

第8回データサイエンスラウンドテーブル会議

ベイズ流手法を利用する際の留意点
ーCID事例に基づく検討

タイムテーブル

時間	所要時間	内容
10:00-10:15	15分	開会の挨拶
10:15-10:45	30分	ベイズ流テーマの説明
10:45-12:15	90分	テーブルディスカッション
12:15-13:15	60分	休憩
13:15-13:35	30分	各テーブルの議論共有
13:35-14:15	30分	ベイズ流手法全体でディスカッション
14:15-14:30	15分	休憩
14:30-15:45	75分	全体発表(各テーマ25分)
15:45-16:00	15分	閉会の挨拶

本テーマのファシリテーター

テーブルディスカッションでのファシリテーターを務めます

- 野村 尚吾 (東京大学)
- 浅野 淳一 (PMDA)
- 田中 優一 (大鵬)
- たぎし 和隆 (千寿)
- 伊庭 克拓 (大塚)
- 白石 亜矢子 (ヤンセン)

ベイズ流テーマの説明

1. テーブルディスカッションの進め方(10分)

2. CID/ベイズを採用する必要性と現状(20分)

野村 尚吾 (東京大学)

3. テーブルに分かれて事例紹介(20分)

DLBCL: 田中 優一(大鵬)

慢性疼痛: たぎし 和隆(千寿)

SLE: 伊庭 克拓(大塚)

今回取り上げる事例

グループ名	対象疾患	事例の特徴	テーブル番号
DLBCL	びまん性大細胞型B細胞リンパ腫	Phase III 外部情報を利用する試験でベイズ流手法を用いた事例	A
慢性疼痛	慢性疼痛	Phase II マスタープロトコルで疾患横断的に情報借用したベイズ流階層モデルを用いた事例	B
SLE	全身性エリテマトーデス	Phase IIb 用量群間の情報を借用したベイズ流階層モデルを用いた事例	C

参考資料: 日本製薬工業協会 (2024). “Complex Innovative Trial Designの現状と事例の整理 - FDAガイダンス邦訳付き”
https://www.jpma.or.jp/information/evaluation/results/allotment/tcjmdm00000010pp-att/DS_202404_CID.pdf

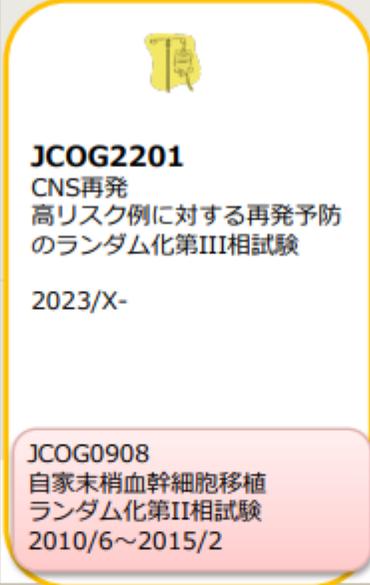
びまん性大細胞型B細胞リンパ腫：
外部情報を利用する試験でベイズ流手法を用いた事例

びまん性大細胞型B細胞リンパ腫 (Diffuse Large B-Cell Lymphoma; DLBCL)

- DLBCLは、白血球のうちリンパ球ががん化する悪性リンパ腫のうち、白血球の一種であるリンパ球のうちBリンパ球ががん化することで発症する疾患
 - 悪性リンパ腫の1つであるホジキンリンパ腫以外の悪性リンパ腫のうち3-4割を占める
 - DLBCLには複数の亜型が存在し、極めて多様な集団を含む不均一な疾患
- DLBCLの治療体系(詳細は次ページ)
 - 一次標準治療はR-CHOP療法であり、病期診断(限局期, 進行期)に従って治療方針が変化
 - R-CHOP療法: リツキシマブ, シクロホスファミド, ドキソルビシン, ビンクリスチン, プレドニゾロンを併用
 - R-CHOP療法は20年以上前に確立された治療法であり安全で効果的な治療であるものの、45%-50%の患者で再発
 - DLBCL患者の10年生存率は約50%([Shi Q, et al. 2018](#))
 - 特定のサブグループ集団では一次標準治療での予後が悪く、一定の治療ニーズが存在

DLBCLの治療開発状況

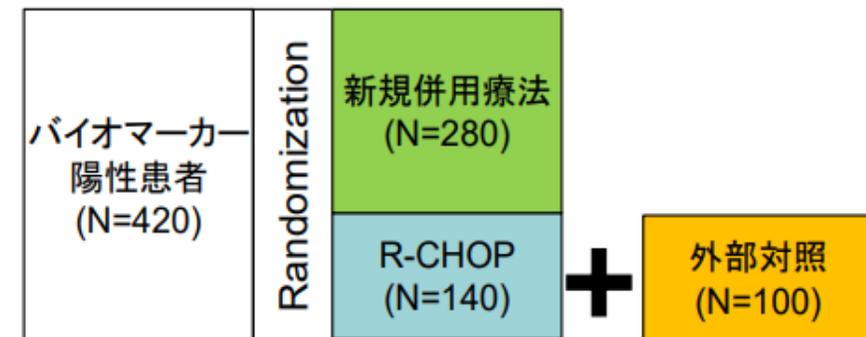
[LSG_02.pdf \(jcog.jp\)](#)

対象 病期 国際予後指標		標準治療	JCOG試験	他組織の試験	主な未承認薬/ 適応外薬			
初 発	限 局 期	リスクなし	R-CHOP (4コース)	 <p>JCOG2201 CNS再発 高リスク例に対する再発予防 のランダム化第III相試験</p> <p>2023/X-</p>	<p>Non GCB DLBCLに対するR-CHOP +/- アカラブルチニブのランダム化第III相試験 (アストラゼネカ) (75歳以下)</p>			
		全リスク	R-CHOP (3コース) + 領域照射					
		低中~高 リスク	R-CHOP (6~8コース) (JCOG0601) ボラツズマブベドチン +R-CHP (6コース)					
	進 行 期	低リスク	R-CHOP (6~8コース) (JCOG0601)				<p>JCOG0908 自家末梢血幹細胞移植 ランダム化第II相試験 2010/6~2015/2</p>	<p>R-CHOP +/- エブコリタマブのランダム化第III相試験 (ジェンマブ・アツヴィ)</p> <p>R-CHOP +/- タファシタマブ + レナリドミドのランダム化第III相試験 (インサイト・バイオサイエンス・ジャパン)</p>
		低中リスク						
		中高リスク	ボラツズマブベドチン +R-CHP (6コース)					
高リスク								
再 発	全 病 期/ 全 リ ス ク	救援化学療法および ASCT併用大量化学療法		<p>自家移植非適応患者に対する救援化学療法とエブコリタマブの第III相試験 (ジェンマブ)</p> <p>ツシジノスタット・リツキシマブの第II相試験 (Meijiファルマ)</p> <p>ロンカスツキシマブ・テシリンの第II総試験 (田辺三菱)</p>	<p>モスネツズマブ+ボラツズマブ・ベドチンと化学療法の第III相試験 (中外)</p> <p>オドロネクスタマブの第II相試験 (REGENERON)</p> <p>タファシタマブ + レナリドミドの第II総試験 (インサイト・バイオサイエンス・ジャパン)</p>	<p>セリネキサ</p> <p>オドロネクスタマブ</p>		
		BR ±ボラツズマブベドチン						
		エブコリタマブ						
		チサゲンレクルユーセル* アキシカブタゲンシロルユーセル*, ** リソカブタゲンマラルユーセル*, **						

R-CHOP (リツキシマブ、シクロホスファミド、ドキシソルピシン、ピンクリスチン、プレドニゾロン)、領域照射: 治療前に病変があった領域のみに照射する方法
*: 初発では2回以上、再発では再発後に化学療法を1回以上施行し、完全奏効が得られなかった又は完全奏効が得られたが再発した場合で自家造血幹細胞移植 (ASCT) の適応とならない又はASCT後に再発した患者に限る。 ** 初回治療不応または初回治療から12ヶ月以内に再発した場合。

試験デザインの概要[1]

- 特定のバイオマーカーが陽性である, DLBCLの一次療法を対象とする第3相非盲検ランダム化比較試験
 - ランダム化比率 試験併用療法群 : 対照群 (R-CHOP) = 2 : 1
 - 重要な副次評価項目である全生存期間 (Overall Survival; OS) の解析にて外部対照データを利用
 - 用いる外部対照データには同一社内では実施中の臨床試験データを利用
 - 傾向スコアマッチングを用いて解析に組み入れる外部対照データを選択
 - ベイズ流手法 (Commensurate prior) を用いて外部対照データの情報を動的に借用
- 主要評価項目 : 無増悪生存期間 (Progression Free Survival; PFS)
 - ハザード比0.6, 有意水準5%, 検出力80%のもとで症例数を算出
 - PFSイベントが80%集まった時点で1回の中間解析を実施予定
- 重要な副次評価項目 : OS
 - ワイブル分布を仮定したベイズ流パラメトリックモデルで解析
 - ハザード比の事後平均及び95%信用区間に基づき効果を推測



試験デザイン背景 1/3

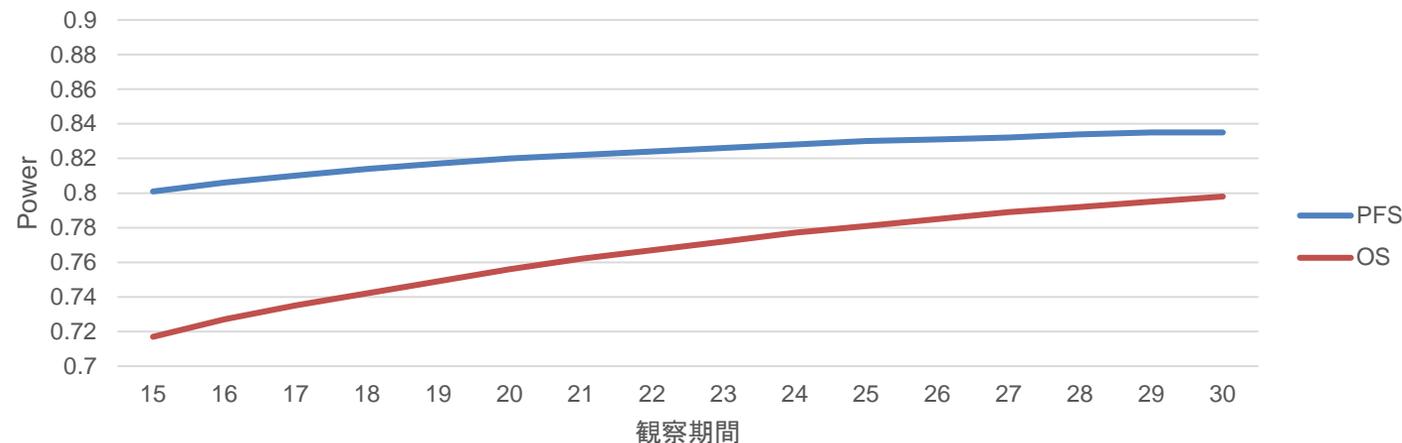
- 抗がん剤開発における評価項目
 - 標準的な評価項目は生存に関する指標でありOSがその代表である
 - 非常に予後のよい集団を対象とする場合には、OSを主要評価項目とすると多くの患者登録や長期の追跡が必要となること等から、PFSや無病生存期間等を主要評価項目とする場合がある

評価項目	Pros	Cons
OS	<ul style="list-style-type: none"> ・測定が容易で正確である ・一般的に客観的かつ定量的な評価に基づく 	<ul style="list-style-type: none"> ・対照群の治療が変更されたり、その後の治療が影響する可能性がある ・長期間のフォローアップが必要とされる ・癌以外の死亡も含まれる
PFS	<ul style="list-style-type: none"> ・OSの評価と比較して一般的に早期に評価でき、サンプルサイズも小さくて済む ・安定した病態の測定も含まれる ・一般的に客観的かつ定量的な評価に基づく 	<ul style="list-style-type: none"> ・オープンラベル試験では評価バイアスの可能性がある ・定義が試験ごとに異なる、頻繁な放射線検査やその他の評価が必要治療群間で評価のタイミングをバランスよく設定することが重要 ・必ずしもOSと関連しない場合がある

一方で、DLBCL患者の10年生存率は約50%であり、一次治療を対象にした試験では、OSの中央値を観測するまでに8年以上かかる場合があり、OSを主要評価項目とした第III相試験を完了するには、登録期間によっては約10年かかる可能性がある。

試験デザイン背景 2/3

- 臨床的有用性を直接反映するOSの結果を添付文書に記載される効能・効果に含めるために、重要な副次評価項目であるOSについても検証されることができることが望ましい
 - 一方で、OSの検出力を保持するためには、さらなる症例数の追加及び長期の観察が必要
- 例：PFSの症例数設定のもとでのOSの検出力の挙動
 - Time to Eventデータではイベントの発生数に基づき検出力が変化し、イベントの発生数は観察する期間の長さに基づいて変化
 - PFS: 試験治療の中央値=10month, 標準治療の中央値=6month (ハザード比:0.60)
登録期間=15month, 観察期間=15month, alpha=0.05, 検出力=80%
 - OS: 試験治療の中央値=16.6month, 標準治療の中央値=10month (ハザード比:0.60)
登録期間=15month, 観察期間=15month~30month, alpha=0.05



ハザード比(群間差)が同じ場合でも観察期間によって検出力が変化するため、PFSと同時期にOSを評価するためには、OSのイベント数の増やす追加が必要がある

試験デザイン背景 3/3^[1]

- 重要な副次評価項目であるOSのイベント数の確保

- 症例数の追加
- 観察期間の延長
- 試験外データの利用

- 本事例では、対照群のOSのデータとして同一社内で行われている臨床試験データが存在

FDAやNCIが収集データ数やモニタリングの負担を減らす取り組みの一環で、OSの長期追跡データを集める負担軽減を検討
<https://www.swog.org/news-events/news/2023/04/12/s2302-pragmatica-lung-opens-enrollment>

- 外部対照データの利用可能性

- 外部対照データ: Pocockの基準^[5]に基づき利用できることを判断
- 外部対照データの利用方針

- 傾向スコアマッチングを利用して、外部対照データの選択に伴い発生するバイアスの程度を調整
 ※当初はワイブル回帰モデルの説明変数に傾向スコアを含めた解析を予定していたが、FDAより得られる結果の解釈可能性に関する指摘を受け、傾向スコアマッチングに切り替え
- ベイズ流手法 (Commensurate prior) を利用して、外部対照データの利用に伴い発生するバイアスの程度を調整

傾向スコアマッチングにも様々な手法が存在
 Yuan J, et al. J Biopharm Stat 2019
 Lin J, et al. Pharm Stat 2018.

