき状況が考えられる。そういった状況において用いることができる、方法について以降の次章にて紹介する。

### 6.1.3 参考資料

1. 日本製薬工業協会 医薬品評価委員会 データサイエンス部会. (2022). 近年のがん第 I 相試験デザインとその選択.  
[https://www.jpma.or.jp/information/evaluation/results/allotment/DS\\_202306\\_oncoP1DE.html](https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202306_oncoP1DE.html)
2. 医薬品医療機器総合機構. (2024). 30日調査照会事項チェックリスト（抗悪性腫瘍剤分野）. <https://www.pmda.go.jp/files/000252155.pdf>
3. 医薬品医療機器総合機構. (2024). 抗悪性腫瘍剤の第 I 相試験計画時の統計学的留意事項-安全性の観点から（Early Consideration）.

<https://www.pmda.go.jp/files/000272424.pdf>

4. Early Consideration「抗悪性腫瘍剤の第 I 相試験計画時の統計学的留意事項—安全性の観点から」説明会開催のお知らせ

<https://www.pmda.go.jp/review-services/symposia/0177.html>

## 6.2 解析で用いる統計モデル

データサイエンス部会2022年度タスクフォース3の成果物<sup>1</sup>では、BLRMの導入と併用療法におけるBLRMの拡張についての解説を行った。本節では、事前データ及び新規試験の観測データにおける異質性を考慮したBLRMの拡張であるBayesian Hierarchical Logistic Regression Model (BHLRM)、及びBHLRMに交換可能(exchangeability)及び非交換可能(Non-exchangeability)な分布を混合した分布を用いるBHLRM+EXNEXを解説する<sup>2,3</sup>。なお、拡張の差異を明確にするためにBLRM、BHLRM、BHLRM+EXNEXの順に解説し、バイズ流解析の理解で重要となる確率変数及び事前分布に指定するパラメータの依存関係などを整理する。また、Neuenschwander, et.al. (2016)<sup>3</sup>では情報借用が可能だと考えられるグループを層(stratum)と呼んでいたが、RのOncobayes2パッケージ<sup>2</sup>のvignetteでは階層モデルの枠組みに合わせて層と呼ばずにグループ(group)と呼ぶようにしているため、ここではグループと呼ぶこととする。そうすることでOncobayes2で情報借用する際に用いるgroup\_idという変数を利用する状況とも一致する。

### 6.2.1 BLRM

用量 $l$  ( $l = 1, \dots, L$ )に対して、その用量のDLT評価可能被験者数を $n_l$ 、DLT発現例数を $Y_l$ とおく。 $Y_l$ を以下の確率法則を満たす確率変数とする。

$$Y_l | \pi(d_l) \sim \text{Bin}(n_l, \pi(d_l))$$

$$\text{logit}(\pi(d_l)) = \log(\alpha) + \beta \log\left(\frac{d_l}{d_*}\right)$$

ここで、 $d_*$ をオッズ比の基準(参照用量)とし、参照用量の決め方については6.2.4節を参考にする。パラメータ $(\log(\alpha), \log(\beta))^t$ の事前分布が、以下の正規分布に従うと仮定する。

$$(\log(\alpha), \log(\beta))^t \sim N\left(\boldsymbol{\mu}, \begin{pmatrix} \tau_\alpha^2 & \rho\tau_\alpha\tau_\beta \\ \rho\tau_\alpha\tau_\beta & \tau_\beta^2 \end{pmatrix}\right)$$

ここで用いられる事前分布のパラメータ $(\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)$ は、ユーザーが指定する。なお、弱情報事前分布となるように、分散パラメータ $(\tau_\alpha, \tau_\beta)$ は $2^2$ など大きな値を指定する。新試験開始前に事前情報となる既存のデータがある場合には、上記の設定の事前分布を既存のデータでバイズ更新して、新試験の事前分布として利用することも出来る。 $(\alpha, \beta)$ は確率変数であり、 $(\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)$ はユーザーが既存又は新試験のデータでバイズ更新する前の分布を指定するために指定するパラメータである(確率変数でない)点に注意する。

BLRMモデルのパラメータ、確率変数の依存関係を以下の図 **Error! Reference source not f**

ound.に示す。なお、確率変数であるノードを黒色で表し、それ以外の固定された値を青色で表している。なお、ユーザーによる $d_*$ の選択は、事前分布のモデルパラメータの解釈及び First-in-human (FIH)試験では試験開始初期の動作特性に影響を与えることを強調するために、ユーザーの入力を表す青色のノードとした。

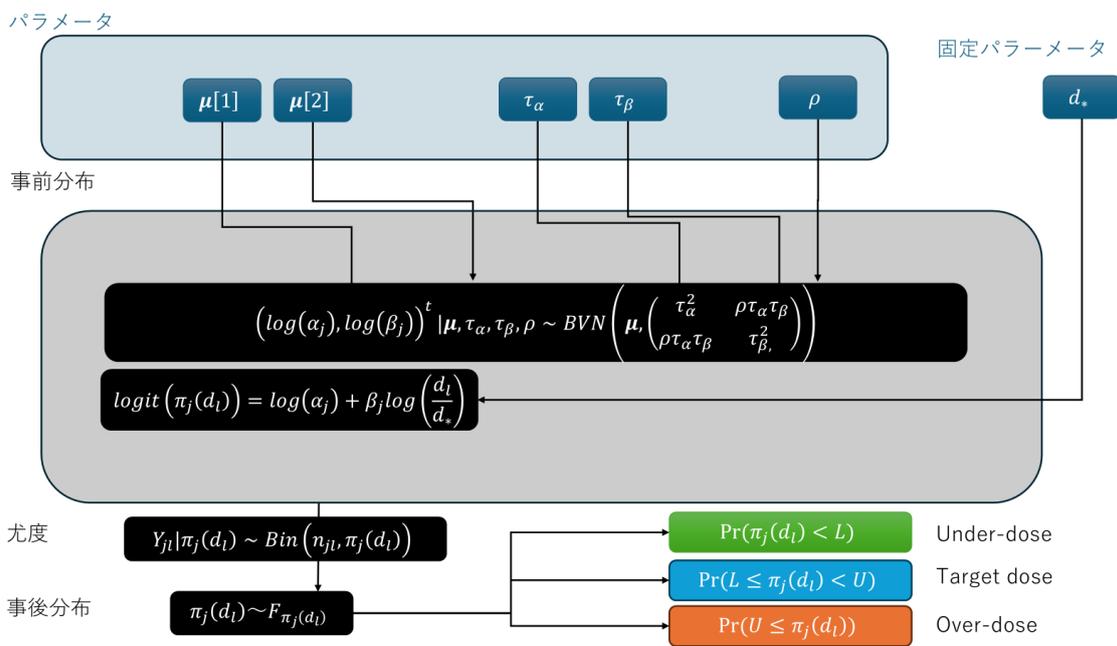


図 6-1 BLRM のパラメータ、確率変数の依存関係

確率変数 $\alpha, \beta$ の関数である $\pi(d_l)$ もまた確率変数であり、ある分布を持つ。事前分布及び事後分布を用いることで、用量 $l (l = 1, \dots, L)$ に対して $\pi(d_l)$ が以下の範囲にある確率を数値化することが可能であり、順に under-dose の確率、target dose の確率、over-dose の確率と解釈される。

$$\Pr(\pi_j(d_l) < L)$$

$$\Pr(L \leq \pi_j(d_l) < U)$$

$$\Pr(U \leq \pi_j(d_l))$$

ここで、 $[L, U)$ はターゲット毒性確率の範囲を表し、 $[0.16, 0.33)$ などが設定される。

### 6.2.2 BHLRM

既存又は新試験が異なる疾患・投与レジメンの症例を含む場合など、データの異質性を考慮する必要がある場合も考えられる。疾患・レジメンは異なっても、DLTの発現傾向はある程度類似していると考えられる場合で、かつ相互の情報借用が望まれる場合などに BHLRM が候補となりえる。ここでは、既存又は新試験の症例で、いくつかのグループがあると仮定してモデルを解説する。なお、冒頭でも述べた通り、情報借用が可能だと考える複数の症例のまとまりを Neuenschwander, et.al.(2016)<sup>3</sup>で記される”strata”(層)ではなく、

グループと呼んでいる点に注意する。

グループ  $j (j = 1, \dots, J)$  の用量  $l (l = 1, \dots, L)$  に対して、その用量の DLT 評価可能被験者数を  $n_{jl}$ 、DLT 発現例数を  $Y_{jl}$  とおく。  $Y_{jl}$  を以下の確率法則を満たす確率変数とする。

$$Y_{jl} | \pi_j(d_l) \sim \text{Bin}(n_{jl}, \pi_j(d_l))$$

$$\text{logit}(\pi_j(d_l)) = \log(\alpha_j) + \beta_j \log\left(\frac{d_l}{d_*}\right)$$

