

第5回 DSRT
医薬品の臨床試験および製造販売後調査における
ベイズ統計学の活用について

【ディスカッション結果】
議題1～5

【ディスカッション結果】

議題1 : Short Discussion

【テーマ1】 ベイズ統計学のメリット・デメリット（結果）

メリット

- 結果が理解しやすい
- 過去の試験結果を事前分布として利用できる
- 確率を逐次更新できる

[その他意見]

- 複雑な分布を扱える
- 意思決定に利用できる

デメリット

- 計算が大変
- 動作特性の確認など、これまで以上に手間がかかる
- 事前分布の設定の際に憑依性が入る
- 結果によっては解釈が難しい

[その他意見]

- まだ認知度が低い
- 検証的でない
- 仮説検定に慣れていて、ベイズに切り替えが大変

【テーマ2】 ベイズ統計学を使ってみたい場面（結果）

検証

- 仮説検定の代替
- 例数設計
- 非劣性試験
- 中間解析

P2試験

- 用量設定

その他

- PMS
- 社内での意思決定
- リスクの段階的評価
- 臨床データ以外のデータリンケージ
- 安全性解析の意思決定
- 希少疾患の評価
- 報告数が少ないAEの評価
- 慢性疾患の長期的評価
- 日常生活

【ディスカッション結果】
**議題 2 : ベイズ流の統計手法と
頻度論の統計手法**

ディスカッションテーマ

- ベイズ流・頻度論の統計手法について議論
 - 二値判断：Bayes factor (BF)と仮説検定
 - 推定：パラメータの事後分布の信用区間と信頼区間
- 「ベイズ流 vs 頻度論を議論の目的とはしない」

ディスカッション結果・二値判断

- **なぜBFはあまり使われていないのか？
なぜ頻度論の方法をよく使うのか？**
 - BFは α エラーを制御できない
 - α エラーを制御できる仮説検定は判断に使いやすいから
- **BFを利用するとしたら、どういう情報が必要か？
BFの使用を考えた際に何が気になるか？何が障壁になるか？**
 - 標準的な基準を定める必要がある（Jeffreys, Kass & Rafteryなど）
 - 逐次的な評価に相性がいい
- **仮説検定の代替になるか？**
 - ベイズを使うには、頻度論で考えていた概念を変えるべきじゃないか？ α エラーをそのまま使えなさそう
- **BFはPhase IIIの検証試験等で使えそうか？**
 - 希少疾患の試験：事前分布をうまく設定できれば例数が少なくなる？中間解析には利用可能かもしれない。ICH E9： α エラー率の制御の問題がある

ディスカッション結果・推定

- **信用区間の使用を考えた際に何が気になるか？**
 - BFと信用区間が対応しない
 - 事前情報の設定を失敗すると、被服確率が下がる
 - 事前情報の選択で恣意性が生まれてしまう
- **信用区間を使う際にはどのような点に注意が必要か？**
 - 事前情報の抽出のルールが必要
 - 事前情報として用いて良いデータがわかると望ましい

まとめ

- BF
 - 使い方が見出されてこなかった
 - 仮説検定と異なり n が増加すると正しい仮説を選択する確率が高くなってゆくため、経時的に情報が蓄積されるようなデザインに適しているかもしれない
- BF・信用区間共通
 - 事前分布の設定が結果に影響を与える可能性がある
 - 標準的な事前分布の設定方法などが必要
 - BFと信用区間は仮説検定と信頼区間のように対応しない点に注意が必要

【ディスカッション結果】
議題3：臨床試験における
意思決定の方法の検討

【論点1】 これまでに、（多重性の調整等は適切に行った上で）有意水準両側5%もしくは片側2.5%以外の基準で検証試験を行ったことがありますか？ある場合、どのような状況でしたか？

- 有意水準を小さくする方向では、治療期と維持期の両方を1試験で検討する場合に、 $\alpha=0.002$ にしたことがある
- ある規制当局に、PIIで結果が極めてよい場合に、PIII試験なしで承認されたことがある
- 有意水準を大きくする方向では、条件付き承認になることが多いのでは？
- Pivotal 試験1つで行う場合についての Points to Consider (CPMP) :
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003657.pdf
- 規制当局に再現性を求められる