ワイブル分布モデルとCommensurate prior

• ワイブル分布モデル

- 外部対照データにおける被験者*i*のOS: $y_{1i} \sim Weibull(\gamma, \exp(\alpha_1))$
- 新規試験データにおけるの被験者*i*のOS: $y_{2i} \sim Weibull(\gamma, \exp(\alpha_2 + \beta z_i))$
 - $\alpha_1, \alpha_2, \beta, \gamma$: ワイブル分布のモデルパラメータ
 - $z_i = 0$: 新規試験の対照群, $z_i = 1$: 新規試験の試験群

• 各パラメータ ($\alpha_1, \alpha_2, \beta, \gamma$) の事後分布 $P(\alpha_1, \gamma | Y_1), P(\alpha_2, \beta, \gamma | Y_2)$

- $P(\alpha_1, \gamma | Y_1) \propto L(\alpha_1, \gamma | Y_1) P(\alpha_1, \gamma)$
- $P(\alpha_2, \beta, \gamma | Y_2) \propto L(\alpha_2, \beta, \gamma | Y_2) P(\alpha_2, \beta, \gamma)$
 - $P(\alpha_1, \gamma) = P(\alpha_1)P(\gamma), P(\alpha_2, \beta, \gamma) = P(\alpha_2)P(\beta)P(\gamma)$ は各パラメータの事前分布

$P(\tau^2)$ に設定する分布に応じて, 外部対照データを利用する程度が変化するため設定には慎重な議論が必要

• Commensurate prior

- 外部データが1個存在する状況を想定して, 新規試験の対照群と外部対照データの乖離度に応じて借用程度を調整する事前分布 ($P(\alpha_2 | \alpha_1, \tau^2)$)
- $P(\alpha_1, \alpha_2, \beta, \gamma | Y_1, Y_2) \propto L(\alpha_1, \gamma | Y_1) L(\alpha_2, \beta, \gamma | Y_2) P(\alpha_2 | \alpha_1, \tau^2) P(\alpha_1) P(\beta) P(\gamma) P(\tau^2)$
 - α_2 の事前分布として平均 α_1 , 分散 τ^2 の正規分布 $N(\alpha_1, \tau^2)$ を設定し, $P(\alpha_2 | \alpha_1, \tau^2)$ と定義
 - パラメータ τ は α_2 と α_1 の違いの程度を反映するパラメータであり, パラメータ τ が小さいほど α_2 と α_1 の違いは小さく, より多くの情報を外部対照データから利用する

CIDのポイント (CID pilot meeting program)^[2]

- 重要な副次評価項目であるOSの比較に外部対照データを利用
 - PFSの主解析時点にてOSを早期に解析することに伴うOSのイベント数不足を補うことが期待される
 - 傾向スコアマッチングにより比較に用いる外部対照データ(集団)を選択
- 外部データの借用にベイズ流手法(Commensurate prior)を利用
 - 傾向スコアマッチングのみではなく, 新規試験と外部対照データのアウトカムの共通性に応じて, 外部対照データの利用程度を調整
- OSの解析にベイズ流パラメトリックモデルを利用
 - Commensurate priorの利用に伴う, OSがワイブル分布に従うと仮定したパラメトリックモデルを用いてベイズ流推定を計画

FDAの懸念事項[2]

- 外部データのエンドポイントに対する懸念
 - 新規試験の実施時期と外部対照データが収集された時期の違いに伴う集団の違い(治療状況, Covid-19の発生状況など)
- 解析方法における懸念
 - ワイブル分布を仮定したパラメトリックモデルの当該疾患領域データへの適合性
 - Commensurate priorの利用の適切性, 解釈可能性
 - 傾向スコアモデルの利用の適切性, 解釈可能性
 - 仮定するモデルからの逸脱に伴う結果の頑健性
- 動作特性に対する懸念
 - 想定と異なるOS分布の影響
 - 未測定の交絡因子の有無
 - 傾向スコアマッチングで利用するモデルの形状の違いによる影響

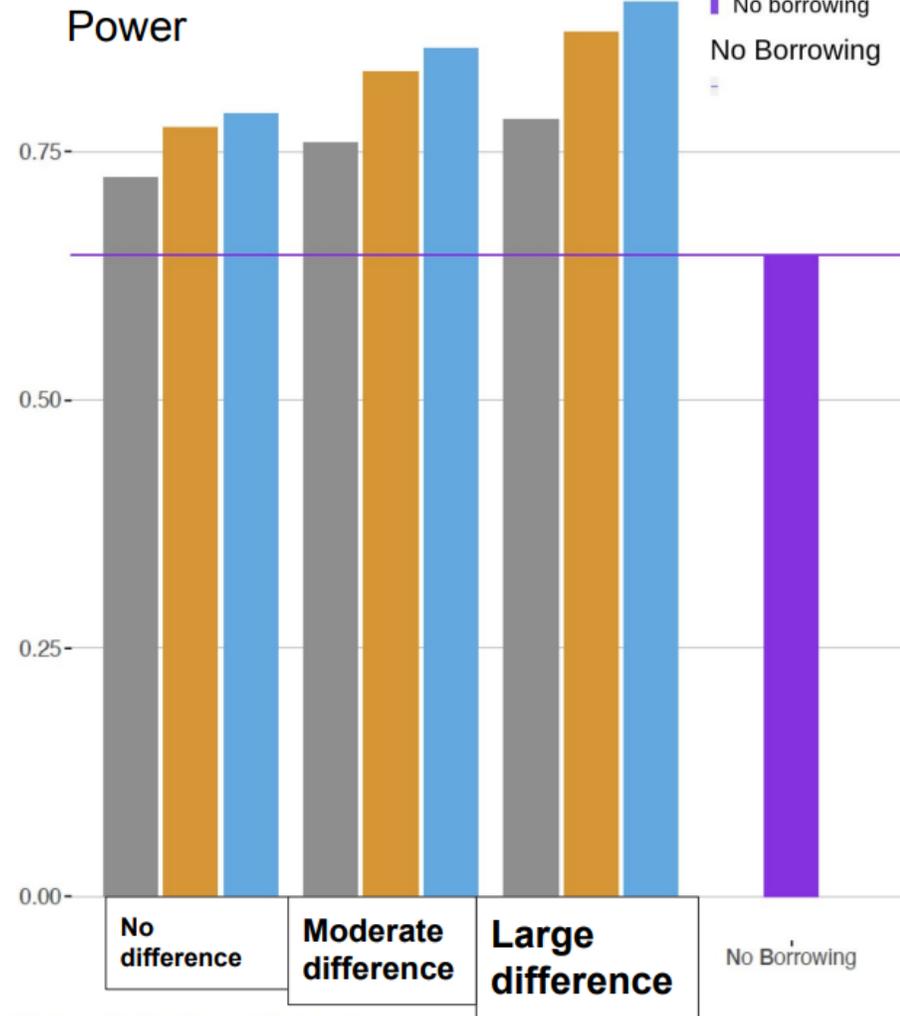
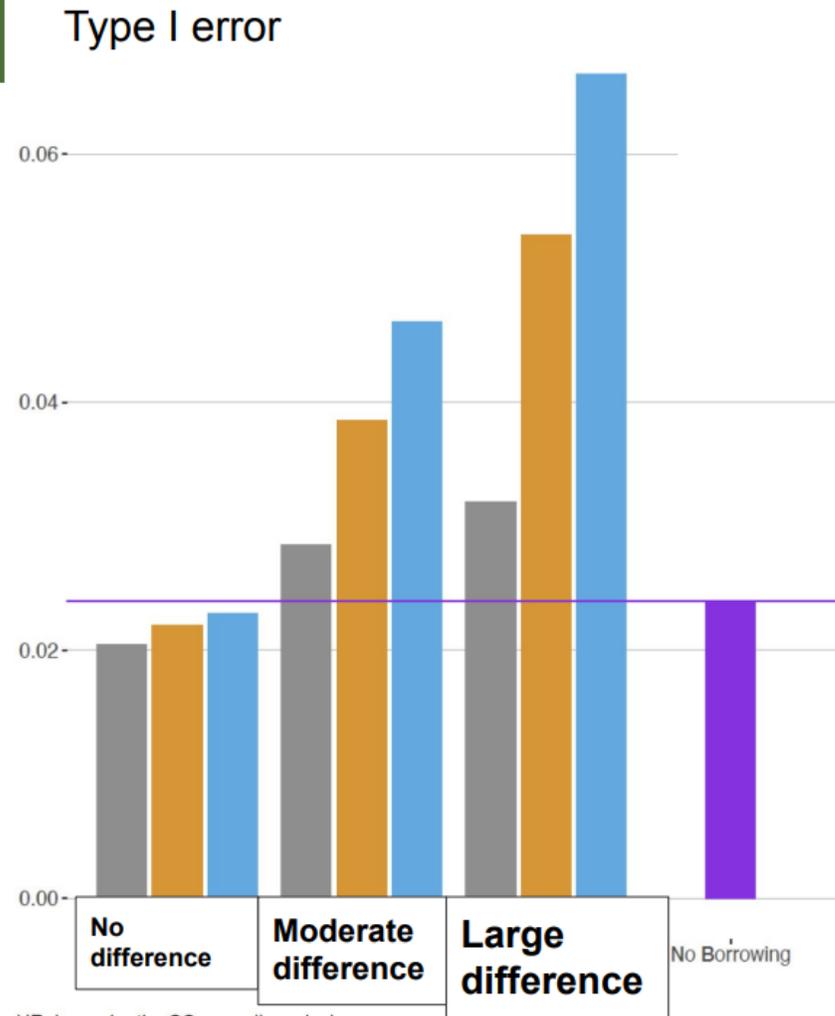
シミュレーション実験^[1]

- シミュレーション実験では，外部対照データの借用方法に応じた第一種の過誤確率，検出力を評価
 - 外部対照データを利用しない場合
 - マッチングされた外部対照データを動的利用するために，Commensurate priorを利用した場合
 - τ^2 の超事前分布 $P(\tau^2)$ に対して，以下の分布を利用
 - ・試験対象群と外部対照データの違いの程度が大きいと仮定した設定：半コーシー分布
 - ・試験対象群と外部対照データの違いの程度が小さいと仮定した設定：ガンマ分布
 - マッチングされたすべての外部対照データを利用する場合
- 加えて，新規試験の対照群と外部対照データの違いの程度に対する第一種の過誤確率，検出力への影響を評価
 - ①新規試験の対照群と外部対照データの患者背景の違い(OSの挙動の違い)
 - ②交絡因子の測定の有無
- ③シミュレーション実験結果
 - 患者背景の違いの程度に応じた各シナリオにおける第一種の過誤確率の平均値，重み付き平均値，最大値を集計し，第一種の過誤確率の動作特性に基づき外部対照データの借用方法を決定

