

日本臨床試験学会 第12回学術集会

シンポジウム1 ●統計的意思決定に基づく柔軟な臨床試験デザイン

1 小規模臨床試験デザイン

京都府立医科大学大学院医学研究科生物統計学（シンポジウムオーガナイザー・座長）

手良向 聰

はじめに

これまでの臨床試験には頻度流統計学が主に用いられていたが、2010年に米国食品医薬品局（Food and Drug Administration: FDA）が「医療機器の臨床試験におけるベイズ流統計学の利用に関するガイダンス」^{1),2)}を発行して以来、その風向きは変わりつつある。

2017年に独立行政法人医薬品医療機器総合機構（PMDA）科学委員会希少がん対策専門部会は、「希少がんの臨床開発を促進するための課題と提言 2017—アカデミア及びレギュラトリーサイエンスの視点から—」³⁾を公表した。本報告書は、幅広い見地から課題が整理され、重要な提言がなされているが、「第4章 希少がんに対する臨床試験のあり方」には、試験デザインの制約を緩和できる条件や状況として、以下の3項目が挙げられている。

- ・単群試験デザインの許容
 - ・奏効率等の代替評価項目の採用
 - ・第I種・第II種の過誤確率の上昇
- また、効率的なデザインとして以下の例示がある。
- ・遺伝子情報に基づいた試験（アンブレラ試験/バスケット試験）
 - ・N-of-1試験などの自己対照試験
 - ・適応的（アダプティブ）デザイン
 - ・ベイズ流デザイン

最近、わが国でも国立研究開発法人日本医療研究開発機構（AMED）医薬品等規制調和・評価研究事業（2019～2021）「希少疾患領域の医薬品開発を効率化するための小規模臨床試験のデザイン・統計解析法の研究開発とその適正使用のための基本的考え方の策定（研究開発代表者：平川晃弘）」において、「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」⁴⁾

が策定された。

ベイズ流統計学は、①「確率」だけですべてを判断するため、わかりやすい、②試験計画に依存せず、いつでもデータ解析をして知りたい確率を計算できるため、適応的デザインに適している、③事前情報はときに意味をもち、事前情報を含むさまざまな情報（外部情報）を明示的に統合できるため、蓄積された情報を生かすことができる、という利点を有する。

1 本シンポジウムの趣旨

本シンポジウムでは、小規模臨床試験における意思決定問題を議論する。小規模臨床試験には、実験的介入を伴うことによる「倫理性」、資源が限られていることによる「効率性」と「柔軟性」が重要である。また、意思決定者（認識主体）はだれか？ ということを意識した議論を行う。すなわち、臨床試験における科学的な問い合わせ「この治療には効果があるのか？ ないのか？」ではなく、認識論的な問い合わせ「この治療に効果がある場合（または、ない場合）、それをどのようにすれば知ることができるのか？」を考える。認識主体としては、医師個人またはコミュニティ、患者個人またはコミュニティ、規制当局、製造販売企業などが想定される。その際、「ベイズ流統計学は内在主義であるが、信念を事実（体系の外部）に近づける必要があり、頻度流統計学は外在主義であるが、正当化はデザインの信頼性に依存する」というような統計学と哲学的認識論を関連付ける議論は有用と考える⁵⁾。

2 小児がん領域の事例

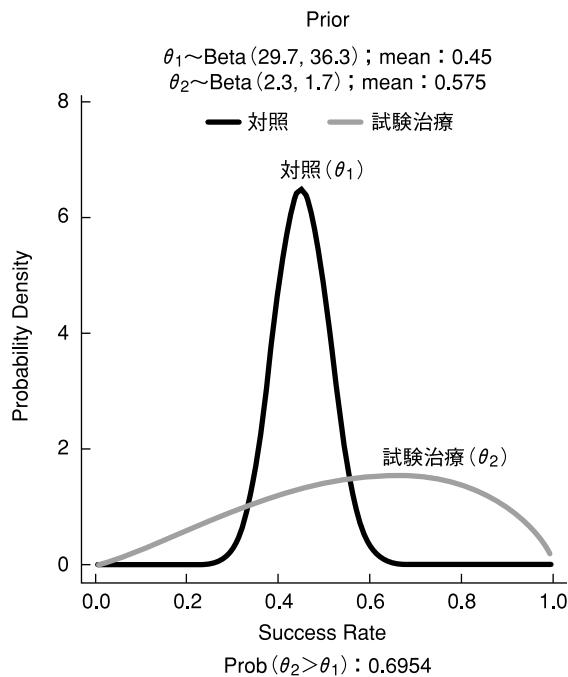
日本小児がん研究グループ（Japan Children's Cancer Group: JCCG）は、2014年12月に設立され、そこには日本的小児がん治療・研究を専門とする大学病院、小児病