ここで、  $d_*$  を参照用量とする。パラメータ  $(\log(\alpha_j), \log(\beta_j))^t$  の事前分布が、以下の正規分布に従うと仮定する。

$$(\log(\alpha_j), \log(\beta_j))^t | \boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho \sim N\left(\boldsymbol{\mu}, \begin{pmatrix} \tau_\alpha^2 & \rho\tau_\alpha\tau_\beta \\ \rho\tau_\alpha\tau_\beta & \tau_\beta^2 \end{pmatrix}\right)$$

このうち、パラメータ  $(\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)$  は BLRM と異なり確率変数となっている点に注意する。さらに、  $(\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)$  は以下の超事前分布に従うものとする。

$$\boldsymbol{\mu} \sim N(\mathbf{m}_\boldsymbol{\mu}, \mathbf{S}_\boldsymbol{\mu})$$

$$\log(\tau_\alpha) \sim N(m_{\tau_\alpha}, \sigma_{\tau_\alpha}^2)$$

$$\log(\tau_\beta) \sim N(m_{\tau_\beta}, \sigma_{\tau_\beta}^2)$$

$$\rho \sim \text{LKJ}(\eta)$$

ここで用いられるパラメータ  $(\mathbf{m}_\boldsymbol{\mu}, \mathbf{S}_\boldsymbol{\mu}, m_{\tau_\alpha}, \sigma_{\tau_\alpha}^2, m_{\tau_\beta}, \sigma_{\tau_\beta}^2, \eta)$  は、ユーザーが指定するパラメータであり、確率変数ではない点に注意する。Lewandowski-Kurowicka-Joe (LKJ) 分布は相関係数の事前分布に使われる分布であり、特に  $\eta=1$  のときは一様分布となる<sup>4</sup>。BLRM と異なり、  $(\alpha_j, \beta_j)$  がグループ  $j$  毎に定義されており、  $(\log(\alpha_j), \log(\beta_j))^t$  はある正規分布からの  $j$  個目の(条件付きの)サンプリングとしている。その正規分布を特徴づけるパラメータ  $(\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)$  にも確率を与えることで、事前又は観測データに依存した正規分布を推測させるモデルであることに注意する。なお、  $(\log(\alpha_j), \log(\beta_j))^t$  は同じ(条件付き)正規分布を仮定しているが、グループごとのパラメータは交換可能であることを仮定している。

グループが異なるレジメンや剤型を表し、  $d_l$  と DLT の関係がグループ毎に異なる場合が考えられる。例えば、異なる剤型の投与量  $(d_l$  と  $d'_l$  と表す) の間に  $d_l = f(d'_l)$  のような変換式が仮定できる場合は、変換を行った後にモデル構築を行うことが推奨される。例えば、BID と QD のようにレジメンが異なるが、1日当たりの投与量を  $d_{jl}$  と置く場合や、異なる剤型の投与量が線形の関係にあると考えられる場合に変換を行う。なお、仮定する変換は超パラメータを不必要に大きくしないことを目的としているため、仮定が適切でない場合は、超パラメータの分散の大きさに反映されることとなる。

確率変数の依存関係を図 6-2 に示す。

図 6-2 BHLRM のパラメータ、確率変数の依存関係

### 6.2.3 BHLRM + EXNEX

BHLRM では、すべてのグループ間で情報借用を行えるように交換可能性を仮定してモデリングしたが、情報借用の程度を明示的にコントロールしたい場合が考えられる。例えば、グループの一部で DLT の傾向が異なる可能性が否定できない場合などが考えられる。そのような場合に、交換可能な分布と非交換可能な分布の混合分布を利用することも考えられる。BHLRM を拡張して、パラメータの分布に混合正規分布を用いたモデルを BHLRM+EXNEX とここでは呼ぶこととする。臨床試験での使用例として Schöffski (2022)<sup>5</sup> などがあげられる。グループ  $j$  ( $j = 1, \dots, J$ ) の用量  $l$  ( $l = 1, \dots, L$ ) に対して、その用量の DLT 評価可能被験者数を  $n_{jl}$ 、DLT 発現例数を  $Y_{jl}$  とおく。  $Y_{jl}$  を以下の確率法則を満たす確率変数とする。

$$Y_{jl} | \pi_j(d_l) \sim \text{Bin}(n_{jl}, \pi_j(d_l))$$

$$\text{logit}(\pi_j(d_l)) = \log(\alpha_j) + \beta_j \log\left(\frac{d_l}{d_*}\right)$$

ここで、 $d_*$  を参照用量とする。パラメータ  $(\log(\alpha_j), \log(\beta_j))^t$  の事前分布が以下の混合正規分布に従うと仮定する。

$$(\log(\alpha_j), \log(\beta_j))^t | \boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho \sim p_j N\left(\boldsymbol{\mu}, \begin{pmatrix} \tau_\alpha^2 & \rho\tau_\alpha\tau_\beta \\ \rho\tau_\alpha\tau_\beta & \tau_\beta^2 \end{pmatrix}\right) + (1-p_j) N(\mathbf{m}_j, \mathbf{S}_j)$$

このうち、パラメータ  $(\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)$  は確率変数とし、以下の分布に従うものとする。また、グループ毎に異なる分布を持つパラメータ  $(\mathbf{m}_j, \mathbf{S}_j)$  及び混合割合を表す  $p_j$  はユーザーが指定する。

$$\boldsymbol{\mu} \sim N(\mathbf{m}_\mu, \mathbf{S}_\mu)$$

$$\log(\tau_\alpha) \sim N(m_{\tau_\alpha}, \sigma_{\tau_\alpha}^2)$$

$$\log(\tau_\beta) \sim N(m_{\tau_\beta}, \sigma_{\tau_\beta}^2)$$

$$\rho \sim \text{LKJ}(\eta)$$

ここで用いられるパラメータ  $(\mathbf{m}_\mu, \mathbf{S}_\mu, m_{\tau_\alpha}, \sigma_{\tau_\alpha}^2, m_{\tau_\beta}, \sigma_{\tau_\beta}^2, \eta)$  は、ユーザーが指定する。  $p_1 = \dots = p_j = 1$  とすると BHLRM モデルとなり、  $p_1 = \dots = p_j = 0$  とすると  $J$  個の個別の BLRM となる。ここで  $p_j = p, \mathbf{m}_j = \mathbf{m}, \mathbf{S}_j = \mathbf{S}$  とおいたとき  $(\log(\alpha_j), \log(\beta_j))^t$  の分布は

$$(\log(\alpha_j), \log(\beta_j))^t | \boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho \sim p N\left(\boldsymbol{\mu}, \begin{pmatrix} \tau_\alpha^2 & \rho\tau_\alpha\tau_\beta \\ \rho\tau_\alpha\tau_\beta & \tau_\beta^2 \end{pmatrix}\right) + (1-p) N(\mathbf{m}, \mathbf{S})$$

となり、グループ  $j$  によらず分布が同一で、並び替えても確率法則を変えないため  $(\log(\alpha_j), \log(\beta_j))^t$  は交換可能となる。一般の場合にはグループ  $j$  によって分布が変わるため非交換可能となる。厳密には交換可能(exchangeable, EX)かつ非交換可能(non-exchangeable, NEX)であるような分布は存在しないが、情報借用すること/しないことを交換可能/非交換可能と読み替えて ENXEX と呼ばれている(以降 EXNEX と呼ぶ)。また、 $p_j$

に対して事前分布をおくことも考えられるが、定数を与える場合と大きく変わらないと考えられており<sup>3</sup>、OncoBayes2でも定数で実装されている。

BHLRM+EXNEX モデルのパラメータ、確率変数の依存関係を図 6-3に示す。なお、BHLRM を拡張した非交換可能性を意図する事前分布のパラメータを点線で図示している。

図 6-3 BHLRM+EXNEX のパラメータ、確率変数の依存関係

## 6.2.4 事前分布の Tips

過去試験・新規試験に限らず、実データを入れる前に事前分布の設定が必要となる。ここでは設定例を紹介するが、この限りでなく、モデルが用いられる実際の試験の設定で期待される動作特性 (Operational Characteristics) が得られる事前分布であれば許容されるため、事前分布設定の際の候補という位置づけとなる。一般に、動作特性は試験全体での挙動を表すために使われるが、特に FIH 試験においては被験者数が少ない状況での挙動も含めて考える必要がある。事前分布及び参照用量( $d_*$ )の設定はデータの少ない状況での挙動に影響を与えるため、その動作特性が期待されるものである設定であることを確認することも推奨される。BLRM の事前分布のパラメータ設定について触れるが、BHLRM や BHLRM+EXNEX についても同様に設定が可能である。ここでは数値例として Neuenschwander, et.al.(2016)<sup>3</sup>における事前分布の設定例を用いる。