シミュレーション実験結果①

[1] [Jiawen Zhu and Herb Pang](#)より抜粋

Simulation results highlights



シミュレーション実験結果②

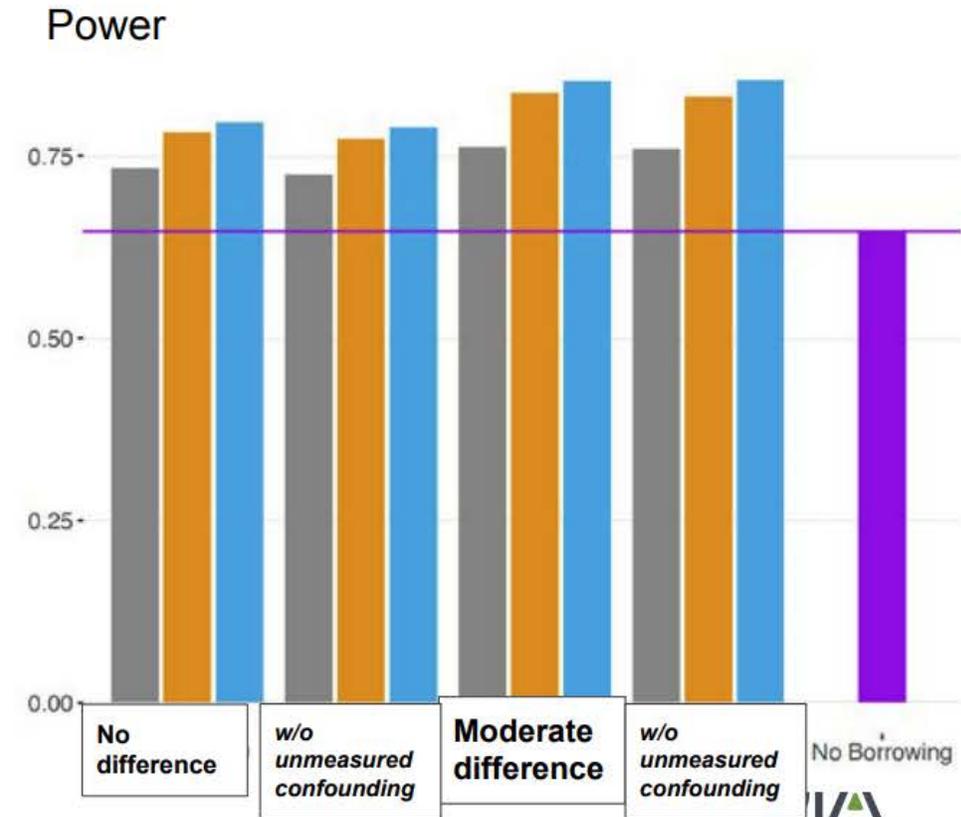
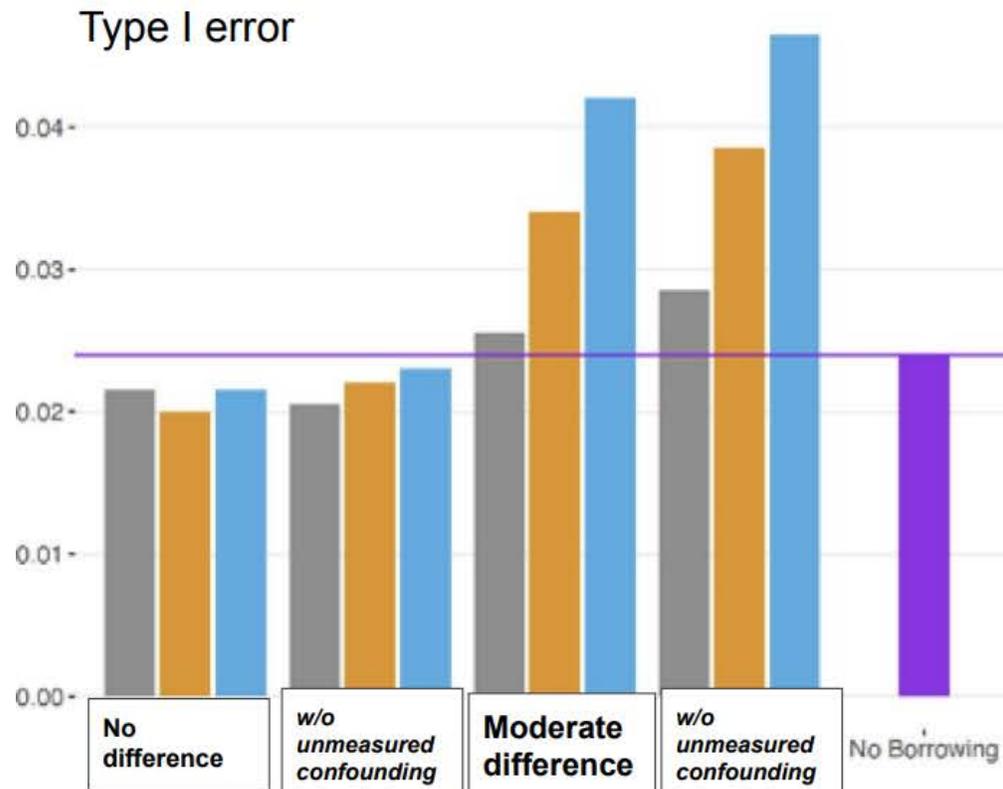
[1] [Jiawen Zhu and Herb Pang](#)より抜粋

Simulation results highlights

Method

Dynamic borrowing with Half-Cauchy
Dynamic borrowing with Gamma
Full borrowing
No borrowing

No Borrowing



シミュレーション実験結果③

[1] [Jiawen Zhu and Herb Pang](#)より抜粋

Simulation results discussion

Summary Table to Compare Method Performance for Differences in Baseline Characteristics Investigations

Approaches		Average Error Rate	Weighted Type I Error Rate*	Max Type I Error Rate
No borrowing (only RCT data)		0.024	0.024	0.024
Dynamic borrowing (with external control)	Conservative prior	0.023	0.023	0.032
	Aggressive prior	0.028	0.026	0.054
Full borrowing (pooling two control arms)		0.033	0.029	0.067

RCT= randomized controlled trial

第一種の過誤確率の平均及び重み付き平均の値が名義水準(0.025)以下に抑えられ、かつ検出力の増加が期待される τ^2 の超事前分布 $P(\tau^2)$ に半コーシー分布を仮定した方法を採用

* Weighted Type I Error Rate is calculated based on the assumed probability on the various scenarios: The probability for "The same" is assumed to be 62.5%, "moderate", 20%, "large" 5%, "moderate reverse" 10%, and "large reverse" 2.5%.

本事例におけるディスカッションポイント

- ① シミュレーション実験に基づき、超事前分布 $P(\tau^2)$ に半コーシー分布を使用したベイズ流ワイブル回帰モデルにcommensurate priorを用いた方法が採択されたが、
第一種の過誤確率が名義水準を超える可能性のある方法を用いてOSを検証することは問題ないか？
- ② 重要な副次評価項目であるOSではベイズ流の解析が計画されているが、外部対照データを利用しなかった場合（頻度流の解析の場合）のOSの結果と差異が生じた場合に、どのようにベイズ流の結果を解釈すべきか？
- ③ ①②の論点も含め、本事例にてCommensurate priorを利用することの妥当性を説明するシミュレーション実験ではどのような条件のもとで、どのような指標を用いるべきか？
シミュレーション実験を実施するうえで、どのような記録が必要であるか？

ディスカッションポイント①

- 検証を目的とする場合、シミュレーション実験にてMax Type 1 error rateも含めて2.5%以下(名義水準以下)になるようにパラメータ設定すべきであるため、本手法の利用はふさわしくない可能性がある。
- Max Type 1 error rateも含めて2.5%以下(名義水準以下)であることが説明できれば、使用することは差し支えない。ただし、第一種の過誤確率の制御について、シミュレーション実験結果にてMax Type 1 error rateが2.5%以下であったとしても、適切に名義水準以下に制御された方法であるかの説明が困難であり、本手法の利用が困難になりうる可能性がある
 - ワイブル分布ではすべての状況をカバーできない等、シミュレーション実験にも限界がある。
- Max Type I error rateを名義水準以下に抑えようとするとう検出力が低下する可能性がある。
- Max Type I error rateが名義水準を超えてしまう事実よりも、RCTのみのデータを利用することよりも検出力が増加することに意義があるため、第一種の過誤確率の上昇程度を鑑みて検出力増加が明らかに期待される場合には、利用することも可能ではないか？
- 今回の事例では副次評価項目の評価のため、厳密に第一種の過誤確率を制御できていなくとも、利用は可能ではないか？
- OSを評価する位置づけによっても利用の是非が変化すると思われる。本事例では副次評価項目OSの検証の背景が添付文書へのラベリングを意識していることを前提にしており、OSも検証できると嬉しい程度であれば、利用可能性があると考えられる。
- 利用目的に応じて、本ベイズ流手法の受け入れ可能性が変化する。
 - 検証することを目的とする場合、原則に基づいて第一種の過誤確率は名義水準以下に制御されるべきである
 - 検証することが目的ではない場合、例えば本事例のようにイベント数不足に伴う結果解釈が困難である状況や、安全性情報の収集等を目的とする場合には、第一種の過誤確率が論点にならない可能性があるため、利用可能性は高いと思われる

ディスカッションポイント②

- 外部データが質的に違う場合には頻度流を採用すべきで、同等と考えられる場合にはベイズ流の結果を評価することで差し支えないと考える。ただし、その判断が困難な状況が発生することを踏まえて、事前にメーカーと規制当局が協議できることが望ましい
 - 本事例の場合、傾向スコアマッチングにて抽出された外部データを用いることから、アウトカムの違いが発生した原因を検討できるような解析が必要である。
- 仮説検定の結果が逆転した場合の解釈も難しい場合がある。
 - 頻度流の解析では有意差がなく、ベイズ流の解析で有意差があった場合など
→単に外部データの利用に伴い推定精度が向上して有意差があるという結果が導かれたのか？その他の要因は考えられるか？
 - ベイズ流の解析で有意差がなく、頻度流の解析では有意差があった場合は？結果に基づいて有意だった頻度流の結果を用いても問題無い？
- 試験計画時にも、解析結果のずれが生じることを考慮した検討が重要と思われる
 - 本事例のシミュレーション結果ではバイアスが報告されていないため、バイアスの評価は必須
 - 計画時の検討もそうだが、結果が出てからの考察も重要。可能であれば得られた結果がシミュレーション実験にて検討されていることが好ましい
- 新規試験データの対照群と外部対照データがLarge differenceのとき、そもそも申請に使うのか？
 - シミュレーション実験のシナリオに含まれているが、そもそも検討すべき対象なのか？
 - 本事例では、OSが勝たないと申請できないというよりはあわよくばラベルにいれたい、という意図。
おそらくLarge differenceの場合に、ベイズ流の結果を用いないことが想像される
- 第一種の過誤確率と同様に評価の利用目的によってベイズ流の結果の利用の是非が変化すると思われる。
 - 試験計画時には想定される状況を踏まえてどちらの結果を用いるのかを事前にすり合わせておくことが好ましい（有利な解析にならないように）
 - 結果が得られた際も、違いによらず手法間の結果を評価して、その評価の適切性を他の解析結果に基づき評価すべき

ディスカッションポイント③

- 第一種の過誤確率、検出力の検討に加えて、推定値のバイアスを評価するシミュレーション実験結果が必要である。
- シミュレーション実験を踏まえて、Commensurate priorがうまくいく・うまくいかない(外部対照データを統合しない)状況の整理しておくことが望ましい
- Commensurate priorを利用出来ると考えられるデータの範囲内のシミュレーションシナリオで十分か？
 - 外部対照データを統合しないような状況のシミュレーション実験結果を提示して、統合しないことの説明もあると好ましい。
 - 可能な限り、治験依頼者側に有利でないことを説明できるようなシミュレーション実験結果があると妥当性の判断がしやすくなる。
- ベイズであれば、信用区間などの指標も考えられる。事前分布の指定を変えた場合も出すべき。
- 本事例では、解析モデル(ワイブル分布)の利用妥当性を評価すべきか？
 - メーカー内で検討されるべき内容ではあるが、その妥当性を説明できれば、シミュレーション実験結果に反映する必要はないと考えられる。

参考文献

[1] Jiawen Zhu and Herb Pang. “Case Study of a Phase 3 Study Design with Hybrid Control in 1L DLBCL – FDA Complex Innovative Designs Pilot Program” <http://www.bayesianscientific.org/wp-content/uploads/2022/09/Bayesian-KOL-session.pptx.pdf>

[2] Food and Drug Administration “CID Case Study: External Control in Diffuse B-Cell Lymphoma” <https://www.fda.gov/media/155405/download?attachment>

- 日本血液学会, 造血器腫瘍診療ガイドライン2023年版. 2023年, 金原出版
- Shi Q, Schmitz N, Ou FS, Dixon JG, Cunningham D, Pfreundschuh M, Seymour JF, Jaeger U, Habermann TM, Haioun C, Tilly H, Ghesquieres H, Merli F, Ziepert M, Herbrecht R, Flament J, Fu T, Coiffier B, Flowers CR. Progression-Free Survival as a Surrogate End Point for Overall Survival in First-Line Diffuse Large B-Cell Lymphoma: An Individual Patient-Level Analysis of Multiple Randomized Trials (SEAL). J Clin Oncol. 2018 Sep 1;36(25):2593-2602.
- Pocock, S. J. The combination of randomized and historical controls in clinical trials. Journal of Chronic Diseases. 1976; 29:175–188.
- J. Normington, J. Zhu, F. Mattiello, S. Sarkar, B. Carlin. An efficient Bayesian platform trial design for borrowing adaptively from historical control data in lymphoma. Contemporary Clinical Trials. 2020; 89
- 野村 尚吾, 大東 智洋, 澤本 涼. 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. 2022-2023; 43 巻 1 号: p. 63-96
- 日本製薬工業協会. “Rare diseaseの治療効果の推測法 - CIDパイロットプログラムとその実例 –. 「Rare diseaseの臨床開発における治療効果推測の現状と課題」シンポジウム(2023) https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS_202302_cid.pdf
- 日本製薬工業協会. “Complex Innovative Trial Designの現状と事例の整理 - FDAガイダンス邦訳付き” 2024年4月 https://www.jpma.or.jp/information/evaluation/results/allotment/tcjmdm00000010pp-att/DS_202404_CID.pdf

慢性疼痛 (Phase II) :
マスタープロトコルで疾患横断的に情報借用した
ベイズ流階層モデルを用いた事例

Chronic Pain (慢性疼痛)について

- 治癒に要すると予測される時間を超えて持続する痛み、あるいは進行性の非がん性疾患に関連する痛み
 - 整形外科疾患や術後に遷延する痛み、帯状疱疹や糖尿病に関連する神経障害性疼痛等が該当する
- **Unmet Needsが高い疾患**
 - 3~6か月痛みが続く
 - 世界の人口の約20%が侵されている
 - 医師の診察の約15~20%を占める
- 疼痛タイプ別に化合物を1つずつを確かめていく方法では、治験で失敗した場合に**開発期間が延長**してしまう
- ある患者の声: meta-analysisではなく、ある薬が**同一のProtocolで評価**されていればよいのと思うことがある