Design of Small Clinical Trials

Satoshi Teramukai: Department of Biostatistics, Graduate School of Medical Science, Kyoto Prefectural University of Medicine

表 1 2 値評価項目・単群臨床試験のデザイン

- ・頻度流
 - ・2項確率に基づく仮説検定
 - ・2(多)段階デザイン Simon 法, SWOG 法
- ・ベイズ流 (標本サイズ設定)
 - ・事後確率に基づく方法 Tan (2002)⁷⁾, Whitehead (2008)⁸⁾
 - ・事前予測確率に基づく方法 (事例 2) Sambucini (2008)⁹⁾
 - ・予測確率に基づく標本サイズ再設定 Sambucini (2010)¹⁰⁾, Teramukai (2012)¹¹⁾
- ・ベイズ流 (中間モニタリング)
 - ・事後確率に基づく方法 (事例 1) Thall (1994)¹²⁾
 - ・事後予測確率に基づく方法 (事例 2) Lee (2008)¹³⁾, Teramukai (2012)¹¹⁾
- ・決定理論に基づく方法



- ・対照の分布
 - ・1年 EFS の閾値 : 45%
 - ・先行試験結果から設定
 - ・Beta (29.7, 36.3)
 - ・平均 45%, 90% 信用区間幅 20%
- ・試験治療の事前分布
 - ・1年 EFS の増分 : 25%
 - ・Beta (2.3, 1.7)
 - ・平均 57.5% (=45%+25%/2),
有効標本サイズ 4

図 1 仮定された対照の分布と試験治療の事前分布

院、総合病院が 200 以上参加し、倫理性、科学性を重視した臨床研究を実施している。私は 2003 年ころから小児がんの臨床研究に携わり、現在は JCCG 生物統計委員会委員として活動している⁶⁾。JCCG で実施されている臨床試験の多くは、希少疾患を対象とした単群臨床試験で

表 2 有効・無効の判断 (確率閾値)

- ・1年 EFS が対照を上回る確率が 95% 以上
 - ・本治療は有効
- ・1年 EFS が対照 (+0.25) を上回る確率が 1% 以下
 - ・本治療は無効
- ・症例数 : 10~18 例

症例数ごとの無イベント生存数

症例数	10	11	12	13	14	15	16	17	18
有効中止 (下限値)	8	8	9	10	10	11	11	12	12
無効中止 (上限値)	2	3	3	4	4	5	5	6	6

ある。ここでは 2 値評価項目・単群試験に焦点を絞り、事例 1 では事後確率に基づくベイズ流中間モニタリング、事例 2 では事前予測確率に基づくベイズ流標本サイズ設定・中間モニタリングを紹介する (表 1)^{7)~13)}。

1) 事例 1 : リンパ腫臨床試験

JCCG リンパ腫委員会が実施している「小児の再発・難治性未分化大細胞リンパ腫に対する骨髄非破壊的前処置を用いた同種造血幹細胞移植の有効性と安全性を評価する多施設共同非盲検無対照試験 (臨床研究実施計画番号 : jRCTs041190066)」の主要評価項目は、移植後 1 年無イベント生存割合 (1 年 EFS) である。1 年 EFS 確率に仮定された対照の分布と試験治療の事前分布を図 1 に示す。この設定のもとで、表 2 に示す有効および無効の基準を規定すると、10~18 例の各症例数について、それらの基準を満たす有効中止の下限値と無効中止の上限値が算出される。この事例では、検出力に基づいて標本サイズを設定することが困難な場合に、実施可能な症例数の範囲でベイズ流の中止基準を設定することで、治療効果の有無を判断することが可能となっている。

2) 事例 2：横紋筋肉腫臨床試験

JCCG 横紋筋肉腫委員会が実施している「横紋筋肉腫高リスク群患者に対する VI/VPC/IE/VAC 療法の有効性及び安全性の評価 第 II 相臨床試験(臨床研究実施計画番号 : jRCTs051180208)」の主要評価項目は、治療開始 54 週後までの進行病変 (PD) の有無である。この試験では、ベイズ流事前予測確率に基づいて標本サイズが設定された。この方法は、pre-posterior analysis (事前事後解析) とよばれ、以下の 2 つの事前分布を用いる。

① fitting prior (analysis prior : 解析事前分布)

解析時に事後分布を計算するために用いる。

② sampling prior (design prior : デザイン事前分布)