以降の章では、正規分布においては平均と中央値は同じであり、かつ  $\log$  や  $\exp$  などの狭義単調増加関数で変換しても中央値やパーセント点は変わらないことに留意する。

切片に対応するパラメータの事前分布の平均値 $\mu[1]$ について

$$\mu[1] = E[\log(\alpha_j)] = \text{Med}[\log(\alpha_j)] = \text{Med}[\text{logit}(\pi_j(d_*))]$$

と表すことができる。従って、 $\mu[1]$ は参照となる用量( $d_*$ )における予想される DLT 発現率の  $\text{logit}$  の中央値が候補と考えることができる。仮に参照用量の DLT 発現率(の中央値)が0.2だと非臨床データなどから推測されるとき、 $\mu[1] = \text{logit}(0.2) = -1.39 \dots$ となる。

切片に対応するパラメータの事前分布の分散 $\tau_\alpha^2$ については $\pi_j(d_*)$ の95%区間を広くとることを考える。 $\text{logit}(\pi_j(d_*))$ の(2.5, 97.5)パーセント点は

$$(\mu[1] - 1.96\tau_\alpha, \mu[1] + 1.96\tau_\alpha)$$

となり、両端を  $\text{logit}$  の逆関数に代入することで $\pi_j(d_*)$ の95%区間が算出でき、十分にその区間が大きくなるような $\tau_\alpha$ を候補と考えることができる。 $\tau_\alpha = 2$ のとき、 $\pi_j(d_*)$ の95%区間は(0.005 ..., 0.926 ...)となるため、十分に幅広く不確実性をとらえていると考えられる。

傾きについては、投与量が2倍になった場合の DLT のオッズ比を考える。 $d_l = d$ と $d_l = 2d$ のときの DLT の  $\text{logit}$ (すなわち  $\log$  オッズ)は

$$\text{logit}(\pi_j(2d)) = \log(\alpha_j) - \beta_j \log\left(\frac{2d}{d_*}\right)$$

$$\text{logit}(\pi_j(d)) = \log(\alpha_j) - \beta_j \log\left(\frac{d}{d_*}\right)$$

であり、差をとることによりオッズ比を求めることができる。

$$\begin{aligned} \log \frac{\text{Odds}_j(2d)}{\text{Odds}_j(d)} &= \beta_j \log(2) = \log(2^{\beta_j}) \\ \frac{\text{Odds}_j(2d)}{\text{Odds}_j(d)} &= 2^{\beta_j} \end{aligned}$$

この中央値を考えると

$$\text{Med} \left[ \frac{\text{Odds}_j(2d)}{\text{Odds}_j(d)} \right] = \text{Med} [2^{\beta_j}] = 2^{\text{Med}[\beta_j]}$$

となる。一方で、

$$\text{Med}[\beta_j] = \text{Med}[\exp(\log(\beta_j))] = \exp(\text{Med}[\log(\beta_j)]) = \exp(E[\log(\beta_j)]) = \exp(\mu[2])$$

であるため、以下の式が得られる。

$$\text{Med} \left[ \frac{\text{Odds}_j(2d)}{\text{Odds}_j(d)} \right] = 2^{\exp(\mu[2])}$$

従って、傾きのパラメータについては、投与量が2倍となった時に想定されるオッズ比の中央値を参考に出来る。仮に $\mu[2] = \log(1)$ としたとき、投与量が2倍となった時のオッズ比の中央値は2となり、 $\pi_j(d_*) = 0.2(\text{Odds}_j(d_*) = 0.25)$ であるならば、参照用量の2倍の投与量の DLT 確率は $\pi_j(2d_*) = 0.333 \dots (\text{Odds}_j(2d_*) = 0.5)$ であることが分かる。

傾きの事前分布の分散パラメータ $\tau_\beta$ の参考としてオッズ比の95%区間を考える。オッズ比の97.5パーセント点は

$$\begin{aligned} 97.5\% \text{tile} \left[ \frac{\text{Odds}_j(2d)}{\text{Odds}_j(d)} \right] &= 2^{\exp(\mu[2] + 1.96\tau_\beta)} = 2^{\exp(\mu[2])} \exp(1.96\tau_\beta) \\ &= \left\{ \text{Med} \left[ \frac{\text{Odds}_j(2d)}{\text{Odds}_j(d)} \right] \right\}^{\exp(1.96\tau_\beta)} \end{aligned}$$

となり、2.5パーセント点についても同様に求めることで以下の95%区間を得る。

$$\begin{aligned} (2.5, 97.5)\% \text{tiles} \left[ \frac{\text{Odds}_j(2d)}{\text{Odds}_j(d)} \right] \\ = \left( \left\{ \text{Med} \left[ \frac{\text{Odds}_j(2d)}{\text{Odds}_j(d)} \right] \right\}^{\exp(-1.96\tau_\beta)}, \left\{ \text{Med} \left[ \frac{\text{Odds}_j(2d)}{\text{Odds}_j(d)} \right] \right\}^{\exp(1.96\tau_\beta)} \right) \end{aligned}$$

この区間が十分に大きくとれるような $\tau_\beta$ を事前分布に用いることも考えられる。仮に $\mu[2] = \log(1)$ とすると、

$$(2.5, 97.5)\% \text{tiles} \left[ \frac{\text{Odds}_j(2d)}{\text{Odds}_j(d)} \right] = (2^{\exp(-1.96\tau_\beta)}, 2^{\exp(1.96\tau_\beta)})$$

となり、 $\tau_\beta = 1$ とすると投与量2倍のときの DLT 確率のオッズ比の95%区間は $(2^{\exp(-1.96)}, 2^{\exp(1.96)}) = (1.103 \dots, 137.123 \dots)$ であり、十分に大きな幅を持っていると考えられる。

一方、相関については自明であることは少ないと考えられるので、0を設定する。

以上、BLRMの事前分布のパラメータ $(\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)$ の設定例を示したが、同様にBHLRM及びBHLRM+EXNEXの事前分布のパラメータ $(\mathbf{m}_\mu, \mathbf{S}_\mu)$ の設定例を示す。BHLRMの文脈では、 $\log(\alpha_j)$ の分散は以下のように表すことができる。

$$\begin{aligned} \text{Var}(\log(\alpha_j)) &= E_{\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho}[\text{Var}(\log(\alpha_j)|\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)] \\ &\quad + \text{Var}_{\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho}[E(\log(\alpha_j)|\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)] \\ &= E_{\tau_\alpha}[\tau_\alpha] + \text{Var}_{\boldsymbol{\mu}}[\boldsymbol{\mu}[1,1]] \\ &= \exp\left(m_{\tau_\alpha} + \frac{\sigma_{\tau_\alpha}^2}{2}\right) + \mathbf{S}_\mu[1,1] \end{aligned}$$

同様に $\log(\beta_j)$ については、以下のように表すことができる。

$$\text{Var}(\log(\beta_j)) = \exp\left(m_{\tau_\beta} + \frac{\sigma_{\tau_\beta}^2}{2}\right) + \mathbf{S}_\mu[2,2]$$

BLRMで $(\tau_\alpha, \tau_\beta)$ を検討した時と同様の手順で $\mathbf{S}_\mu$ の検討を行い、グループ間の違いがどの程度分散を増加させる想定かに応じて $(m_{\tau_\alpha}, m_{\tau_\beta})$ を設定する。パラメータ $(m_{\tau_\alpha}, \sigma_{\tau_\alpha}^2, m_{\tau_\beta}, \sigma_{\tau_\beta}^2, \eta)$ の設定方法についてコンセンサスはないが、Neuenschwander, et.al.(2016)<sup>3</sup>では以下の設定例を紹介している。 $\mathbf{S}_\mu[1,1] = 2$ と設定した場合、 $m_{\tau_\alpha} = \log(0.5)$ と設定することは、相対的にグループ間の違いが分散を25%程増加させると解釈できる。 $\sigma_{\tau_\alpha}^2 = (\log(4)/1.96)^2$ のように設定することは、分散を4倍に増加させることに対応し、 $\tau_\alpha$ の95%区間を $(0.125, 2)$ と設定することを意味する。

グループ間の乖離度	$m_{\tau_\alpha}$	$m_{\tau_\beta}$
Small	$\log(0.125)$	$\log(0.0625)$
Moderate	$\log(0.25)$	$\log(0.125)$
Substantial	$\log(0.5)$	$\log(0.25)$
Large	$\log(1)$	$\log(0.5)$

