

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

2 分子標的薬を用いた医師主導治験における バスケット試験デザインの利用

北海道大学大学院 医学研究院 医学統計学教室/北海道大学病院 医療・ヘルスサイエンス研究開発機構
プロモーションユニット データサイエンスセンター

横田 勲 岡田 和史

北海道大学大学院 医学研究院 医学統計学教室/独立行政法人医薬品医療機器総合機構

高木 諒

はじめに

1990年代以降登場した分子標的薬は、特定のたんぱくや遺伝子といったがんの増殖に関わる分子を攻撃することで、がん細胞の増殖を抑える。有効性は、遺伝子変異のような分子マーカーの有無に応じて得られるものと考えられ、臓器横断的にみられる特徴がある。たとえば、PARP (ポリADP-リボースポリメラーゼ) 阻害薬では、BRCA1/2 遺伝子変異を有するがんの効果があるとされる。PARPは細胞増殖におけるDNAの複製におけるDNA一本鎖切断を修復する酵素であり、阻害されれば、DNA二本鎖切断に至る。正常細胞とは異なり、BRCA1/2 遺伝子変異を有するがん細胞では二本鎖切断修復が行われない。こうしてPARP阻害薬は、BRCA関連遺伝子変異を有するがんに特異的に作用するメカニズムをもつ。

複数の臓器に有効性が期待でき、治療方法や評価項目が同一であるにもかかわらず、従来の臓器別に評価する臨床試験の枠組みでは、別々の試験として計画せねばならず、非効率である。そこでいくつかの臓器に対し、共通する分子マーカーを有する患者集団において単一のプロトコルを用いて実施するバスケット試験デザインが提案された。試験遂行のための手順を共通化して効率化を図るだけでなく、臓器間で同様な有効性が確認された場合には類似性を考慮した解析手法によって統計的効率を

向上させることも可能である。有効性を左右する分子マーカーを有する対象者の割合は、臓器、分子マーカーおよび人種のような集団の特徴に応じて大きく異なる。たとえば、BRCA関連遺伝子変異は日本の卵巣がん患者ではおよそ15%にみられる¹⁾一方、消化器がんではまれにみられる程度である。希少性のある対象集団であれば、統計的効率が良いデザインを採用し、サンプルサイズを小さくできれば、開発意欲を高めることにつながるであろう。

本稿では、バスケット試験を含む、分子マーカーと臓器の組み合わせを部分試験としてとらえ、全体をひとつのプロトコルに表すマスタープロトコル試験について述べた後、医師主導治験の適用を念頭に置いた数値例を用いながらバスケット試験デザインを紹介する。さらに近年登場したデザインであるために、バスケット試験デザインを採用するにあたり、スポンサー内および規制当局との検討した経験はあまり共有されていない現状を踏まえ、とくにfamily-wise error rateを制御する観点を紹介し、バスケット試験デザインの利用を積極的に進めるうえでの課題について検討したい。

1 マスタープロトコル試験

従来の一臓器一薬剤に対する臨床試験を部分試験としてとらえたマスタープロトコル試験は、バスケット試

Use of Basket Trial Design in Investigator Initiated Clinical Trials for Molecular-Targeted Agents

Isao YOKOTA, Kazufumi OKADA: Department of Biostatistics, Graduate School of Medicine, Hokkaido University/Data Science Center, Promotion Unit, Institute of Health Science Innovation for Medical Care, Hokkaido University Hospital; Ryo TAKAGI: Department of Biostatistics, Graduate School of Medicine, Hokkaido University/Pharmaceuticals and Medical Devices Agency

験, アンブレラ試験, プラットフォーム試験の3つに大別される^{2),3)}。大まかな定義は以下のとおりであるが, いまだにあいまいな部分も大きい。

①バスケット試験

複数のがん種に, 1つの分子マーカーと薬剤の組み合わせを実施する試験

②アンブレラ試験

1つのがん種に, 複数の分子マーカーと薬剤の組み合わせを実施する試験

③プラットフォーム試験

複数のがん種に, 複数の分子マーカーと薬剤の組み合わせを実施する試験

図1aに示すようなバスケット試験は, 同一の化合物を用いるため, 効果を期待できる臓器のスクリーニング目的を併せ持つような proof of concept 評価の早期第II相試験で利用されやすい。それぞれの部分試験では, 腫瘍に対する反応の有無を評価項目とする単群試験となることが多い。生存時間アウトカムは, 臓器間での異質性の影響が大きく, 過去情報などから落とし込んだ閾値生存曲線の妥当性について解釈上の困難を伴うことから, あまり主要評価項目として採用されることはない。