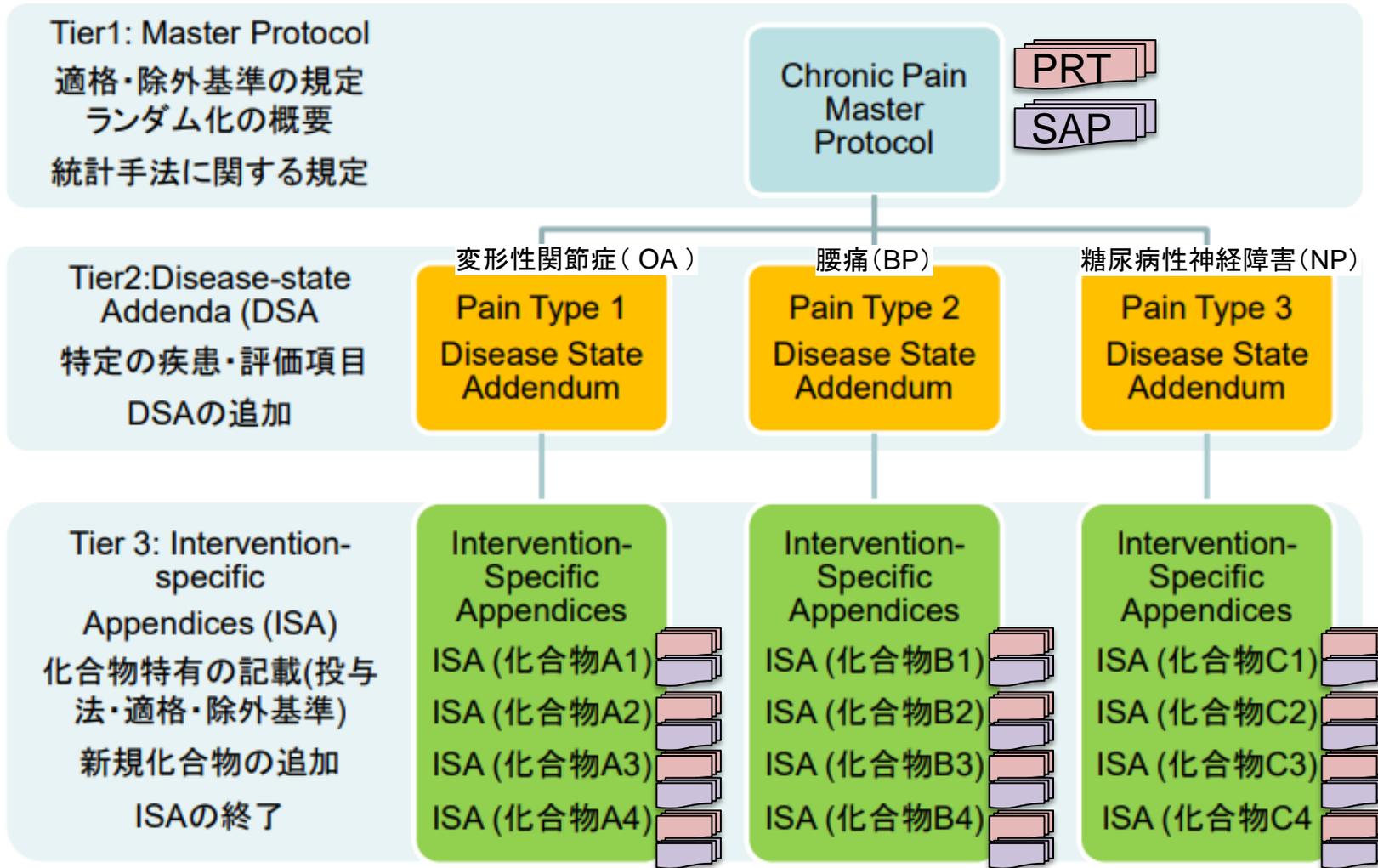
The Voice of the Patient

A series of reports from the U.S. Food and Drug Administration's (FDA's)
Patient-Focused Drug Development Initiative

Chronic Pain

Public Meeting: July 9, 2018
Report Date: March 2019

試験の概要



- Phase IIa (POC)
- Basket trial and Platform trial (期間は5~7年)
- 各sub-studyでの割付比:
試験薬:プラセボ=2:1
- 主要評価項目:
Numerical Rating Scale
- 主解析:
ベイズ流MMRM
- 各sub-studyでBayesian primary critical success factor (CSF)を規定

タイムライン

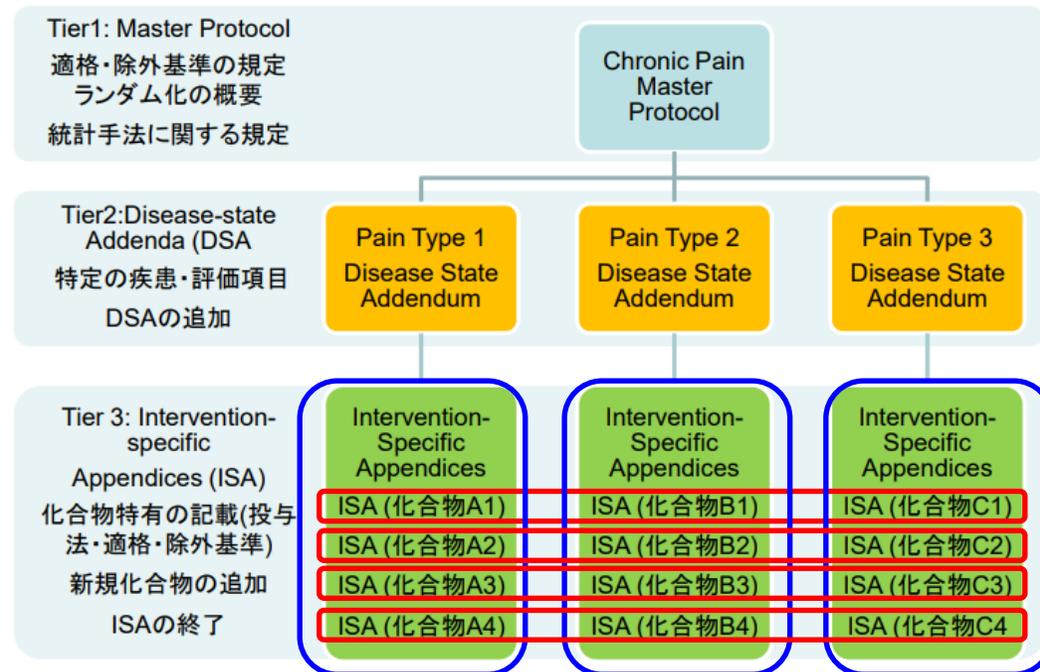
Year	2019		2020				2021				2022				2023			
Month	7-9	10-12	1-3	4-6	7-9	10-12	1-3	4-6	7-9	10-12	1-3	4-6	7-9	10-12	1-3	4-6	7-9	10-12
Lilly press release																		
Master Protocol																		
OA01																		
BP01																		
NP01																		
OA03																		
BP03																		
NP03																		
OA02																		
BP02																		
NP02																		
OA05																		
BP05																		
NP05																		

Clinicaltrial.govの情報を基に作成

- 異なる疾患(DSA)に対する同じ化合物(ISA)の試験を同時期に実施

FDAが革新的と判断した点

- 同一の疾患 (DSA) で実施された、他の化合物 (ISA) の **プラセボ群情報** を借用する
- 異なる疾患 (DSA) で実施された、同じ化合物 (ISA) の **治療効果の情報** を借用する



FDAのCID meeting programのケーススタディ資料から転載

本デザインで考慮すべき点

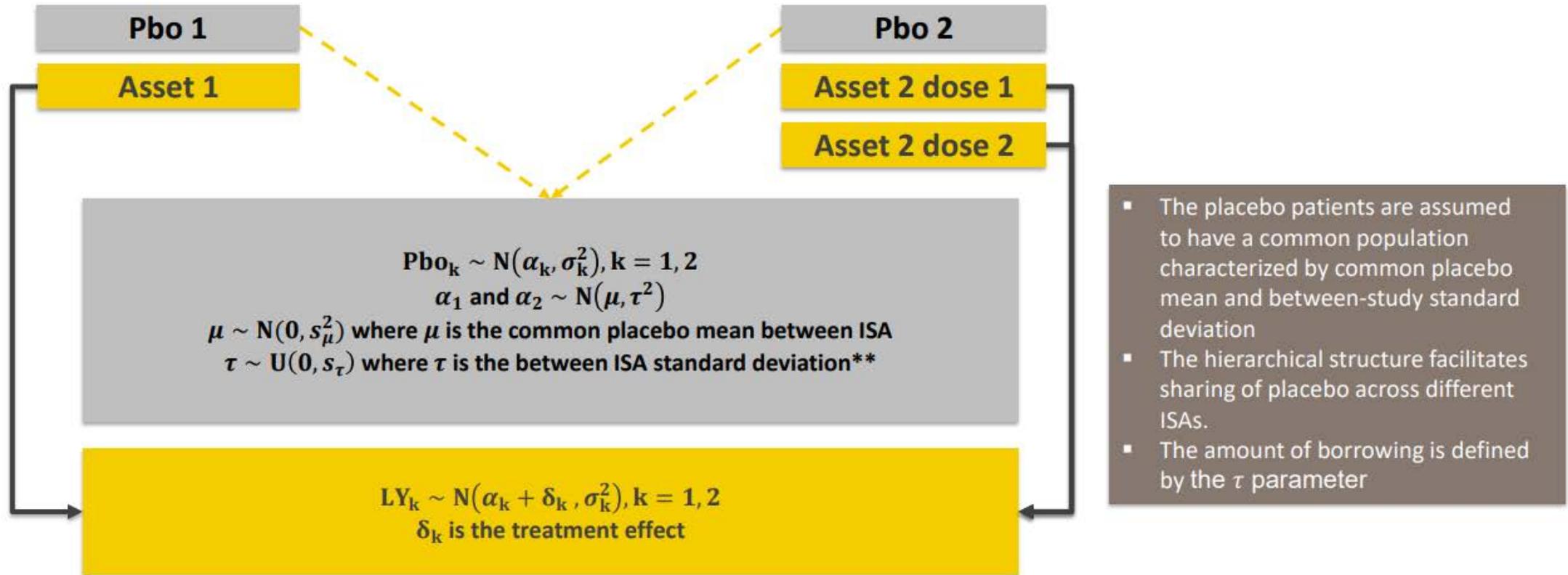
- 借用データの影響はどのようなものがあるか
- 借用データに対する重み付けの手法の違いは何か
- プラセボ効果が時間的に変化した場合、何が起こるか
- より類似性が高いと思われるデータをより強く借用するアプローチはあるか
 - 例えば、古い研究よりも最近の研究に重みを付けたり、同一タイプの疾患に重みをつける

シミュレーション

- 開発を継続すべき化合物の見極めという試験目的を勘案し、バイアス、治療効果の標準誤差、平均二乗誤差を評価
- 借用データとcurrentデータの類似性を考慮する、階層ベイズモデルおよびcommensurate priorを用いた