デザイン時に “what if” spirit (「仮に〇〇だったら」の精神) から生じる。

先行試験の結果から成功 (非 PD) の目標値を 80% とし、解析事前分布は無情報一様分布 (Beta (1,1))、デザイン事前分布は平均 95%、標準偏差 2.5% のベータ分布 Beta (71.25, 3.75) とした。有効確率が目標値 80% を上回る事後確率が 0.9 を超える場合に、「本治療は有効である」と判断することとし、ベイズ流検出力は 0.9 以上とした。この仮定のもとで目標登録症例数を 60 例と設定した。ちなみに、この計算ツールは京都府立医科大学・生物統計学ホームページ (<http://www.f.kpu-m.ac.jp/k/biostat/>) において利用可能である。

また、中間モニタリング計画として、「最初の登録症例 20 例のデータ固定時点で、事後予測分布を推定し、最終的に「有効である」と判断される事後予測確率が 0.1 未満であれば、試験の早期中止を効果安全性評価委員会において検討する」という無効中止の基準を設定した。

おわりに

Kadane JB & Seidenfeld T¹⁴⁾ は、実験を「experiments to learn (学習のための実験)」と「experiments to prove (証明のための実験)」に分類している。前者において、意思決定者は個人や狭いコミュニティであり、主観ベイズ流の方法や決定理論が適しており、後者において、意思決定者は広いコミュニティや規制当局であり、客観ベイズ流や頻度流の方法が適していると考える。

小規模臨床試験はさまざまな状況で実施されているが、試験統計家には、意思決定者を想定して臨床試験デザインを策定することによって、「必要な情報を必要な

人々に届ける」という発想が求められているのではないかと思う。

【謝 辞】

本稿で紹介した事例の研究代表者および統計解析責任者・担当者である JCCG リンパ腫委員会：深野玲司氏（山口大学小児科）、JCCG 横紋筋肉腫委員会：細井創氏（京都府立医科大学小児科）、小川淳氏（新潟県立がんセンター新潟病院小児科）、JCCG 生物統計委員会：嘉田晃子氏（名古屋医療センター生物統計研究室）、橋本大哉氏（名古屋市立大学病院臨床研究開発支援センター）、横田勲氏（北海道大学医学統計学）に感謝いたします。

文 献

- 1) FDA. Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials, 2010.
<https://www.fda.gov/media/71512/download>
(2021 年 5 月 19 日アクセス)
- 2) 手良向聰. なぜベイズを使わないのか!?-臨床試験デザインのために. 金芳堂. 2017.
- 3) PMDA 科学委員会希少がん対策専門部会. 希少がんの臨床開発を促進するための課題と提言 2017—アカデミア及びレギュラトリーサイエンスの視点から—.
<https://www.pmda.go.jp/files/000223815.pdf>
(2021 年 5 月 19 日アクセス)
- 4) 平川晃弘, 佐藤宏征, 井桁正堯, 藤川桂, 堀口剛, 大門貴志, 手良向聰. 希少疾患領域の治験におけるベイズ流アプローチの利用可能性と留意事項. 投稿中.
- 5) 大塚淳. 統計学を哲学する. 名古屋大学出版会. 2020.
- 6) 手良向聰. 希少がんに対する臨床試験デザイン. 日小児血がん会誌 2019; 56: 425-8.
- 7) Tan SB, Machin D. Bayesian two-stage designs for phase II clinical trials. Stat Med 2002; 21: 1991-2012.
- 8) Whitehead J, Valdés-Márquez E, Johnson P, Graham G. Bayesian sample size for exploratory clinical trials incorporating historical data. Stat Med 2008; 27: 2307-27.
- 9) Sambucini V. A Bayesian predictive two-stage design for phase II clinical trials. Stat Med 2008; 27: 1199-1224.
- 10) Sambucini V. A Bayesian predictive strategy for an adaptive two-stage design in phase II clinical trials. Stat Med 2010; 29: 1430-42.
- 11) Teramukai S, Daimon T, Zohar S. A Bayesian predictive sample size selection design for single-arm exploratory clinical trials. Stat Med 2012; 31: 4243-54.
- 12) Thall PF, Simon R. Practical Bayesian guidelines for phase II b clinical trials. Biometrics 1994; 50: 337-49.
- 13) Lee JJ, Liu DD. A predictive probability design for phase II cancer clinical trials. Clin Trials 2008; 5: 93-106.
- 14) Kadane JB, Seidenfeld T. Randomization in a Bayesian perspective. J Stat Plan Inference 1990; 25: 329-45.