アンブレラ試験は, 1つの臓器において, さまざまな分子マーカーを同時に測定し, その結果に応じて, 部分試験に対象者を進めるような試験である。そのため各部分試験は, 検証的な目的を有することがあり, 単群試験だけでなくランダム化並行群間デザインが採用されることもしばしばみられる。

プラットフォーム試験は複数の臓器にさまざまな分子マーカーを測定し, 分子マーカーの内容に応じて部分試験を実施する。臓器を特定せず, 標準治療不応のすべてのがんのような広い対象集団の定義である場合も含む。試験開始後にヒトへの臨床応用が進んだ新規化合物を部分試験において追加することもあり, 長期にわたってさまざまな臨床試験に被験者を紹介するような体制のような試験ともいえる。たとえば, 図1bに示すように, 国立がん研究センター東病院が主導する SCRUM-Japan は, 肺がんと消化器がんに端を発し, 現在では広範な固形がんを対象に, 血液検体から複数のがん遺伝子変異を探索し, 発見された遺伝子変異に対する部分試験が登録中であれば被験者が紹介される仕組みで実施されている。

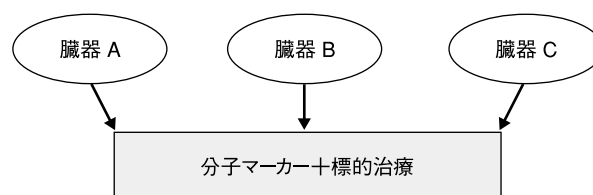
2 バスケット試験デザイン

各臓器において反応の有無を評価する結果をもとに, 以下の3とおりの意思決定手順が考えられる。

①臓器を無視し, 試験全体における反応割合を解析

②臓器別に反応割合を解析

a バスケット試験



b SCRUM-Japan を例にしたプラットフォーム試験

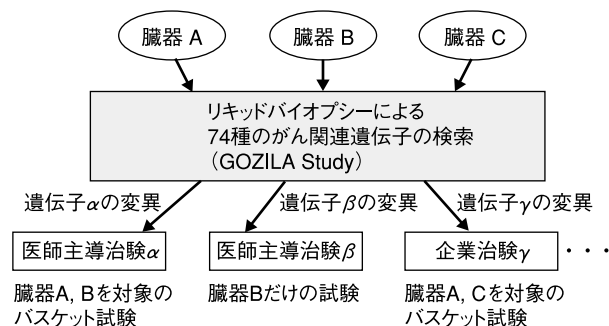


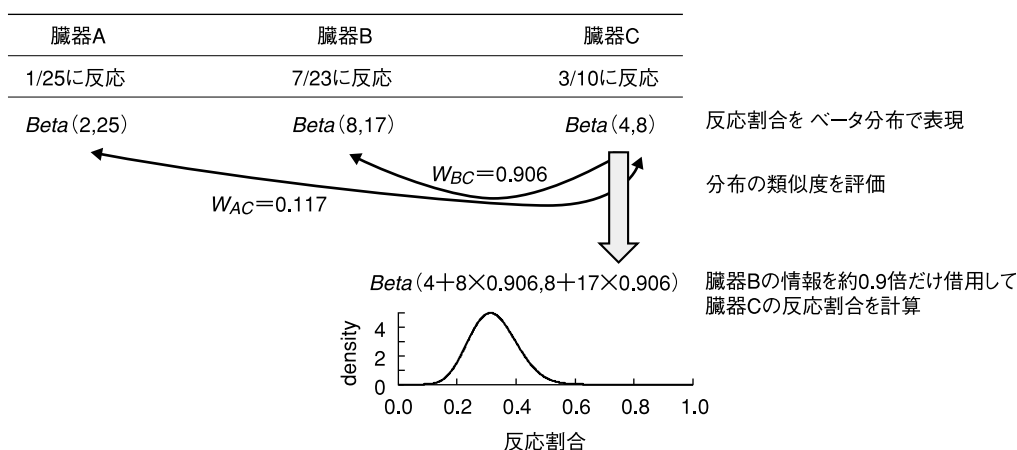
図1 試験デザインの略図

③臓器別に反応割合を解析するものの, 他の臓器における情報の一部を利用

臓器を無視し, 試験全体における反応割合を解析する場合, すべての臓器に有効か否かの all or none の意思決定となる。ここで, 臓器間の反応割合に異質性がみられた場合, 反応割合が悪い臓器にも有効と意思決定を下してしまうため, 解釈困難となるおそれがある。一方, 臓器それぞれにおける反応割合を解析するならば, 臓器それぞれに十分なサンプルサイズを設定する必要がある, 希少性のある集団における開発が困難である。個別の意思決定となるため, 検定の多重性に対する調整が必要にもなる。このことは, 次節で詳しく述べたい。

上記2つの意思決定方針の弱点を克服するため, 臓器別に反応割合を解析するものの, 他の臓器における情報の一部を利用する統計的手法がいくつか開発された。詳しくはいくつかのレビューに任せ, ここでは, 試験計画に用いた Fujikawa ら (2020)⁴⁾の方法について, 数値例を用いて概説する。