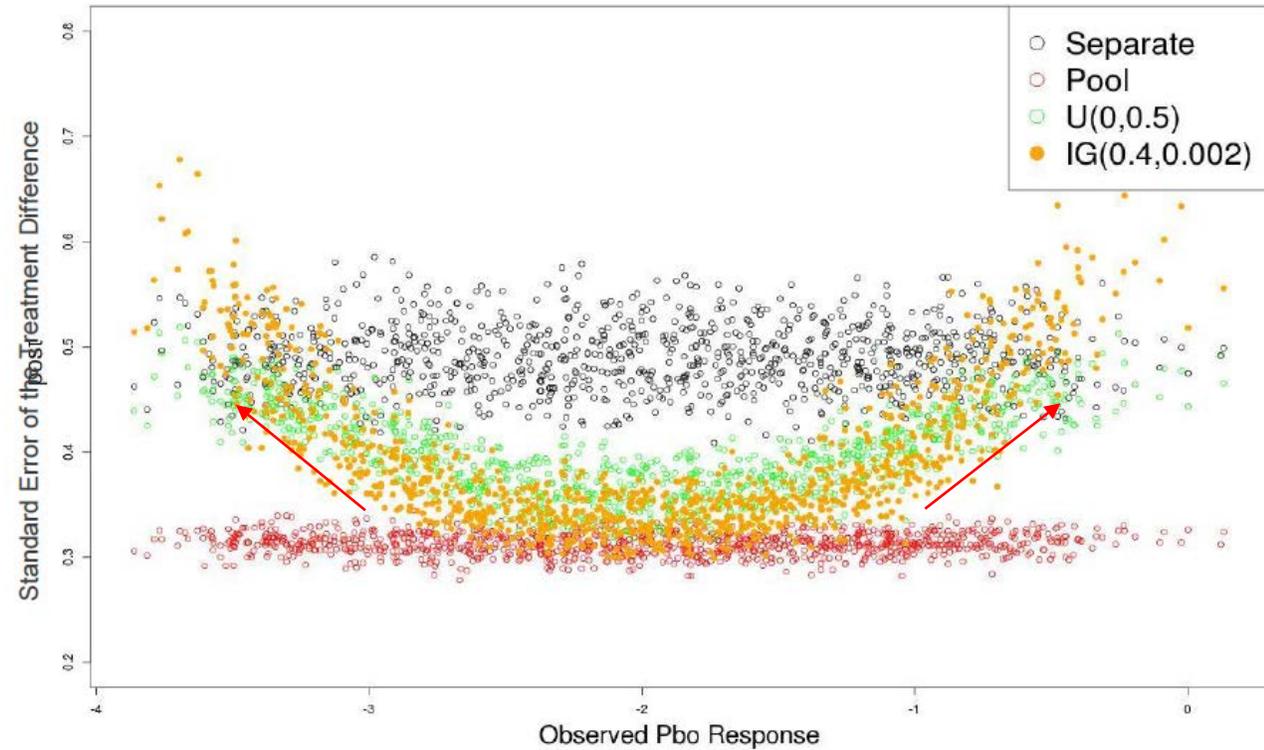
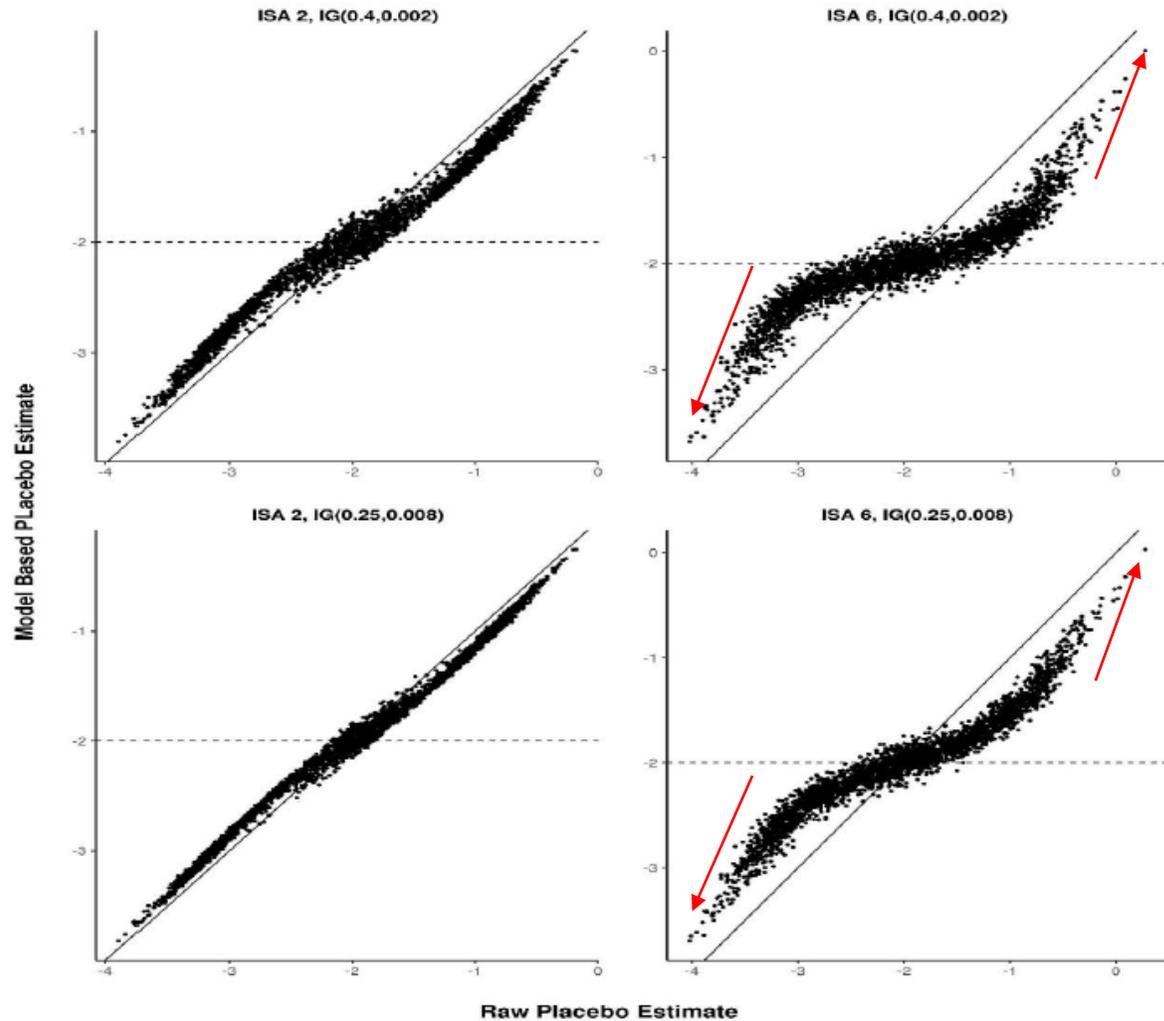
階層ベイズモデル: パラメータは同じ分布からのランダムサンプルと仮定
commensurate prior: パラメータ間に差が存在すると仮定し、その差に対して設定する事前分布

階層ベイズモデルによるプラセボ情報の借用



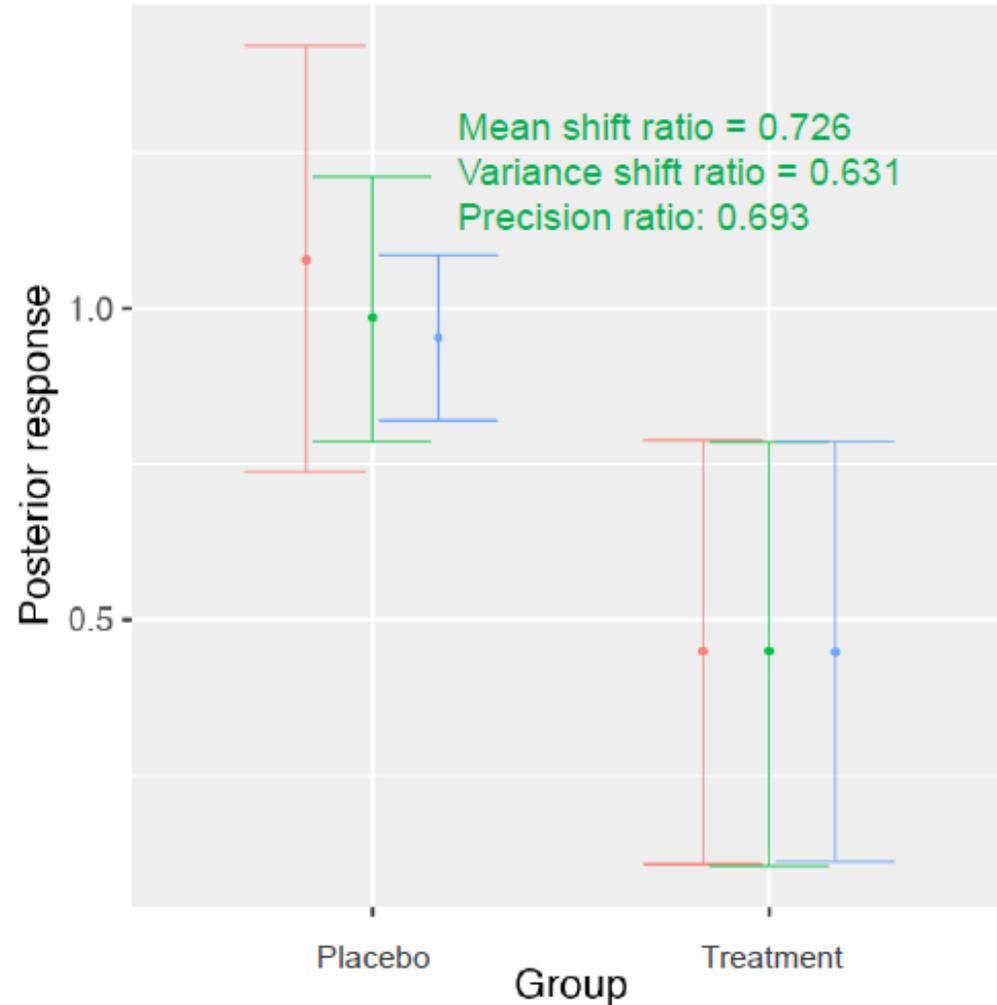
**An inverse gamma prior can also be used.

プラセボの推定値と治療効果のSEの挙動



FDAは、Dynamic borrowingにより“over-borrowing”を防止できる点を支持

借用の定量化



- Placebo borrowing is about 63%-73% according to the borrowing metrics.
- The mean shift ratio and variance shift ratio are visible in the graph.

Model

- Independent
- Hierarchical
- Pooled

Notation:

- $u_t^{(m)}$: posterior mean of the placebo mean.
- $v_t^{(m)}$: posterior variance of the placebo mean.
- $n_t^{(m)}$: number of observed data points.
- t : time point.
- m : 1 for the hierarchical model, 2 for the independent model, and 3 for the pooled model.

Mean shift ratio

$$\frac{u_t^{(1)} - u_t^{(2)}}{u_t^{(3)} - u_t^{(2)}}$$

Variance shift ratio

$$\frac{v_t^{(1)} - v_t^{(2)}}{v_t^{(3)} - v_t^{(2)}}$$

Precision ratio

$$\frac{\frac{1}{\tau_t^2}}{\frac{1}{\tau_t^2} + \frac{n_t}{34\sigma_t^2}}$$

FDAとの議論ポイント

- 交換可能性を仮定したモデルは妥当ではないかもしれない
 - 集団の背景/レスキュー療法の影響
 - 共通のプラセボ
 - ISA01:静注、ISA03:経口BID、ISA02:経口SID、ISA05:経口 (Clinicaltrial.gov より)
 - 評価する時期の影響 など
- 利用可能な過去データによる評価
 - 長い歴史として、プラセボ反応の一貫性がある場合、プラセボ群の情報借用を支持
 - 痛みの種類によって治療効果が異なる場合は、治療効果の情報借用は妥当ではない

データ借用に関するSAPの記載

- CPMP SAP

The ISA SAP will contain the exact specification of the model and parameters for the respective ISA. The specific decisions on the borrowing of placebo and treatment-effect information, including how much information to borrow and the specific methodology used to borrow, will be defined in each ISA. These decisions may vary over the course of the master protocol, as some borrowing strategies may perform better than others as more data are available to borrow.

借用の程度・方法は各ISA-SAPで規定

- OA02 SAP

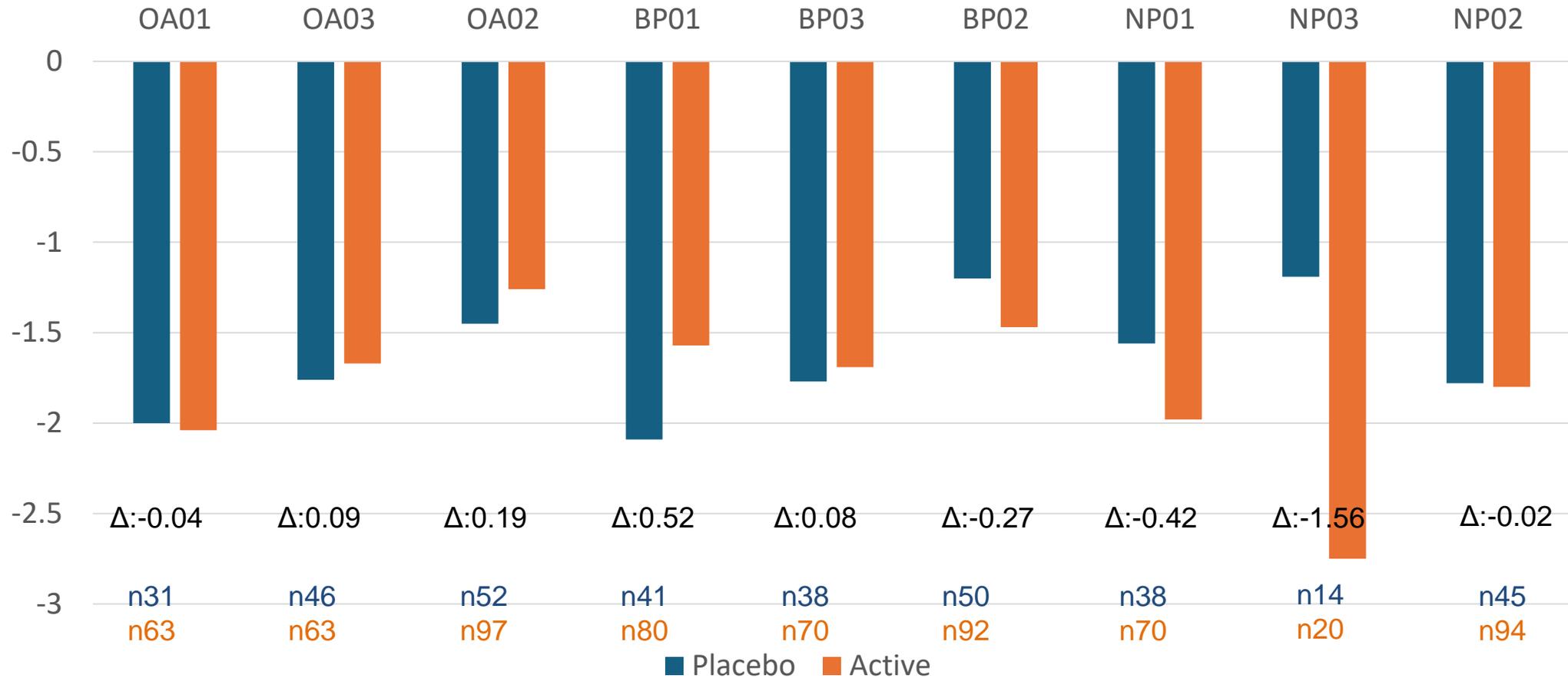
6.10.2. Additional Analyses of the Primary Outcome

Borrowing placebo information by pooling from Studies H0P-MC-OA01 and H0P-MC-OA03 for the evaluation of treatment effect on the mean change from baseline of API-NRS (and secondary outcomes of worst pain intensity as measured by the NRS, WOMAC® pain subscale, and VAS) will be performed following the description in CPMP SAP Version 5, Section 6.12.1.