【論点2】 今回のように、Type 2 harm を考慮して、試験ごとに有意水準やサンプルサイズ計算に用いる方法について、賛成ですか、反対ですか？

<賛成意見>

- Type 2 harm を考えること自体はよいのでは？
 - α を変更する場合は、(社内, 医師, 規制当局含め)様々なハードルがありそう
- 具体的な設定は難しいが、概念としては面白い
 - 「薬を世に出すことにより、状況がよくなる」ことを評価することは重要
 - 実際にどう設定する？
 - 社内や臨床家への説明が難しいか？

<反対意見>

- 対象患者の人数が大きくなるとコストが大きくなる今回の方法では、majority の人のメリットがあるようになる
 - 対象患者の多い疾患に有利で、希少疾患にはむしろ不都合になる？
 - 会社のポートフォリオとして、「開発の優先順位が下がる」方向になる？
 - 希少疾患の「1人当たりのコスト C_1, C_2 」が大きい場合は、被験者数が少なくてもコストが増えて優先順位が上がるか？

【論点3】 Montazerhodjat et al. (2017)で提案された方法の メリット・デメリットはどのようなものが考えられますか？

メリット

- 有意水準に関する議論が活発化するのでは？
- 有意水準両側5%が厳しすぎると考えるときに、「少し大きく」して試験を行いたい場合の正当化の方法としてはよいのでは？
- 社会的な状況（試験外の患者さんら）を考慮に入れられる点はよい
- 「今後発症する患者さんの状況」等を可視化できる点は魅力的
 - 社内での検討の際に、このような考え方は有益かもしれない
- 有効性に従って柔軟に α や n を設定できるのはよい

デメリット

- α が大きくなりすぎた場合（30%以上など）、結果の「納得性」が得にくいのでは？
 - メーカーとしての見解を考えるのが難しそう
- パラメータ設定が難しい。
 - T_1 の値は適切に設定できるか？
- パラメータの設定により α や n が大きく変動する
 - 感度分析の結果、 p_0, γ の設定に強く依存した
 - 本論文よりもシンプルなモデルを用いることで、より安定したモデルを作ることは可能かもしれない
- 式が難しすぎる

【ディスカッション結果】

議題4：製造販売後調査におけるベイズ統計学の利用に関する検討

議題

重要な潜在的リスクに関するRQ「各事象は本当にリスクと言えるほど発現確率が高いのか？」の検討を目的とした場合、

1. ベイズ統計学を利用することで得られるメリットについて、以下の視点から考えよう！
 - 会社にとって、どんなメリットがあるだろうか？
 - 医師にとって、どんなメリットがあるだろうか？
 - 患者にとって、どんなメリットがあるだろうか？
 - 当局にとって、どんなメリットがあるだろうか？
2. ベイズ統計学を利用する際に障害となりうるものは何か、考えよう！

1. ベイズ統計学を利用することで得られるメリット まとめ

- 当日は患者，医師，会社，当局それぞれのメリットについてdiscussionしてもらったが、どのメリットも患者，医師，会社，当局共通のメリットになりうる、という結論になった。

ベイズ統計学の特長	期待できるメリット
既存情報を（事前分布として）利用可能	→ <ul style="list-style-type: none">・ 効率化（症例数、時間、リソース）・ より豊富な情報量に基づく評価が可能
定量的な評価が可能	→ <ul style="list-style-type: none">・ 結果の解釈が容易
リスクの逐次的な評価が可能	→ <ul style="list-style-type: none">・ タイムリーな添付文書改訂・ タイムリーなリスク評価、注意喚起・ 再評価期間の短縮
従来使われてきた手法よりも科学的なアプローチ	→ <ul style="list-style-type: none">・ 企業・当局間で合意しやすい試験デザインの立案