### 6.2.5 まとめ

過去の試験、新規試験、新規試験と同時期に行われる試験に含まれる症例が、同じDLTの発現傾向を想定できる状況であれば、BLRMを使うことで十分である。投与レジメンが異なる、がん腫が異なる、剤型が異なる、過去試験が複数あるなど、グループごとにDLTの発現傾向が異なるかもしれないが情報借用が望まれる場合は、BHLRMを使うことで、より広く得られたデータを考慮しての意思決定を可能とする。BHLRMでは $(\log(\alpha_j), \log(\beta_j))$ に対してグループ間で交換可能を仮定したが、グループの一部で若干

異なる DLT の発現傾向が考えられる場合は、混合分布を用いた BHLRM+EXNEX の利用が考えられる。モデルを複雑にすることで、利用可能なデータ量が増えるというメリットがある一方、計算コストが増えるというデメリットも考えられる。どのモデルを用いても、試験計画段階で動作特性を評価できるため、モデルに必要な仮定と真の確率法則が異なるリスクについては、動作特性に設定する真の DLT 確率で事前評価可能である。どのモデルを使うべきかは、利用可能なデータ量、パラメータの交換可能性、動作特性などを総合的に評価して選択することが推奨される。

Parameter 設定例

パラメータ	Neuenschwander, et.al. 2016	example-single-agent	example-combo2*	example-combo3*
$m_\mu$	logit(0.13), log(1)	logit(1/2), log(1)	logit(0.2) , log(1)	logit(1/3), log(1)
$S_\mu$	diag(1.98^2, 0.99^2)	diag(2^2, 1^2)	diag(2^2, 1^2)	Diag(2^2, 1^2)
$m_{\tau\alpha}$	log(0.25)	0	log(0.25)	log(0.25) for s=1, log(0.5) for s=2
$\sigma_{\tau\alpha}^2$	(log(2)/1.96)^2	0	(log(4)/1.96)^2	(log(4)/1.96)^2
$m_{\tau\beta}$	log(0.125)	0	log(0.125)	log(0.125) for s=1, log(0.25) for s=2
$\sigma_{\tau\beta}^2$	(log(2)/1.96)^2	0	(log(4)/1.96)^2	(log(4)/1.96)^2
$\eta$	1	1	1	1
$m_j$	logit(0.13) ,log(1)	NA	NA	logit(1/3),log(1)
$S_j$	diag(2^2, 1^2)	NA	NA	diag(2^2, 1^2)
$p_j$	0.0, 0.9, 1.0	NA	NA	0.9

モデルの比較

BLRM	BHLRM	BHLRM + EXNENX
Random variables $Y_l \text{ for } l = 1, \dots, L$ $\pi(d_l) \text{ for } l = 1, \dots, L$ $\alpha, \beta$	Random variables $Y_{jl} \text{ for } j = 1, \dots, J, l = 1, \dots, L$ $\pi_j(d_l) \text{ for } j = 1, \dots, J, l = 1, \dots, L$ $\alpha_j, \beta_j \text{ for } j = 1, \dots, J$ $\mu[1], \mu[2], \tau_\alpha, \tau_\beta, \rho$	Random variables $Y_{jl} \text{ for } j = 1, \dots, J, l = 1, \dots, L$ $\pi_j(d_l) \text{ for } j = 1, \dots, J, l = 1, \dots, L$ $\alpha_j, \beta_j \text{ for } j = 1, \dots, J$ $\mu[1], \mu[2], \tau_\alpha, \tau_\beta, \rho$
User-defined parameters $\mu[1], \mu[2], \tau_\alpha, \tau_\beta, \rho$ $d_*$	User-defined parameters $m_\mu[1], m_\mu[2], S_\mu[1,1], S_\mu[2,2]$ $m_{\tau_\alpha}, \sigma_{\tau_\alpha}^2, m_{\tau_\beta}, \sigma_{\tau_\beta}^2, \eta$ $d_*$	User-defined parameters $p_j, m_j[1], m_j[2], S_j[1,1], S_j[2,2] \text{ for } j = 1, \dots, J$ $m_\mu[1], m_\mu[2], S_\mu[1,1], S_\mu[2,2]$ $m_{\tau_\alpha}, \sigma_{\tau_\alpha}^2, m_{\tau_\beta}, \sigma_{\tau_\beta}^2, \eta$ $d_*$

## 6.2.6 参考文献

1. 日本製薬工業協会 医薬品評価委員会 データサイエンス部会. (2022) . 近年のがん第 I 相試験デザインとその選択.  
[https://www.jpma.or.jp/information/evaluation/results/allotment/DS\\_202306\\_oncoPIDE.html](https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202306_oncoPIDE.html)
2. Weber, S., Widmer, L., Bean, A. (2023). \_OncoBayes2: Bayesian Logistic Regression for Oncology Dose-Escalation Trials\_. R package version 0.8-9.
3. Neuenschwander, B., Wandel, S., Roychoudhury, S., and Bailey, S. (2016) Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharmaceut. Statist.*, 15: 123-134.
4. Lewandowski, Daniel; Kurowicka, Dorota; Joe, Harry (2009). "Generating Random Correlation Matrices Based on Vines and Extended Onion Method". *Journal of Multivariate Analysis*. 100 (9): 1989–2001.
5. Schöffski, P., Tan, D. S., Martín, M., et al. (2022). Phase I/II study of the LAG-3 inhibitor ieramilimab (LAG525)±anti-PD-1 spartalizumab (PDR001) in patients with advanced malignancies. *Journal for immunotherapy of cancer*, 10(2).

## 6.3 シミュレーション

### 6.3.1 シミュレーションの目的

BHLRM+EXNEX を用いた用量探索デザインを適用した場合の動作特性を確認するため、仮想的な第 I 相用量漸増試験を設定し、シミュレーションを実施する。なお、シミュレーションプログラムは本成果物の GitHub にて公開されている。

### 6.3.2 試験の設定

#### 6.3.2.1 仮想的な試験設定

- 試験目的

薬剤 A の単剤療法（1日1回 [QD] 又は1日2回 [BID]）における漸増用量の安全性及び忍容性について評価し、MTD を決定する。本シミュレーションでは QD での投与のデータがある状態で、BID での投与に切り替える状況を想定し、BID コホートを追加した場合の検討を実施することとした。

- QD投与のデータは集積済み
- これからBID投与での用量漸増を開始

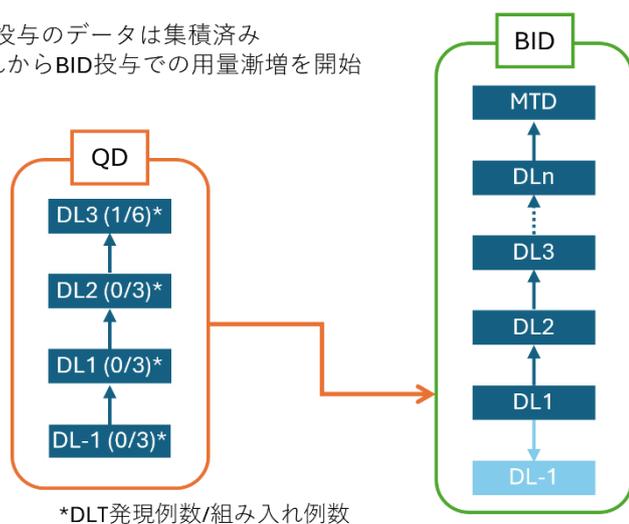


図 6-4 試験概略図

- 試験デザイン

患者数の設定：

最大で36例まで患者を組み入れることとする。コホートサイズは3～6例とする。

漸増・減量基準：

EWOC 原則（当該用量が OD となる確率が25%未満である）に基づく BHLRM+EXNEX を用いる。全ての用量レベルについて DLT 発現確率が以下に示す3つの用量区間（過少用量 [under

dose: UD]、目標用量 [target dose: TD]、過大用量 [over dose: OD]) に入る確率の事後確率を計算する。

過少用量 (UD) :	[0, 0.16)
目標用量 (TD) :	[0.16, 0.33)
過大用量 (OD) :	[0.33, 1]

新たに組み入れる患者へ投与する用量レベルは、EWOC 原則に基づき決定する。

#### 用量レベル :

検討する用量レベルは、以下の表 6.1に示す用量とする。

表 6.1 用量漸増における暫定的な薬剤 A の用量レベル

用量レベル (DL)	用量 (mg/day)
DL-1	10 (BID 1回用量 : 5 mg)
DL1 (開始用量)	20 (BID 1回用量 : 10 mg)
DL2	40 (BID 1回用量 : 20 mg)
DL3	80 (BID 1回用量 : 40 mg)
DL4	120 (BID 1回用量 : 60 mg)
DL5	160 (BID 1回用量 : 80 mg)

#### MTD/中止の基準 :

以下の基準の両方を満たす場合には、用量 $d_l$ を MTD とみなし試験を中止する。

- 少なくとも6例の患者が用量 $d_l$ で投与を受けた。
- 用量 $d_l$ で以下のいずれかを満たしている。
  - 用量 $d_l$ が TD である確率が50%を超えている。
  - 少なくとも9例の患者が本試験で投与を受けており、用量 $d_l$ が EWOC 基準を満たす最大用量である。