試験のAdditionalな解析でプラセボ情報借用 (pooling)

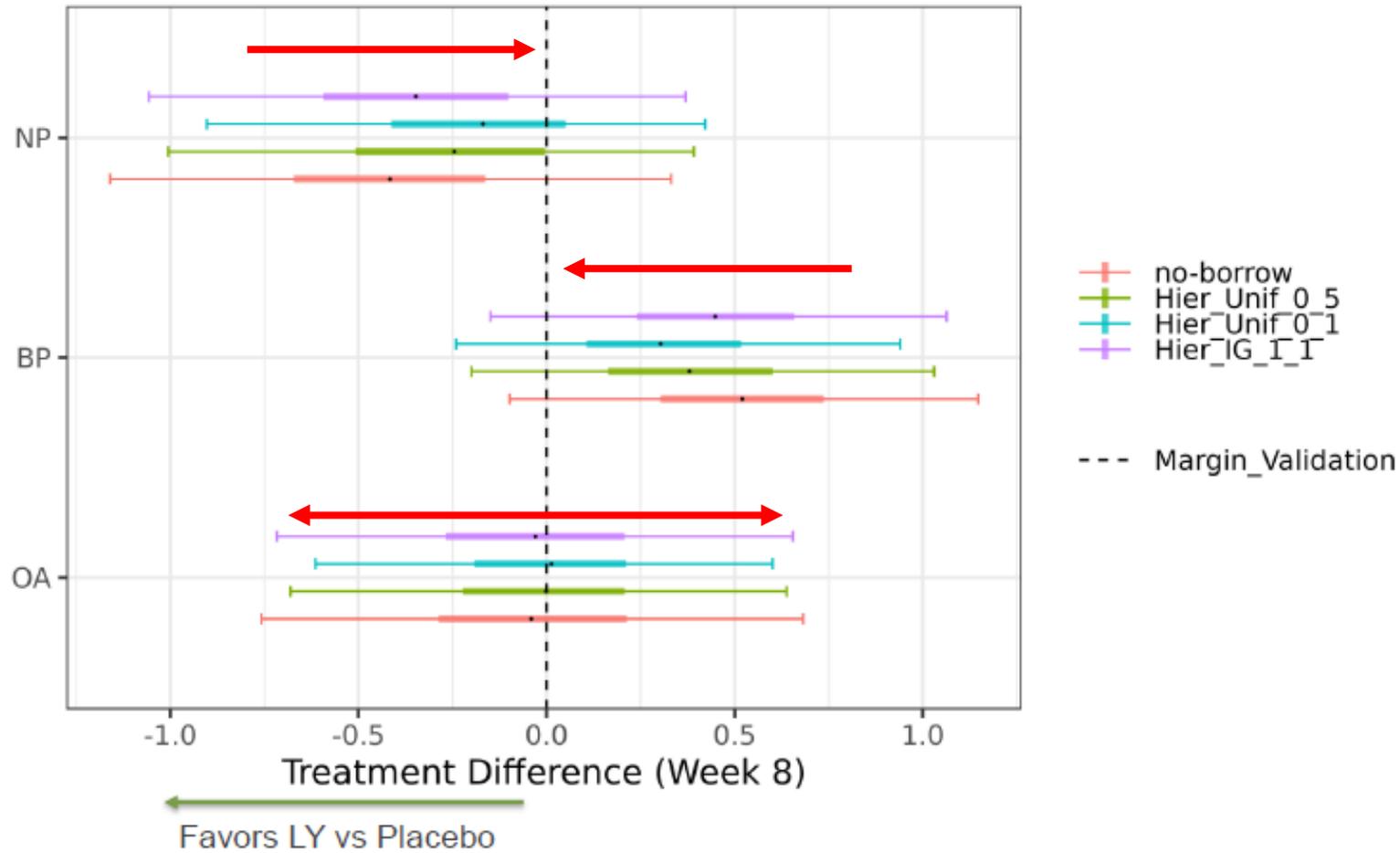
各sub-studyの結果(05を除く)

Average change from Baseline to Week 8 endpoint,
Average Pain Intensity –Numeric rating scale



同一DSAのプラセボ効果及び同一ISAの治療効果は類似していない？

OA01, BP01, NP01の治療効果の情報借用



Key Takeaways

1. Borrowing shrinks the estimated TE in the direction observed in the other disease types.
2. The neutral effect in OA does not change much after borrowing, but the shorter credible interval leads to more confidence in the treatment effect estimates.

本事例におけるディスカッションポイント

- ① 事例において、ベイズ流階層モデルが提案されたことは妥当と考えられるか？（キーワード：交換可能性、試験の位置づけなど）
- ② 事例において、情報借用する妥当性を説明するために必要なシミュレーションの設定条件は？（キーワード：評価項目、シナリオなど）
- ③ FDAから提示された点（スライドp6）以外で解析担当者として留意すべき点はあるか？

ディスカッションポイント①

- 対照疾患についてある程度症例数が確保できそうのため、ベイズを使う妥当性が低い
- 治験依頼者としては、コスト面でのメリットが大きい
 - 被験者数を抑えられる
 - 試験期間の短縮
 - オペレーションの観点
 - プラセボの情報の借用ができれば、実薬が投与される患者の割合が増える
 - P2のPOC試験も加味
- プラセボ群の情報借用
 - 患者背景の調整等で妥当性の説明が可能では
 - 慢性疼痛の特性により、時期の影響は少ないのでは
 - 標準治療が変わる可能性の検討は必要

ディスカッションポイント①

- 交換可能性の説明は難しい
 - ISAの経路が異なる点は懸念点
 - 過去試験の情報を活用したいが難しそう
 - プラセボ効果が出ないような疾患なら本デザインを適応できるのでは
 - 客観的な評価項目に対しては、ベイズ流階層モデルを使うことが適切であるかもしれない
 - 時間的な乖離は重み付け等で調整できないか
 - より発展的なモデルも活用したい(EXNEXモデル)
 - 治療効果の情報借用は、プラセボの情報借用と比べてハードルが高いのでは
 - 事前規定には限界がある
 - POCであるため、データが出た後に事後的に解析することが許容されるのでは
 - 疼痛スコアの差の解釈が難しい
 - 意思決定の際に参考情報とする

ディスカッションポイント②

- α エラーやpowerは見るべきか
 - POCであることを踏まえると、バイアスやMSEが優先されるのでは
 - P2bを実施するかどうかで変わりそう
 - SAP上、各試験での参考情報として確認はしていそう
- プラセボの有効性をばらつかせて、交換可能性が仮定できないシナリオで評価
- モデルを誤特定した際の挙動を確認

ディスカッションポイント③

- NRS以外の評価指標も情報借用検討すべきか
- POC試験における、意思決定方法の明確化
 - 事前に規定しSAPに記載すべき
 - 方針を説明するために、ある程度の基準は設けるべき
 - リスクベネフィットや競合情報なども含めて判断されることになるのでは
 - 事後確率による評価方法は複数あっても良いのでは
- データの秘匿性の観点でオペレーション上の制御が必要
 - 後から開始された試験が、それ以前の試験の結果を確認できる状況は好ましくない
- 交換可能性が説明できない場合でも対応できるような借用モデルも検討する
 - 試験背景を踏まえてモデルを選択
- シミュレーション実行に必要な時間を社内に共有する
 - 必要なリソースの確認

参考文献

- Food and Drug Administration “CID Case Study: Master Protocol to Study Chronic Pain”
<https://www.fda.gov/media/155403/download?attachment>
- DIA BSWG KOL Series
 - 資料: http://www.bayesianscientific.org/wp-content/uploads/2022/10/DIA-BSWG-KOL-CPMP-21OCT2022_final.pdf
 - Youtube: <https://www.youtube.com/watch?v=ex5BgsBKjWo>
- Bayesian Biostatistics Conference 2022 “Bayesian methods for placebo borrowing in master protocols”
<https://bayes-pharma.org/wp-content/uploads/2022/12/Landau-2022-10-13-20m.pdf>
- Clinical trial.gov

	PRT / SAP	NCT Number	Clinical trial.gov
Master Protocol	PRT addendum/ SAP	NCT05986292	Link
OA01	PRT / SAP	NCT04456686	Link
OA03	PRT/ SAP	NCT04627038	Link
OA02	PRT / SAP	NCT05080660	Link
OA05	NA / NA	NCT05620563	Link
BP01	PRT / SAP	NCT04529096	Link
BP03	PRT / SAP	NCT04874636	Link
BP02	PRT / SAP	NCT05086289	Link
BP05	NA / NA	NCT05630196	Link
NP01	PRT / SAP	NCT04476108	Link
NP03	PRT / SAP	NCT04707157	Link
NP02	PRT / SAP	NCT05177094	Link
NP05	NA / NA	NCT05620576	Link

参考文献 (Con't)

- 【日本製薬工業協会シンポジウム】 Rare diseaseの臨床開発における治療効果推測の現状と課題 セッション2: Rare diseaseの治療効果の推測法-CIDパイロットプログラムとその実例 -
https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS_202302_cid.pdf
- Eli Lilly and Company press release” Lilly's Pain Clinical Trial Protocol Selected for FDA Complex Innovative Trial Designs Pilot Meeting Program”
<https://investor.lilly.com/news-releases/news-release-details/lillys-pain-clinical-trial-protocol-selected-fda-complex>
- Beckman RA, Natanegara F, Singh P, Cooner F, Antonijevic Z, Liu Y, Mayer C, Price K, Tang R, Xia A, Apostolaros M, Granville C, Michaels D. Advancing innovative clinical trials to efficiently deliver medicines to patients. Nat Rev Drug Discov. 2022 Aug;21(8):543-544.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9834420/>
- 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(案)
<https://public-comment.e-gov.go.jp/servlet/PcmFileDownload?seqNo=0000273914>
- 日本製薬工業協会 (2024). “Complex Innovative Trial Designの現状と事例の整理 - FDAガイダンス邦訳付き”
https://www.jpma.or.jp/information/evaluation/results/allotment/tcjmdm00000010pp-att/DS_202404_CID.pdf

全身性エリテマトーデス (Phase IIb)
用量群間の情報を借用したベイズ流階層モデルを用いた事例

全身性エリテマトーデス (SLE)

- 自己免疫疾患であり，多臓器の関与と免疫学的異常が特徴で，症状は患者間での異質性があり，発熱，全身倦怠感，およびその他の全身症状，関節痛，発疹，光線過敏，脱毛，アフタ性潰瘍が含まれるため，適格性や治療に対する反応の評価が複雑
- 日本では難病に指定されており，SLE患者数は64,304人（令和3年度）と少なく，更に臨床試験では評価可能な疾患活動性があり，かつ重篤ではない患者を組み入れる必要がある
- 重症度に応じた副腎皮質ステロイドや免疫抑制薬による治療が行われるが，感染症をはじめとした副作用の問題がある
- SLEをはじめとする炎症性疾患では，単調な用量反応関係が観察されないことがある

Phase IIb試験のデザイン

- 目的: 3用量の被験薬の有効性をプラセボと比較評価し、後続の第III相試験のための最適な投与量を特定する
 - 結果によっては検証試験の位置付けにしたい
- 主要評価項目: 52週時点のSLE Responder Index 4 (SRI4)を達成した被験者の割合 (SRI4反応率)

・SELENA SLEDAI (全般的な疾患活動性スコア) の減少
 ・PGA (医師による全般評価) の悪化なし
 ・BILAG (各臓器の悪化) で悪化した臓器系がない

活動性SLEのある
18~75歳の患者を
320例をランダム化

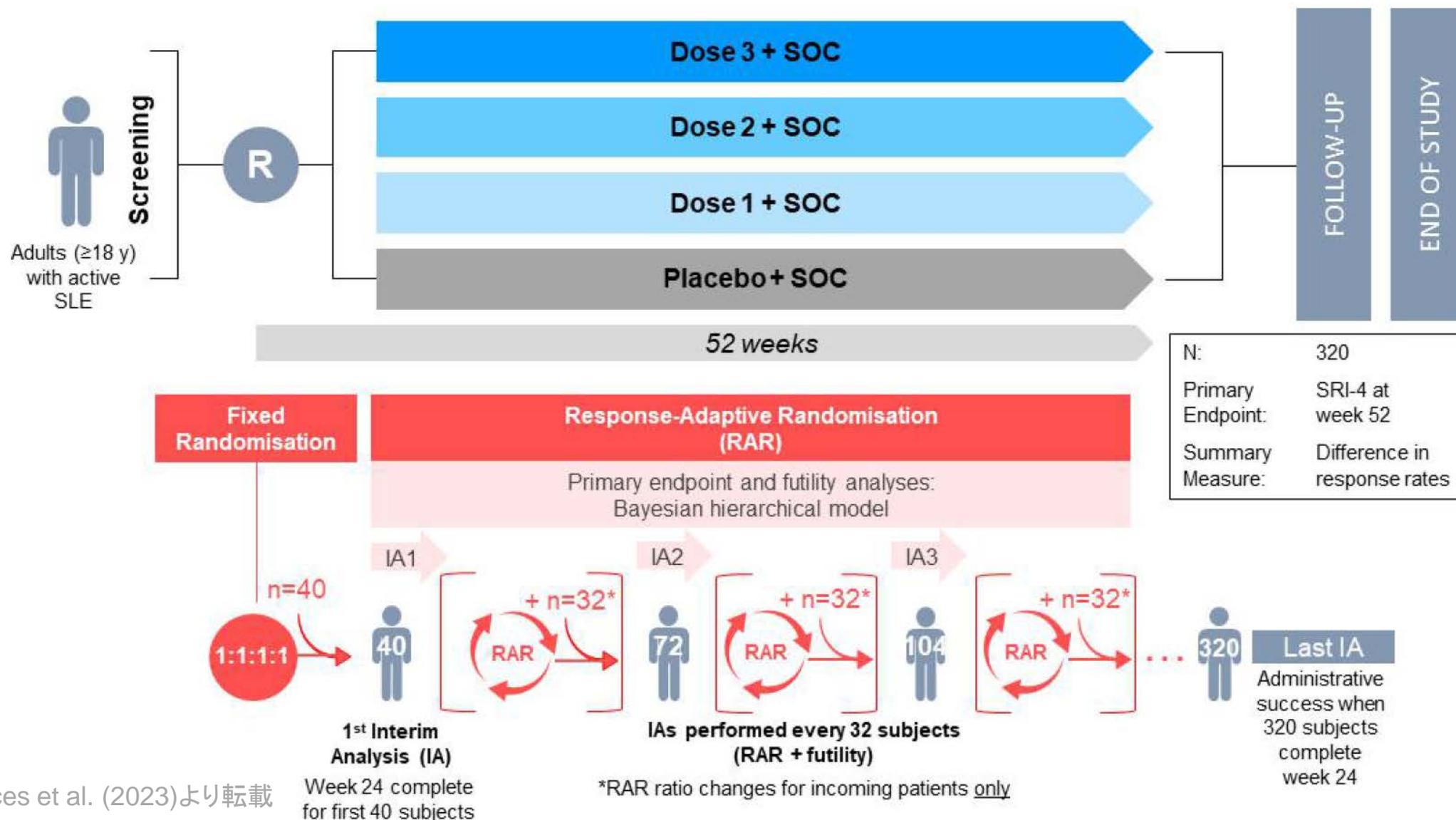
Randomization	プラセボ + 標準治療	安全性 フォローアップ
	実薬低用量 + 標準治療	
	実薬中用量 + 標準治療	
	実薬高用量 + 標準治療	