2. ベイズ統計学を利用する際に障害となりうるもの まとめ

計画段階

- Research questionの決定方法
- 事前分布の設定方法
 - 事前情報とPMSで背景情報が異なる可能性
 - 事前情報に利用する情報を選択してよいか・どうやって選択するか
- (リスクかどうかを判定することを目的とした場合、) 判定に用いる閾値発現率の決定方法

インフラ

- PMS統計担当者のベイズ統計学に関する理解
- 解析用ソフトウェア
- 結果を利用する人 (企業/医師/患者/当局) における結果解釈に関する理解

【ディスカッション結果】
議題5：PII非劣性試験における
ベイズ統計学の利用可能性

2試験の情報から作られた事前分布について

- 事前分布に利用された2つの過去試験が、「似ている」と判断されたため併合しているが、「似ている」と判断するにはどのような検討がなされるべきか？
 - 併合する際に、異質性の検討をすべき
 - 体重の類似性などが検討されるべきかもしれない。例えば今回の2試験では背景表で体重が10kg近く異なる。
 - SPORTIF III ではアジア人が少ない
 - 間違った併合をした場合のリスクの大きさも認識しながら調整
 - この試験を国際共同試験として計画するかどうかも検討されるべき
 - 抗がん剤では、薬剤の特徴が検討されるべき
- 併合しないとすれば、どのような方法が考えられるか？
 - 2試験のデータの混合分布を用いて重みづけを行い、事前分布とする

Prior data conflict(こついで)

- Prior data conflict が起きることを想定して、何を検討しておけばよいか？
 - 動作特性を予め算出し、プロトコルに記載する
 - 症例数の下で、どれだけずれていたら影響が大きくなるかについて、プロトコルに記載しておく
 - ずれた場合の対策も含めて記載しておく
- a_{WH} の事前に決める際に注意すること
 - 影響度合いを見て決める。そのために動作特性を参考にできるのでは
 - 試験の性質にもよるのでは（非劣性・同等性試験であれば、重み大きくできるのでは？）

独立モニタリング委員会による中間解析

- 独立モニタリング委員会による中間解析について
 - 閾値をどう決めるか？（数値を事前規定するか，手順を事前規定するか？）
 - 臨床的な観点から設定（これ以上イベント発現したらダメ）
 - 精度を維持できるように設定（動作特性が大きくなりすぎることを閾値に設定）
 - その他のリスク（割付け比率のズレが頻度論流の評価の際検出力低下に影響）
 - 非盲検でなくても，盲検下でprior data conflict の検討は十分できる可能性がある
 - 中間解析のタイミングは事前に検討しておくべき
 - 途中で結果を見たところで，最終時点でPrior data conflictが起きないことは保障できない

事前に決めておいた重み a_{wH} の変更

- 事前に決めておいた重み a_{wH} を途中で変更する場合に考慮すべき点
 - 8,000例が急に12,000例に変わるとリクルートの面で困難である
 - 途中で割り付け比率を変更することはできるが、そうであれば1:3よりは1:2の方が実施しやすい
 - 先に頻度論を計画しておいて(たとえば1群6,000例), 頻度論の中間解析のように8,000例での有効中止のときのみ事前分布を使用するのは?
 - 中間解析のような形で, 最終解析の α を事後的にコントロールすべき
- 結果が得られてから a_{wH} を決めるのは?
 - Adaptiveなpriorを使った場合, 検出力不足の可能性はある

動作特性のシミュレーションの設定

- α エラー・検出力の定義は正しい？
 - 頻度論流に準じて設定していて、同じものとはいえないが、納得できる
 - そもそも、頻度論流の検定に合わせて評価しなくてもいいのでは？
- 過去試験の得られたデータを固定してシミュレーションするべきか？確率変数としてサンプリングすべきか？
 - 状況次第で動作特性の意味が変わる。
 - 今回のように過去試験が「すでに手に入っている」場合、今回の試験で用いられる decision rule としての意思決定手法としての動作特性を検討している場合は、固定すべきと考える
 - 方法論としてこの手法の動作特性を検討するという目的であれば、確率変数とすることも可