なお、これらの条件を満たさない場合でも、検討する用量レベルがすべて EWOC 基準を満たさなかった場合、又は最大被験者数に達した場合には、MTD を決定せずに試験を中止する。

### 6.3.2.2 統計モデル

#### • BHLRM+EXNEX モデル

用量レベル  $l$  ( $l = -1, 1, \dots, 5$ ) に対応する用量を  $d_l$  とする。用量  $d_l$  での薬剤 A の DLT 発現率について、用量-DLT モデルとして以下のロジスティックモデルを仮定する。なお、用量レベルは1日

あたりの総量で換算するものとする。

$$\text{logit}(\pi_j(d_i)) = \log(\alpha_j) + \beta_j \log\left(\frac{d_i}{d_*}\right)$$

$d_*$ を80mg/day (DL3) とし、薬剤 A の参照用量とする。また、 $j$  ( $j=1, 2$ ) は投与スケジュール ( $1 = \text{QD}, 2 = \text{BID}$ ) を表すものとする。なお、一般的に参照用量は非臨床データなどにに基づき決定される。

• 事前分布の設定

パラメータ  $(\log(\alpha_j), \log(\beta_j))^t$  の事前分布について、以下の混合正規分布に従うと仮定する。

$$(\log(\alpha_j), \log(\beta_j))^t | \mu, \tau_\alpha, \tau_\beta, \rho \sim p_j N\left(\mu, \begin{pmatrix} \tau_\alpha^2 & \rho\tau_\alpha\tau_\beta \\ \rho\tau_\alpha\tau_\beta & \tau_\beta^2 \end{pmatrix}\right) + (1 - p_j) N(\mathbf{m}_j, \mathbf{S}_j)$$

ここで、パラメータ  $(\mu, \tau_\alpha, \tau_\beta, \rho)$  は確率変数であり、以下の分布に従うとする。

$$\mu \sim N(\mathbf{m}_\mu, \mathbf{S}_\mu)$$

$$\log(\tau_\alpha) \sim N(m_{\tau_\alpha}, \sigma_{\tau_\alpha}^2)$$

$$\log(\tau_\beta) \sim N(m_{\tau_\beta}, \sigma_{\tau_\beta}^2)$$

$$\rho \sim \text{LKJ}(\eta)$$

また、パラメータ  $(\mathbf{m}_j, \mathbf{S}_j)$  は投与スケジュールごとに異なる分布を持つ。

モデルパラメータの事前分布の設定を、以下のように与えることとした。

表 6.2 モデルパラメータの事前分布

パラメータ	値	意味
$m_\mu[1]$	$\text{logit}(0.3)$	$\mu[1]$ の事前平均
$m_\mu[2]$	$\log(1)$	$\mu[2]$ の事前平均
$S_\mu[1,1]$	2	$\mu[1]$ の標準偏差
$S_\mu[2,2]$	1	$\mu[2]$ の標準偏差
$m_{\tau_\alpha}$	$\log(0.25)$	$\tau_\alpha$ の事前平均
$m_{\tau_\beta}$	$\log(0.125)$	$\tau_\beta$ の事前平均
$\sigma_{\tau_\alpha}$	$\log(2)/1.96$	$\tau_\alpha$ の標準偏差
$\sigma_{\tau_\beta}$	$\log(2)/1.96$	$\tau_\beta$ の標準偏差
$\eta$	1	相関 $\rho$ の事前パラメータ (LKJ分布)
$m_j[1]$	$\text{logit}(0.3)$	$\alpha_j$ の非交換可能な部分に対する事前平均
$m_j[2]$	0	$\beta_j$ の非交換可能な部分に対する事前平均
$S_j[1,1]$	2	$\alpha_j$ の非交換可能な部分に対する標準偏差

パラメータ	値	意味
$S_j[2,2]$	1	$\beta_j$ の非交換可能な部分に対する標準偏差
$p_j$	0.8	各投与スケジュール間での交換可能事前分布の混合割合

さらに、これまでに得られている QD 投与におけるデータとして、以下を用いた。

表 6.3 QD 投与のデータ

投与スケジュール	用量 (mg/day)	被験者数	DLT 発現例数
QD	10	3	0
QD	20	3	0
QD	40	3	0
QD	80	6	1

表 6.2の値を表 6.3の情報を用いて更新することにより得られたモデルパラメータの事前分布を、これから実施する BID コホートでのモデルパラメータの事前分布として用いることとした。更新されたモデルパラメータの事前分布を表 6.4に示す。また、各用量の DLT 発現確率の事前分布を表 6.5に示す。

表 6.4 更新されたモデルパラメータの事前分布

パラメータ	値
$m_\mu[1]$	-1.7
$m_\mu[2]$	0.27
$S_\mu[1,1]$	1.3
$S_\mu[2,2]$	0.99
$m_{\tau\alpha}$	0.27
$m_{\tau\beta}$	0.13
$\sigma_{\tau\alpha}$	0.10
$\sigma_{\tau\beta}$	0.05
$m_j[1]$	-1.51
$m_j[2]$	2.04
$S_j[1,1]$	1.58
$S_j[2,2]$	2.55

表 6.5 DLT 発現率の事前分布

用量	DLT 発現率が区間内となる	平均	SD	分位点
----	----------------	----	----	-----

(mg/day)	事前確率							
	UD	TD	OD			2.5%	50%	97.5%
10	0.888	0.049	0.063	0.069	0.157	0.000	0.011	0.641
20	0.849	0.065	0.086	0.094	0.178	0.000	0.024	0.733
40	0.768	0.109	0.233	0.138	0.204	0.000	0.057	0.818
80	0.495	0.257	0.248	0.246	0.236	0.013	0.162	0.902
120	0.311	0.238	0.451	0.371	0.288	0.021	0.287	0.972
160	0.239	0.211	0.550	0.450	0.315	0.027	0.387	0.994

### 6.3.3 シミュレーション設定

#### 6.3.3.1 シナリオ設定

動作特性を確認するにあたり、以下の3つのシナリオについてシミュレーションを実施した。

表 6.6 真の DLT 発現確率のシナリオ

シナリオ	各用量レベルにおける真の DLT 発現確率					
	10 mg/day	20 mg/day	40 mg/day	80 mg/day	120 mg/day	160 mg/day
1	0.05	0.1	0.2	0.3	0.4	0.6
2	0.33	0.4	0.5	0.6	0.7	0.8
3	0.025	0.05	0.1	0.15	0.2	0.3

シナリオ1は、各用量レベルにおける真の DLT 発現確率が UD、TD、OD の全ての区間に分散されるように設定したシナリオである。シナリオ2は、検討する最低用量でも OD となるような、過毒性の場合のシナリオである。シナリオ3は、検討する最大用量でも OD とならない場合のシナリオである。

#### 6.3.3.2 シミュレーションプログラム

R の OncoBayes2 Package<sup>1</sup>を利用し、BHLRM+EXNEX の動作特性を示すプログラムを作成した。

- **OncoBayes2 Package**

OncoBayes2 Package は、抗がん剤領域におけるベイジアンロジスティック回帰モデルを用いた用量漸増試験に向けて開発されたパッケージであり、オプションとして EXchangeability-NonEXchangeability (EXNEX) パラメータモデルを使用することで、ヒストリカルデータ及び並行して行われる試験のデータなどから柔軟な情報借用を可能としている。詳細は

OncoBayes2 Package の vignette<sup>2</sup>を参照されたい。

### 6.3.3.2.1 シミュレーションプログラムの設定

OncoBayes2 Package に含まれる関数を用い、動作特性シミュレーションを実施するプログラムを作成した。プログラムは、動作特性シミュレーションを実行するものの他、任意に与えたデータから事後分布を算出する `bhlm_demo.R` を用意した。動作特性シミュレーションのプログラムは大きく3つのパート（①試験デザイン等を設定する `bhlm_setup.R`、②1回分の試験について設定・実行する `bhlm_oc.R`、③反復試行の設定及び結果を要約する `bhlm_oc_itr.R`）に分けて構成されている。以下に、各パートごとに設定した変数を説明する。