実薬: Efavaleukin Alfa (AMG 592)

革新的な特徴

- ベイズ流階層モデルによる用量群間の情報借用
 - ➡ 用量群間で反応が類似している場合に情報借用して検出力を上げたい
- 複数回の中間解析(初回は40例, 2回目以降は+32例が24週を完了した時点)
 - 反応適応的無作為化(初回以降の中間解析)
 - ➡ 反応が大きい用量群の割付比を増やすことで, 検出力を上げたい
無効な治療を受ける患者を減らしたい
 - 無益性の評価(2回目以降の中間解析)
 - ➡ 無効な治療を受ける患者を減らし, また, 他の臨床試験にリソースを譲りたい
 - Administrative success(最終の中間解析での第Ⅲ相試験の成功確率の予測)
 - ➡ 開発期間を短縮したい

革新的な特徴



ベイズ流階層モデルによる用量群間の情報借用

ベイズ流階層モデル:

$$X_d \sim \text{Binomial}(p_d, N_d)$$

$$\log\left(\frac{p_d}{1-p_d}\right) = \alpha_d \sim N(\alpha_{\text{treatment}}, \sigma^2)$$

$$d \in \{low, medium, high\}$$

用量群間の反応の類似度

X_d : 用量群 d のSRI4反応例数

p_d : 用量群 d のSRI4反応率

N_d : 用量群 d の例数

主解析の成功基準:

$$\Pr(p_d > p_{\text{placebo}} | \text{Data}) > \text{high value threshold}, \text{ for any } d \in \{low, medium, high\}$$

無益性の判断基準:

$$\max \Pr(p_d - p_{\text{placebo}} > \text{target treatment effect} | \text{Interim Data}) < \text{low value threshold}, d \in \{low, medium, high\}$$

反応適応的無作為化

- 初期の割付比は1:1:1:1であり, 最終を除く中間解析で52週時点のSRI4反応率に基づき, 各実薬群の割付比を変更可能(プラセボ群の割付比は固定)

$$Allocation_d \propto \Pr(p_d = \max_c p_c \mid \text{Interim Data}) \quad c, d \in \{low, medium, high\}$$

- 反応適応的無作為化でのSRI4反応率は非階層モデルに基づく

$$X_d \sim \text{Binomial}(p_d, N_d)$$

$$\log\left(\frac{p_d}{1-p_d}\right) = \alpha_d$$

$$d \in \{low, medium, high\}$$

試験デザインのPros/Cons

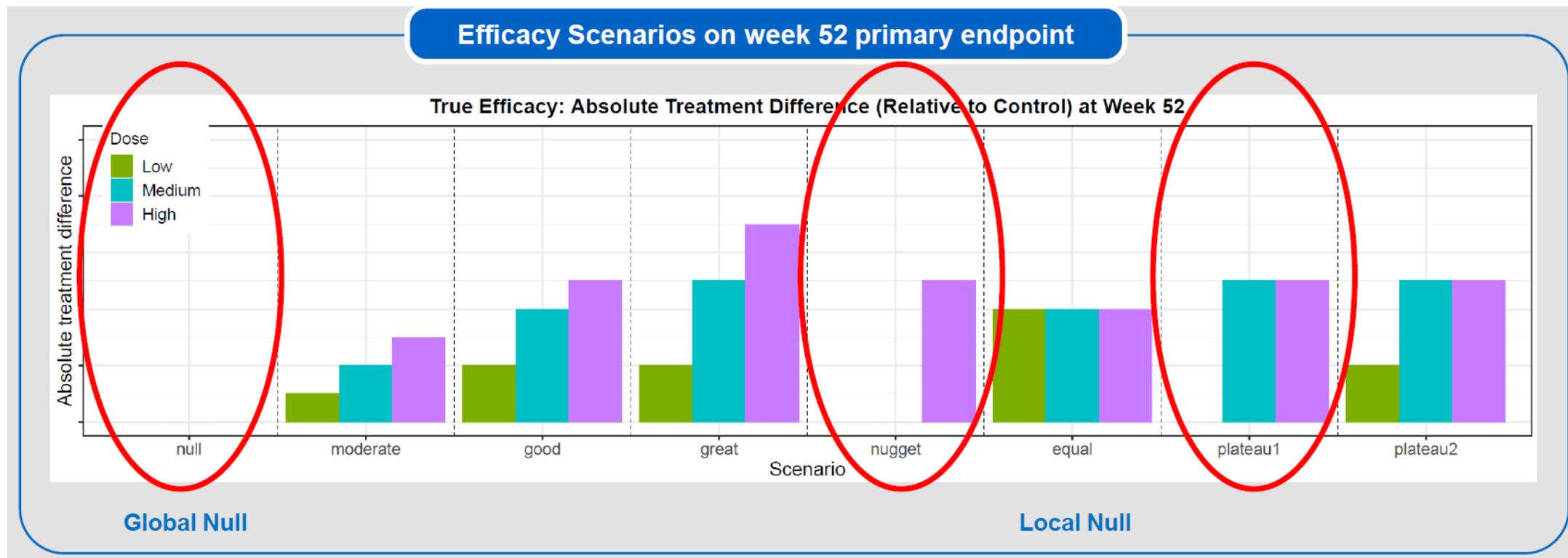
	Pros	Cons
ベイズ流階層モデル	<ul style="list-style-type: none"> 用量反応に関する仮定を置くことなく、用量群間の反応の類似性に基づいて、動的に情報借用できる 	<ul style="list-style-type: none"> 縮小(α_dが$\alpha_{treatment}$に引き寄せられる)により反応率の推定値にバイアスが生じる
反応適応的無作為化	<ul style="list-style-type: none"> より多くの患者を最も効果がある用量に割付ける傾向がある 最も効果がある用量とプラセボとの比較における検出力を増加させる 	<ul style="list-style-type: none"> 治療群間の患者背景が不均衡になる可能性がある 効果が低い用量群に割付けられる患者が少なくなり、用量反応関係の特定が難しくなる可能性がある
無益性の評価	<ul style="list-style-type: none"> 患者が有効ではない治療を受けるリスクを低減するとともに、試験が失敗した場合のコストを削減する 	<ul style="list-style-type: none"> 第2種の過誤確率の増加

CID pilot meeting programの主な論点

- 治療群間の患者背景の比較可能性に対する反応適応的無作為化の影響
- 真の用量反応関係の仮定およびそれ以外の試験結果に影響を及ぼす項目*のもっともらしい設定を組み合わせた場合における試験デザインの動作特性の適切性
- Arm-droppingアプローチと比較した反応適応的無作為化の性能
- 主解析において、競合する従来の方法と比較したベイズ流階層モデルの性能
- 中間解析結果の機密性および試験の完全性を維持するための手順

*プラセボ群の反応率, 登録スピード, 評価時点間の相関など

シミュレーションによる動作特性の評価



用量反応関係以外のパラメータ:

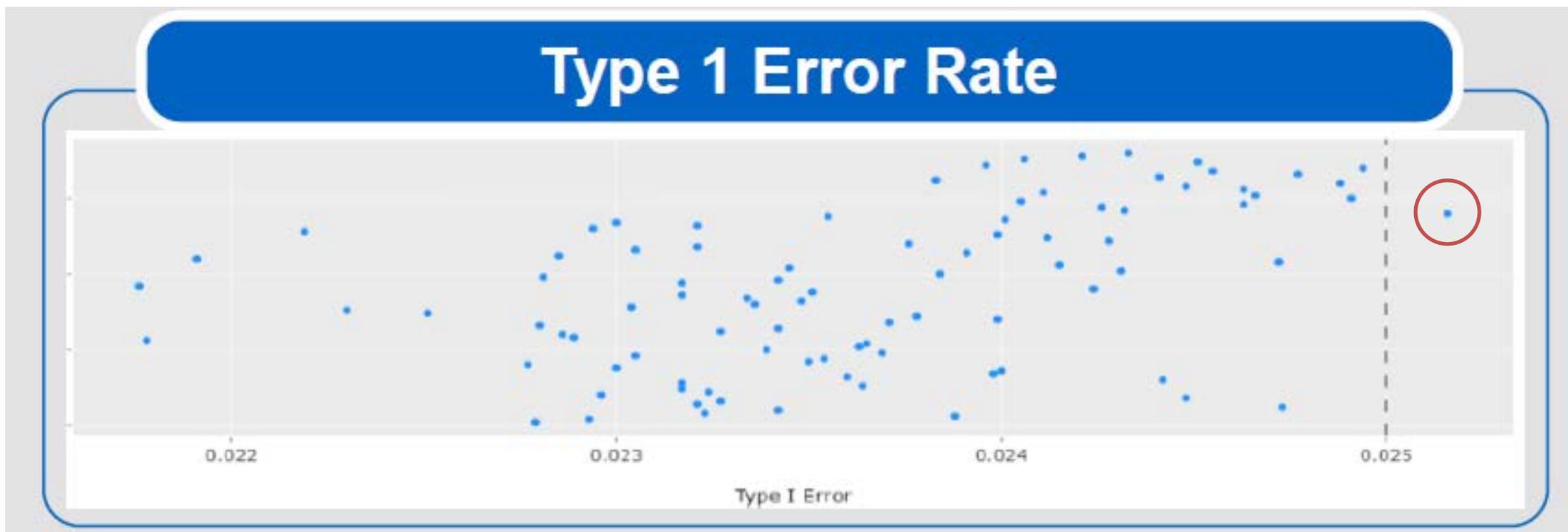
プラセボ群の反応率, 登録スピード, 評価時点間の相関

評価された重要な動作特性:

第1種の過誤確率, 検出力, 点推定値の信頼性など

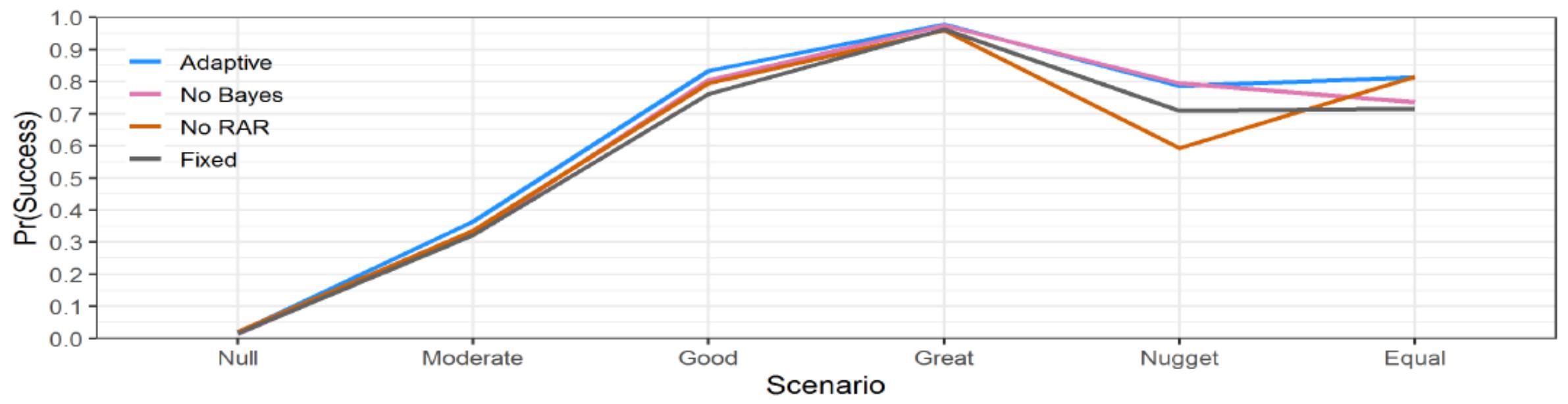
DIA Bayesian Scientific Working Group資料から転載