#### ①**bhlm\_setup.R**

ここでは試験デザイン、薬剤の情報、ヒストリカルデータ、事前分布のパラメータ等の初期設定を行う。

#### 試験デザイン設定

基本的な試験のデザインについては、以下の変数を用いて設定する。

変数名	意味	設定した値
<code>group_ids</code>	交換可能性を考慮する際のグループを定義する。モデルにおける部分集団 $j$ の設定であり、検討するすべてのグループを列挙する。本シミュレーションでは投与スケジュールに対応する。	<code>c("QD","BID")</code>
<code>stratum_ids</code>	異なる情報割引を考慮する際に必要となる層を定義する。ここでの「層」はより高レベルのグループ構造を定義しており、各グループは単一の層にのみ割り当てられなければならない。詳細は OncoBayes2 Package の vignette <sup>2</sup> を参照されたい。本シミュレーションでは単一の値を用いている。	<code>c("single_stratum")</code>
<code>group_id</code>	新規試験において用いられるグループ。	<code>factor("BID", group_ids)</code>
<code>stratum_id</code>	新規試験において用いられる層。	<code>factor("single_stratum",stratum_ids)</code>

#### ヒストリカルデータの設定

tibble 型の変数 `hist_data` に、他試験のデータなど既に得られている情報を設定することができる。

hist\_data の持つ要素は以下の通りであり、過去試験におけるコホートごとにデータを入力可能である。

変数名	意味	設定した値
stratum_id	過去試験において用いられる層。	"single_stratum",
group_id	過去試験において用いられるグループ。	"QD"
drug_A	当該コホートにおける薬剤の用量（1日あたりの総量換算）。	表 6.3参照のこと
num_patients	当該コホートにおける被験者数。	表 6.3参照のこと
num_toxicities	当該コホートにおける DLT 発現例数。	表 6.3参照のこと
cohort_time	当該コホートが何番目かを示す。過去試験の場合は0とする。	0

### **薬剤情報 : drug info**

tibble 型の変数 drug\_info に、薬剤の基本情報を設定することができる。drug\_info の持つ要素は以下の通りである。

変数名	意味	設定した値
drug_name	薬剤名。hist_data 及びその他の薬剤の用量を設定する変数名と同じ値とする。	"drug_A",
dose_ref	試験における参照用量。	80
dose_unit	用量の単位。	"mg/day",

### **用量レベル設定 : dose info**

tibble 型の変数 dose\_info に、新規試験で検討する用量レベルを設定することができる。dose\_info の持つ要素は以下の通りである。

変数名	意味	設定した値
stratum_id	新規試験において用いられる層。	factor("single_stratum", stratum_ids)
group_id	新規試験において用いられるグループ。	factor(c("BID"), group_ids)
drug_A	検討するすべての用量（1日あたりの総量換算）。	c(10, 20, 40, 80, 120, 160)

### **開始用量設定 : starting doses**

tibble 型の変数 starting\_doses に、新規試験で検討する開始用量を設定することができる。starting\_dose の持つ要素は以下の通りである。

変数名	意味	設定した値
-----	----	-------

stratum_id	新規試験において用いられる層。	factor("single_stratum", stratum_ids)
group_id	新規試験において用いられるグループ。	factor(c("BID"), group_ids)
drug_A	開始用量。	20

### 事前分布のパラメータ設定

モデルの事前パラメータについては、以下の変数で設定する。

変数名	対応するパラメータ	設定した値
prior_mu_mean	$m_\mu[1], m_\mu[2]$	c(logit(0.3), log(1))
prior_mu_sd	$S_\mu[1,1], S_\mu[2,2]$	c(2, 1)
prior_tau_mean	$m_{\tau\alpha}, m_{\tau\beta}$	c(log(0.25), log(0.125))
prior_tau_sd	$\sigma_{\tau\alpha}, \sigma_{\tau\beta}$	c(log(2)/1.96, log(2)/1.96)
prior_prob	$p_j$	0.8
prior_glm_u_mean	$m_j[1], m_j[2]$	c(logit(0.3), log(1))
prior_glm_u_sd	$S_j[1,1], S_j[2,2]$	c(2,1)

### ②bh1rm\_oc.R

ここでは許容する増量幅、最大被験者数及び最小被験者数、コホートあたりの被験者数、中止基準 (MTD 評価に必要な最小被験者数、用量 d を MTD として決定する際に超えるべき TD の確率) を設定する。また、シミュレーションにおけるシナリオ (真の DLT 発現確率) もここで設定する。設定すべき変数は以下のとおりである。

変数名	意味	設定した値
max_percent_inc	許容する増量倍率。	2
max_patients	組み入れる最大被験者数。	36
min_patients	MTD 決定時に必要となる試験全体で組み入れられた最小の被験者数。	9
cohort_sizes	各コホートあたりの組み入れ被験者数。本試験では3~6例であるため、シミュレーションでは3~6例がランダムに組み入れられる設定としている。	3:6
cohort_sizes_probs	コホートあたりの組み入れ被験者数が複数である場合、何例になるかの確率。	rep(1/length(cohort_sizes), length(cohort_sizes))
patients_on_dose_for_MTD	MTD として決定される用量に最	6

	低限投与されるべき被験者数。	
prob_target_cond	用量 $d_i$ を MTD として決定する際に超えるべき TD の確率。	0.5
tureDLTs	シミュレーションシナリオにおける真の DLT 発現確率。検討する用量レベルと同じ長さであること。	表 6.6参照のこと。

### ③bhlrm\_oc\_itr.R

ここでは反復試行についての設定をする。設定すべき変数は nTrials のみであり、ここで試行回数を設定することで、反復試行のシミュレーションを実行することが可能である。本シミュレーションでは1000回の試行を実施することとした。

## 6.3.4 シミュレーション結果

表 6.6で設定した3つのシナリオに対する BHLRM+EXNEX を用いたシミュレーション実験の結果を以下に示す。なお、交換可能事前分布の混合割合 $p_j$ は表 6.2に示した通り0.8とした。また、結果は3つの用量区間及び用量レベルごとに提示している。

表 6.7 各シナリオに対する BLRM の動作特性

シナリオ	MTD として選択される割合 (%)			中止の割合 (%)		平均被験者数				平均 DLT 発現例数			
	UD	TD	OD	過毒性 <sup>a</sup>	最大被験者数 <sup>b</sup>	UD	TD	OD	計	UD	TD	OD	計
1	28.5	64.7	0.7	5.9	0.2	7.0	7.6	0.5	15.1	0.7	1.8	0.2	2.7
2	-	-	23.1	76.9	0.0	-	-	8.3	8.3	-	-	3.4	3.4
3	71.2	26.5	-	1.7	0.6	15.2	3.6	-	18.8	1.4	0.9	-	2.3

a.検討する最低用量でも EWOC 基準が満たされず、中止となった割合。

b.最大被験者数に到達したが MTD が決定できず、中止となった割合。

表 6.8 用量ごとの各シナリオに対する BLRM の動作特性 (MTD 選択割合)

シナリオ/ 用量(mg/day)	MTD として選択される割合 (%)					
	10	20	40	80	120	160
1	0.8	27.7	52.4	12.3	0.6	0.1

シナリオ/ 用量(mg/day)	MTD として選択される割合 (%)					
	10	20	40	80	120	160
2	5.2	15.6	2.3	0.0	0.0	0.0
3	0.1	8.6	38.2	24.3	9.8	16.7

表 6.9 用量ごとの各シナリオに対する BLRM の動作特性 (平均被験者数)

シナリオ/ 用量(mg/day)	平均被験者数						
	10	20	40	80	120	160	計
1	0.6	6.4	5.6	2.0	2.7	2.3	15.1
2	1.6	5.9	0.8	0.0	0.0	0.0	8.3
3	0.2	5.5	5.6	3.8	1.6	1.9	18.8

表 6.10 用量ごとの各シナリオに対する BLRM の動作特性 (平均 DLT 発現例数)

シナリオ/ 用量(mg/day)	平均 DLT 発現例数						
	10	20	40	80	120	160	計
1	0.0	0.6	1.2	0.6	0.1	0.1	2.7
2	0.5	2.4	0.4	0.0	0.0	0.0	3.4
3	0.0	0.3	0.5	0.6	0.3	0.6	2.3

シナリオ1では、TD が適切に MTD として選択される割合が64.7%と最も高い結果であり、用量別では40mg/day (DL2) が MTD として選択される割合が52.4%と最も高かった。また、OD が MTD として選択される割合は0.7%と低く、OD に割り付けられた平均被験者数は0.5例であり、過大用量への投与は抑制できていると考えられる。

検討するすべての用量が OD となるシナリオ2では、検討する最低の用量レベルでも過毒性となり中止となる確率が76.9%と高く、また平均被験者数は8.3例、平均 DLT 発現例数は3.4例と少ない被験者数に抑えられていた。

すべての用量が OD に満たないシナリオ3では、UD が MTD として選択される確率が71.2%と最も高かったが、これは UD に相当する最高用量における DLT 発現確率が0.15と、TD に相当する DLT 発現確率の下限に近かったことが一因と考えられる。また、事前分布に用いた QD のデータにおいて、UD である80mg/day (DL3) の用量で6例中1例に DLT の発現があったことも影響を与えている可能性がある。

以上の結果より、シナリオの設定によってはUDが選択される可能性が高くなる場合もあり、やや保守的な傾向はあるものの、本デザインを用いることでMTDは適切に評価でき、また過大用量への投与についても抑制できると考えられる。

### 6.3.5 交換可能事前分布の混合割合を変えた場合の動作特性

BHLRM+EXNEXの特徴として、交換可能事前分布の混合割合 $p_j$ を設定することができる点が挙げられる。この値を1とした場合は単純な階層ベイズを用いたBHLRMとなり、0とした場合にはグループ間の情報の借り入れを考慮せず独立に各試験を評価するBLRMとなる。実際の試験においては、交換可能事前分布の混合割合は臨床的な観点などを総合的に判断して決定される。ここでは参考として、 $p_j$ の値を1及び0とした場合のシミュレーション結果を示す。なお、プログラムの動作の都合上、prior\_probに0を設定できなかったため、本シミュレーションでは0.000001と設定した。

表 6.11  $p_j$ を変えた場合の各シナリオに対する BLRM の動作特性

シナリオ	MTD として選 択される割合 (%)			中止の割合 (%)		平均被験者数				平均 DLT 発現例数			
	UD	TD	OD	過毒 性 <sup>a</sup>	最大 被験 者数 <sup>b</sup>	UD	TD	OD	計	UD	TD	OD	計
1 ( $p_j = 1$ )	29.1	63.6	0.9	6.4	0.0	6.9	7.5	0.6	15.0	0.7	1.7	0.3	2.7
2 ( $p_j = 1$ )	-	-	25.6	74.4	0.0	-	-	8.7	8.7	-	-	3.4	3.4
3 ( $p_j = 1$ )	74.0	24.1	-	1.0	0.9	15.1	3.5	-	18.6	1.4	0.9	-	2.3
1 ( $p_j = 0$ )	29.5	63.8	1.2	5.5	0.0	7.0	7.3	0.5	14.8	0.7	1.7	0.2	2.6
2 ( $p_j = 0$ )	-	-	24.8	75.2	0.0	-	-	8.4	8.4	-	-	3.4	3.4
3 ( $p_j = 0$ )	75.7	22.7	-	1.1	0.5	15.2	3.1	-	18.3	1.5	0.8	-	2.2

a. 検討する最低用量でも EWOC 基準が満たされず、中止となった割合。

b. 最大被験者数に到達したが MTD が決定できず、中止となった割合。

表 6.12  $p_j$ を変えた場合の用量ごとの各シナリオに対する BLRM の動作特性 (MTD 選択割合)

シナリオ/ 用量(mg/day)	MTD として選択される割合 (%)					
	10	20	40	80	120	160
1 ( $p_j = 1$ )	1.3	27.8	49.8	13.8	0.7	0.2
2 ( $p_j = 1$ )	7.6	16.3	1.7	0.0	0.0	0.0
3 ( $p_j = 1$ )	0.6	7.7	39.7	26.0	9.6	14.5
1 ( $p_j = 0$ )	1.1	28.4	50.9	12.9	0.7	0.5
2 ( $p_j = 0$ )	5.1	16.4	3.1	0.2	0.0	0.0
3 ( $p_j = 0$ )	0.3	8.2	42.8	24.4	8.4	14.3

表 6.13  $p_j$ を変えた場合の用量ごとの各シナリオに対する BLRM の動作特性 (平均被験者数)

シナリオ/ 用量(mg/day)	平均被験者数						
	10	20	40	80	120	160	計
1 ( $p_j = 1$ )	0.5	6.4	5.4	2.1	0.3	0.3	15.0
2 ( $p_j = 1$ )	1.8	6.0	0.8	0.0	0.0	0.0	8.7
3 ( $p_j = 1$ )	0.3	5.4	5.7	3.7	1.6	1.9	18.6
1 ( $p_j = 0$ )	0.6	6.5	5.5	1.8	0.3	0.2	14.8
2 ( $p_j = 0$ )	1.7	5.9	0.8	0.1	0.0	0.0	8.4

シナリオ/ 用量(mg/day)	平均被験者数						計
	10	20	40	80	120	160	
3 ( $p_j = 0$ )	0.3	5.5	5.8	3.7	1.4	1.7	18.3

表 6.14  $p_j$ を変えた場合の用量ごとの各シナリオに対する BLRM の動作特性 (平均 DLT 発現例数)

シナリオ/ 用量(mg/day)	平均 DLT 発現例数						計
	10	20	40	80	120	160	
1 ( $p_j = 1$ )	0.0	0.7	1.1	0.6	0.1	0.2	2.7
2 ( $p_j = 1$ )	0.6	2.4	0.4	0.0	0.0	0.0	3.4
3 ( $p_j = 1$ )	0.0	0.3	0.6	0.5	0.3	0.6	2.3
1 ( $p_j = 0$ )	0.0	0.6	1.2	0.5	0.1	0.1	2.6
2 ( $p_j = 0$ )	0.6	2.4	0.4	0.0	0.0	0.0	3.4
3 ( $p_j = 0$ )	0.0	0.3	0.6	0.6	0.3	0.5	2.2

今回のシナリオの設定の下では、交換可能事前分布の混合割合を変えた場合でも、動作特性に大きな違いは見られなかったが、 $p_j = 0.8$ とした BHLRM+EXNEX モデルを用いた場合の方が、平均被験者数や平均 DLT 発現確率は同程度でありながらも、シナリオ1及び3では TD が適切に MTD として選択される割合が高かった。また、すべての用量が過毒性となるシナリオ2では、過毒性として適切に中止となる割合が高かった。実際の試験においては、交換可能事前分布の混合割合はグループ間の異質性や臨床的な観点などを総合的に判断して決定されるべきであろう。

### 6.3.6 事後分布の生成による仮想的な用量漸増シナリオ

実際の試験を計画する上では、データが得られた下での仮想的な用量漸増シナリオを確認することも有用である。bhlrm\_demo.R は、任意のデータを与えた下での事後分布を算出するプログラムであり、仮想的な用量漸増シナリオを作成することができる。以下にいくつかのシナリオを示す。

表 6.15 仮想的な用量漸増シナリオ

シナリオ	用量 (mg/day)	DLT/N	次の推奨用量 (ND) (mg/day)	ND が TD 区間に入る確率	ND が OD 区間に入る確率
1	20	0/3	40	0.164	0.092
2	20	1/3	10	0.264	0.196

シナリオ	用量 (mg/day)	DLT/N	次の推奨用量 (ND) (mg/day)	ND が TD 区間に 入る確率	ND が OD 区間に 入る確率
3	20	2/3	中止	-	-
4	20	0/4	40	0.150	0.073
5	20	1/4	20	0.325	0.222
6	20	1/5	20	0.312	0.152
7	20	1/6	20	0.293	0.101
8	20	2/6	10	0.355	0.193
9	20	0/3	80	0.208	0.194
	40	0/3			
10	20	0/3	40	0.340	0.190
	40	1/3			
11	20	1/3	20	0.322	0.149
	10	0/3			
12	20	1/3	中止	-	-
	10	1/3			
13	20	0/4	80	0.208	0.182
	40	0/3			
14	20	0/4	40	0.333	0.163
	40	1/3			
15	20	1/4	40	0.349	0.249
	20	0/3			
16	20	1/4	10	0.334	0.131
	20	1/3			

シナリオ1は、開始用量である20mg/day (DL1) で3例の参加者を組み入れ、DLT発現が0例であったというデータが得られた場合の次の推奨用量 (ND) を示している。このデータが得られた場合、次の用量としては40mg/day (DL2) が推奨されることとなる。同様に、開始用量で3例中1例にDLTが認められた場合にはNDは10mg/day (DL-1) となり (シナリオ2)、3例中2例にDLTが認められた場合には過毒性により試験の中止が推奨される (シナリオ3)。シナリオ4~8は、開始用量への組み入れ被験者数が4、5、6例であった場合のいくつかの仮想的な用量漸増シナリオを示している。また、シナリオ9~16は、開始用量でのデータから推奨されたNDでの投与を続け、さらに次のコホートで推奨される用量を示した用量漸増シナリオである。例えば、シナリオ9では開始用量である20mg/day (DL1) で組み入れられた3例でDLT発現が0例であり、続くコホートでNDとされた40mg/day (DL2) に3例を組み入れ、DLT発現が0例であった場合のシナリオであるが、この時

のNDは80mg/day (DL3) となった。

このように、実際の試験を計画する際は、仮想的な用量漸増シナリオをいくつか検討し、適切に増量/減量の判断ができるかどうかについても、必要に応じて検討すべきであろう。

### 6.3.7 参考文献

1. Sebastian, W., Lukas, AW., Andrew, B. Package ‘OncoBayes2’. <https://cran.r-project.org/web/packages/OncoBayes2/index.html>
2. Andrew, B., Sebastian, W., Lukas, W., Guiding Oncology Dose-Escalation Trials. <https://cran.r-project.org/web/packages/OncoBayes2/vignettes/introduction.html>