第1種の過誤確率



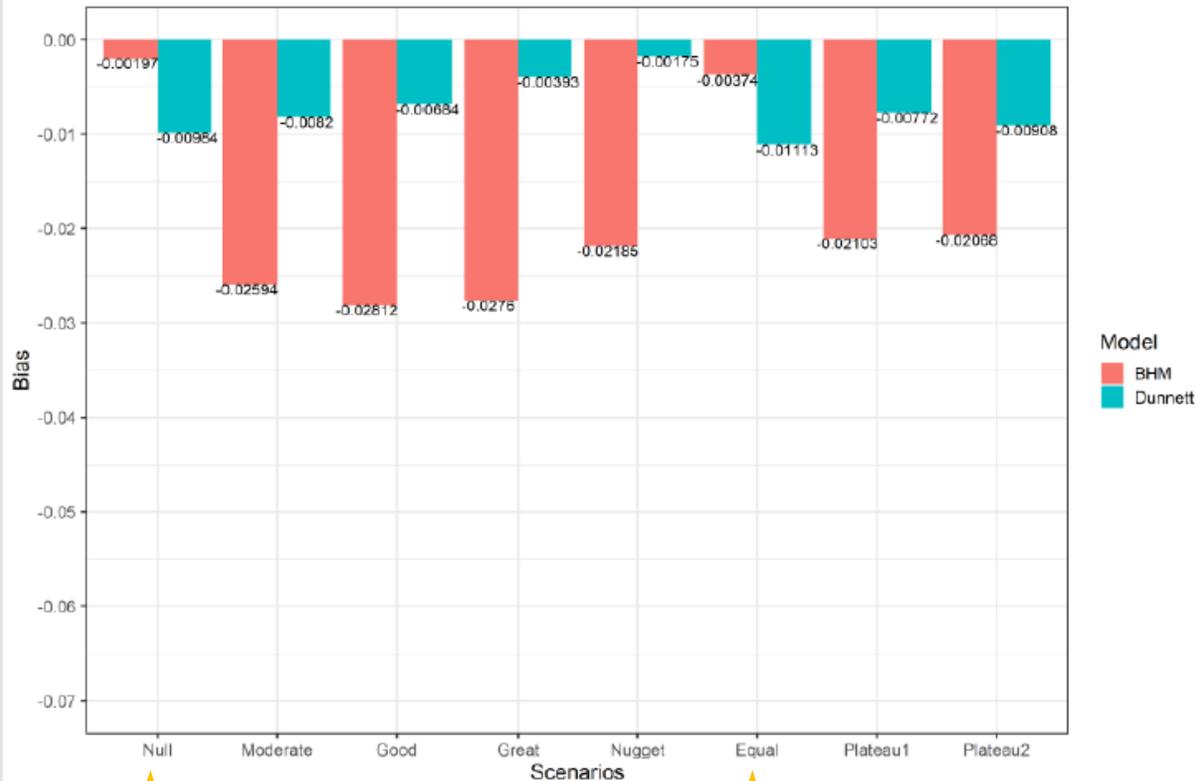
Global Null(どの用量群も有効ではないシナリオ)

検出力

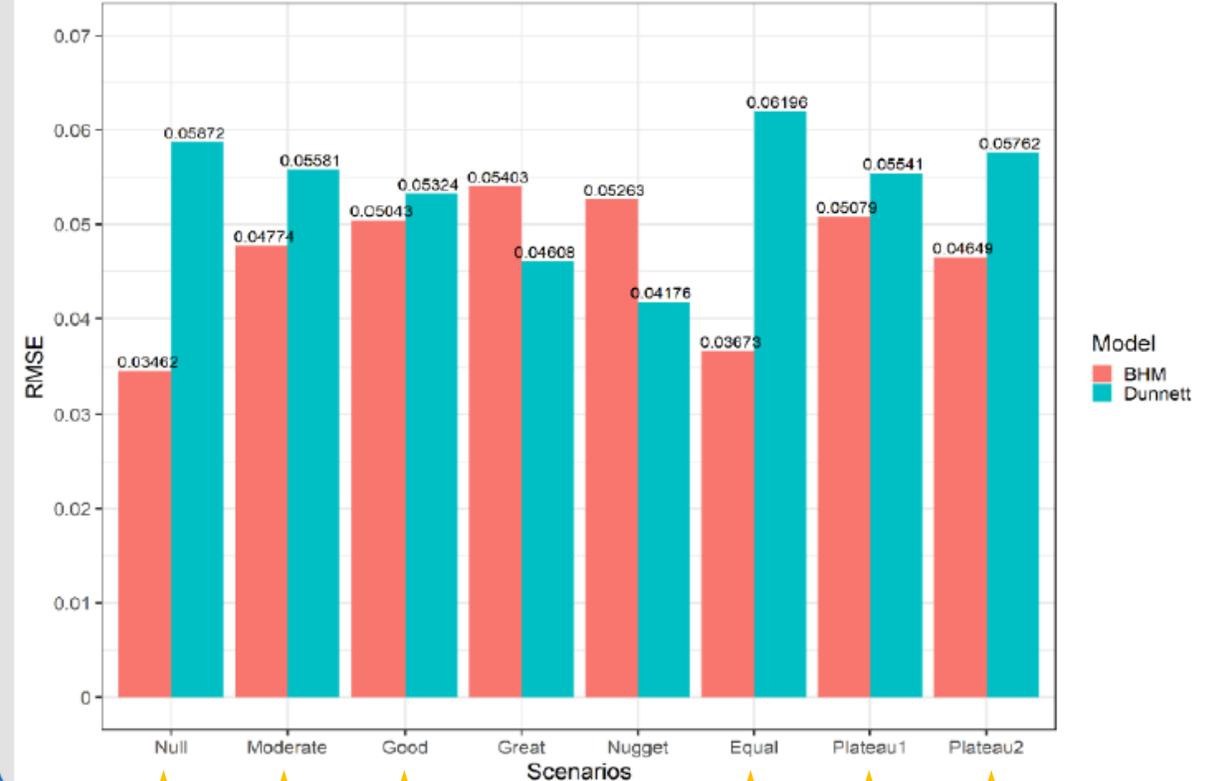


推定値の信頼性(バイアスおよびRMSE)

Bias



RMSE



★ : ベイズ流階層モデルがwin

本事例におけるディスカッションポイント

- ① 事例の状況において、ベイズ流階層モデルや反応適応的無作為化などが提案されたことは妥当と考えられるか？
- ② 事例のシミュレーション結果では、ベイズ流階層モデルのバイアスは大きい傾向、RMSEは小さい傾向だが、ベイズ流階層モデルは有用と考えられるか？
- ③ 事例において、試験デザインの動作特性の妥当性をシミュレーション結果で説明する際に押さえておくべきポイント(評価項目や考慮するパラメータなど)は何か？

ディスカッションポイント①

- 単調な用量反応関係が観測されないことがある疾患なのに、シミュレーションで有効性が単調に増加するシナリオのみ仮定することで良いか疑問に感じる
 - 精神疾患では用量を上げた場合に副作用が症状に悪影響を及ぼす可能性も考えられるが、がんなど他の領域では一般的には用量を上げた場合に有効性が下がることは仮定しないと考えられる
- 40例で反応適応的無作為化を行っているが、推定精度を確保するために、もっと先の間解析(例えば、60例以降)で行うことも考えられる
- 反応適応的無作為化では、安全性は考慮されないため、高用量の安全性が悪かった場合に問題が生じないか疑問に感じる
 - 有効性と安全性の両方を考慮した反応適応的無作為化を検討することも考えられる
- 用量反応関係の評価を目的としているのに、ベイズ流階層モデルによってバイアスが入ることが許容されるのか疑問に感じる

ディスカッションポイント①

- ベイズ流階層モデルは用量間の反応が類似した場合に有効であるが、反応適応的無作為化では用量反応関係がある場合に有効であり、用量反応に対して相反する2つの方法が用いられている
- 検証的な位置づけにもしたいとあるが、どのような条件が満たされたら検証的と判断するか疑問である
- 検証的試験と位置付けたいために320例の比較的大規模な試験となっているが、もっと少ない例数でシンプルなPh2を行うことも考えられる(実は効率的ではないかも知れない)
- 適用することは理解はできるが、妥当かどうか明確に判断しづらい
 - ベイズ流階層モデルと反応適応的無作為化を用いることに対して、用量反応関係の仮定が曖昧
 - 用量反応関係を評価したいのか、有効性を検証したいのか曖昧

ディスカッションポイント②

- 縮小によってバイアスが生じるため、用量反応関係がある場合に、ベイズ流階層モデルは相性が悪い
- バイアスだけでなく、Type I error rateが制御されていることも重視される
- バイアスとRMSEに加えて、エフェクティブサンプルサイズ(ESS)も評価することが有用と考えられる
 - どの用量から何例分の情報が借用されているか分からないと思われるが、推定精度がどれくらい向上しているかは分かると考えられる
- この事例では不明だが、用量の選択に使用した解析と有効性(治療効果)の判定に使用した解析が異なると解釈が難しくなる

ディスカッションポイント③

- Type I error rateが確認できるだけのシミュレーション回数が必要(例えば, 10万回)
 - Seed値を変えたときに結果が変わると困る
- シミュレーションプログラムの提出が必要となる可能性がある
- Rのパッケージのバージョンやソフトウェア(RとSAS)の違いによる影響がないか注意が必要
- スポンサーやCROでのシミュレーションのプログラムのQCをどのように行うのかも検討する必要がある
- 局外パラメータを複数パターン設定して, 最大のType I error rateを確認することが重要である
 - 最大のType I error rateを押さえようとする, 結局のところ頻度論ベースの例数に落ち着くことがある
- シミュレーションと実際の試験の運用(試験計画の変更に必要な期間など)が異なると問題が生じる可能性がある

参考文献

- 日本製薬工業協会. “Complex Innovative Trial Designの現状と事例の整理 - FDAガイダンス邦訳付き”
<https://www.jpma.or.jp/information/evaluation/results/allotment/tcjmdm0000010pp-att/DS_202404_CID.pdf>
- Food and Drug Administration “CID Case Study: A Study in Patients with Systemic Lupus Erythematosus”
<<https://www.fda.gov/media/155404/download?attachment>>
- DIA Bayesian Scientific Working Group KOL Series “Delivering Transformative Medicines to Patients by Leveraging Complex Innovative Designs AMG 592 SLE Case Study”
<http://www.bayesianscientific.org/wp-content/uploads/2022/08/CID_BSWG_AUG2022.pdf>
- Garces, Sandra, et al. "Improving resource utilisation in SLE drug development through innovative trial design." *Lupus Science & Medicine* 10.2 (2023): e000890.

テーマ全体でのディスカッションポイント

各事例でのディスカッションを踏まえて

- ① ベイズを利用した試験デザインを採用する背景としてどのような特徴が挙げられるか
- ② ベイズ流手法を利用する際の留意点をどう考えるべきか
 - デザイン特性、バイアス・第一種の過誤確率
 - シミュレーションで動作特性を評価することの是非/範囲
- ③ 今後ベイズを利用した試験デザインを計画・実施する際に産官学でどのように議論していくべきか

ディスカッションポイント①

- ベイズを利用した試験デザインを採用する背景としてどのような特徴が挙げられるか
 - 用量間で情報借用してサンプルサイズを抑える
 - 同一疾患や化合物間で情報借用してサンプルサイズを減らす
 - 最小限のサンプルサイズを確保して、試験期間を短くする
 - サンプルサイズを抑える目的でベイズを利用することは適切か → サンプルサイズを小さくするなど試験を効率的に行うことは、ベイズを利用する動機としてはありえるが、疾患や薬理的な背景からの裏付けが必要
 - 試験を効率的に行うことで患者さんのメリットになる可能性があることは良いと思われるが、ベイズを利用するための臨床的・薬理的な妥当性を説明する必要があるため、統計だけではなく開発全体で議論が必要

ディスカッションポイント②

- ベイズ流手法を利用する際の留意点(デザイン特性, バイアス・第一種の過誤確率など)
 - 階層ベイズではバイアスが入るが、どの程度であれば許容できると判断するのが難しい
 - ESSがどれくらいになるかの検討も重要
 - モデルの誤特定やパラメータ設定の影響にも注意する必要がある(実施側に有利にならないように)
 - シミュレーションに時間がかかるので、計画時に必要な工数も検討しておく必要がある
 - シミュレーションと実際の運用に乖離が生じる可能性にも注意する必要がある(例えば、試験デザインの変更にかかる期間など)
- シミュレーションで動作特性を評価することの是非/範囲
 - 第一種の過誤確率を制御することは重要だが、過度に抑えようとするベイズを利用する意義が薄れる可能性もある。実際外部データとの大きな乖離が示唆された場合には申請しない可能性があることを加味すると、全ての設定でシミュレーションを実施して、Worst Caseまで考慮して第一種の過誤確率を制御する必要があるのかは疑問
 - 規制当局の観点からは、全ての状況で第一種の過誤確率を制御してほしいが、臨床的にあり得ない状況まで確認する必要はないかもしれない。ただし実際の試験と乖離がでる可能性はある

ディスカッションポイント③

- 今後ベイズを利用した試験デザインを計画・実施する際に産官学でどのように議論していくべきか
 - 規制当局の観点からは、まずは事前面談で相談してほしい
 - アカデミアの観点からは、ベイズを利用した試験デザインなどに挑戦したいと思う場面はあるが、企業側の考えもあるので自由にできるわけではなく保守的になることも多い
 - 検証試験でベイズを利用することは可能か → 実例での議論を重ねる必要あり
 - ベイズを利用した方がより妥当と説明できればCIDを採用することもありえると思うが、相談するには企業側、規制当局側も労力がかかる → 自由に議論できる場があればよい
 - 事前面談などの正式な相談ではなく、もっとハードルの低い相談の枠組みがあるとよい