## 7. おわりに

本成果物では、ベイズ流試験デザインを適用する際に、その動作特性を測るためのシミュレーションの方法をまとめることで、規制当局と議論となりうるポイントをまとめた。ベイズ流試験デザインを適用することで、試験のデザインと解析に柔軟性を提供でき、複数のエビデンスを体系的に組み合わせることが可能になる一方で、事前分布や意思決定の基準によっては誤った結論を導く可能性の増加があり得る。そのため、本報告書で紹介したようなシミュレーションによる動作特性の確認は重要である。ただし、実際の規制当局との論点や受け入れ可能性は、当該薬剤の開発状況や対象疾患等によって異なると考えられるため、各状況に応じて判断していただきたい。本報告書がベイズ流試験デザインでシミュレーションをする際の、議論ポイントやその実装方法について、理解を深める一助となれば幸いである。

## Appendix1 : ベイズ解析の基礎

ベイズ流解析は、ベイズの定理が基礎となっている。ベイズの定理を簡潔に説明すると、興味のあるパラメータの事前確率（事前に研究者が設定する）と尤度（実際に得られたデータに基づき計算される）を結合し、パラメータの事後確率を得るものである。臨床試験では、得られた事後分布を用いて、治療効果の代表値についての信用区間を算出したり、あるいは事前に規定した閾値を超える確率を求めるなど、治療効果についての確率を用いた定量的な評価を行うことができる。事前分布は、情報のない事前分布（無情報事前分布：特定のパラメータの範囲にピークを持たない漠然とした分布）と **Informative** な事前分布に大別される<sup>1</sup>。推定したいパラメータが多くなった場合等には、事後分布を数理的に解くことができない場合があるが、その対処法として、事後分布に基づく乱数を発生させ近似的に事後分布を得る Markov Chain Monte Carlo (MCMC) 法等がある。一方で推定すべきパラメータが一つの場合等のシンプルな状況では、共役性の性質を利用することで事前分布と尤度から事後分布の計算が著しく簡単になる場合がある。

事前分布と事後分布が同じ型の分布になるという性質を共役性と呼び、例えば事前分布をベータ分布、尤度を二項分布とすれば、ベイズの定理により事後分布はベータ分布となることが知られている。具体的には、確率変数  $X$  の尤度が二項分布  $X \sim B(n, p)$  に従い（ $n$  : 被験者数、 $p$  : 成功確率）、二項分布のパラメータ  $p$  の事前分布を  $p \sim \text{Beta}(\alpha, \beta)$ （ベータ分布の期待値： $\frac{\alpha}{\alpha+\beta}$ ）とすれば、 $x$ （成功例）が与えられた下での  $p$  の事後分布は  $P \sim \text{Beta}(\alpha + x, \beta + n - x)$  となる（ベータ分布の事後期待値： $\frac{\alpha+x}{(\alpha+x)+(\beta+n-x)}$ ）。

解析時に無情報事前分布を使用した場合には、実際に得られたデータに依存して事後分布が構成される傾向になる。一方で **Informative** な事前分布を使用した場合には、その情報量（ $\alpha + \beta$ ）を多く持たせるほど、得られたデータが事後分布に与える影響は小さくなる傾向になる。例えば、成功確率  $p$  の事前分布として、 $\text{Beta}(1, 1)$ （期待値： $\frac{1}{1+1} = 0.5$ 、情報量： $1 + 1 = 2$ ）を採用し、二項分布の下で15例中10例が成功したというデータを得た場合、事後分布は  $\text{Beta}(11, 6)$ （事後期待値： $\frac{11}{11+6} \cong 0.65$ 、情報量： $11 + 6 = 17$ ）となる。頻度論に基づく期待値は  $\frac{10}{15} \cong 0.67$  であるのに対し、ベイズ流解析に基づく事後期待値は0.65であることから、得られたデータが事後分布に与える影響が大きいことが分かる。一方で、成功確率  $p$  の事前分布として  $\text{Beta}(2, 9)$ （期待値： $\frac{2}{2+9} \cong 0.18$ 、情報量： $2 + 9 = 11$ ）を採用し、同じく二項分布の下で15例中10例が成功したというデータを得た場合、事後分布は  $\text{Beta}(12, 14)$ （事後期待値： $\frac{12}{12+14} \cong 0.46$ 、情報量： $12 +$

14 = 26) となる。頻度論に基づく期待値は $\frac{10}{15} \cong 0.67$ であるのに対し、ベイズ流解析に基づく事後期待値は0.46であることから、得られたデータが事後分布に与える影響が小さいことが分かる。

また尤度 (情報量) が大きいほど、設定した事前分布に関わらず尤度 (情報量) に依存して事後分布が構成される傾向になる。例えば、成功確率  $p$  の事前分布として、 $Beta(1,1)$ を採用し、二項分布の下で150例中10例が成功したというデータを得た場合、事後分布は $Beta(11,141)$  (事後期待値： $\frac{11}{11+141} \cong 0.072$ 、情報量： $11 + 141 = 152$ ) となる。一方で成功確率  $p$  の事前分布として $Beta(2,9)$ を採用し、同じく二項分布の下で150例中10例が成功したというデータを得た場合、事後分布は $Beta(12,149)$  (事後期待値： $\frac{12}{12+149} \cong 0.075$ 、情報量： $12 + 149 = 161$ ) となる。よって、頻度論に基づく期待値は $\frac{10}{150} \cong 0.067$ であるのに対し、ベイズ流解析に基づく事後期待値はそれぞれ0.072及び0.075であることから、設定した事前分布に関わらず尤度 (情報量) に依存して事後分布が構成されることが分かる。

さらに、使用する事前分布は同じとして、尤度 (情報量) が大きい場合と小さい場合の事後分布を比較すると、前者の方が尤度 (情報量) が事後分布に与える影響が大きくなる傾向になる。例えば、成功確率  $p$  の事前分布として $Beta(2,9)$ を採用し、二項分布の下で150例中100例が成功したというデータを得た場合、事後分布は  $Beta(102,59)$  (事後期待値： $\frac{102}{102+59} \cong 0.63$ 、情報量： $102+59=161$ ) となる。一方で同じく成功確率  $p$  の Informative な事前分布として  $Beta(2,9)$ を採用し、二項分布の下で15例中10例が成功したというデータを得た場合、事後分布は  $Beta(12,14)$  (事後期待値： $\frac{12}{12+14} \cong 0.46$ 、情報量： $12+14=26$ ) となる。よって、尤度 (情報量) が異なることで与える情報量によって事後分布への影響が異なることが分かる。

一般的に知られている共役性の例を以下に示す。

事前分布	尤度	事後分布
ベータ分布	二項分布	ベータ分布
ガンマ分布	ポアソン分布	ガンマ分布
正規分布	正規分布の平均	正規分布
逆ガンマ分布	正規分布の分散	逆ガンマ分布

<参考部文献>

1. 竹林由武. (2018). しなやかな臨床試験デザイン：適応型デザインによる効率化. *Japanese Psychological Review*, 61(1), 86-100.

## 執筆者・タスクフォースメンバー

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

2024年度タスクフォース3-2 STAT 新時代

ベイズ流試験デザインにおけるシミュレーションの利用法

### 【チームメンバー（本稿執筆者，50音順）】

相澤 愛奈	ファイザーR&D 合同会社	担当：6.3章
石田 弘輔	ノバルティスファーマ株式会社	担当：6.2章
伊庭 克拓	大塚製薬株式会社	担当：4章
遠藤 麻衣	田辺三菱製薬株式会社	担当：3.1章
栗城 正朗	協和キリン株式会社	担当：4章
高戸 俊一	杏林製薬株式会社	担当：3.2章（3.2.5章を除く）
田中 大樹	鳥居薬品株式会社	担当：2章（2.2.1章，2.3章を除く），2.3章の SAS プログラム
田中 優一	大鵬薬品工業株式会社	担当：3.2章（3.2.5章を除く）
畑山 知慶	ヤンセンファーマ株式会社	担当：4章
東川 正晃	EA ファーマ株式会社	担当：3.1章
松田 裕也	中外製薬株式会社	担当：4章、5章
吉田 和正	日本ベーリンガーインゲルハイム株式会社	担当：6.1章
吉本 拓矢	中外製薬株式会社	担当：2.2.1章, 2.3章, 3.2.5章

### 【タスクフォースリーダー，推進委員】

竹ノ内 一雅	日本イーライリリー株式会社	
豊泉 樹一郎	ヤンセンファーマ株式会社	担当：1章、5章、7章

### 【担当副部長】

菅波 秀規	興和株式会社	
-------	--------	--