図2に示すように, 3つの部分試験(臓器)を含めるバスケット試験において, 臓器Aでは25例中1例の反応, 臓器Bでは23例中7例の反応, 臓器Cでは10例中3例の反応が観察されたとする。事前分布に無情報事前分布のひとつである一様分布($Beta(1,1)$)をおいた Bayes 推定を行うと, それぞれの臓器の反応割合は, $Beta(2,25)$, $Beta(8,17)$, $Beta(4,8)$ という事後分布で表される。臓器別に反応割合を解析する場合であれば, 以上の推定手順となるが, 提案法ではさらに類似度に基づい



意思決定基準 (例)

反応割合に関して閾値10%を超える確率が99%を超えた場合、有効と判断

図 2 数値例による Fujikawa ら (2020)⁴⁾ デザインによる意思決定手順

た情報借用を行う。臓器Bの事後分布 $Beta(8,17)$ は期待値 0.32, 臓器Cの事後分布 $Beta(4,8)$ は期待値 0.33 であり, 分布の位置は類似している。これを 1-Jensen & Shannon divergence という, 完全に一致すれば 1, まったく一致しなければ 0 をとる類似度指標で表現すれば, 0.906 と求まる。一方, 臓器Aの事後分布 $Beta(2,25)$ は期待値 0.07 であり, 臓器Cの事後分布との類似度は 0.117 にすぎない。そこで, 類似度指標が一定の基準を超えたならば, 臓器間の反応割合は「類似している」と判定し, 互いの事後分布の引数に類似度を乗じた値を加え, 事後分布を更新する。たとえば, 臓器Bの事後分布であれば $Beta(8+4 \times 0.906, 17+8 \times 0.906)$, 臓器Cの事後分布であれば $Beta(4+8 \times 0.906, 8+17 \times 0.906)$ とする。こうして最終的に計算された事後分布について, 閾値反応割合を超える確率が事前に定めた水準以上であれば, 有効という判断を下す。

3 Family-Wise Error Rate の制御とサンプルサイズ設計

頻度流仮説検定の枠組みにおいて, 本当は差がない仮説のうち, 1つ以上の仮説で有意と判断する確率を family-wise error rate (FWER) という。複数の仮説検定を行う状況において, 第一種の過誤 (α エラー) 率を拡張した考えである。ベイズ流デザインを採用した場合でも, 本当は有効でない部分試験のうち, 少なくとも 1つ以上の部分試験において有効と判断する割合を制御することが, 規制上の点から重要視された。有効と判断された臓器について, 承認に向けての議論が行われるため, 有効でない臓器のうち, 誤って有効と判断してしまう状況が発生する確率を制御したい要請は自然なものと考えられる。

たとえば, 3つの部分試験を含むバスケット試験において, 有効でない臓器が 1つ, 2つ, 3つそれぞれの場合, FWERを制御する必要が生じる。試験計画を行ううえで, さまざまなシミュレーション実験を通した結果, 有効な臓器が 2つ, 有効でない臓器が 1つである場合, 有効な臓器から有効でない臓器に情報借用され, 有効と判断する状況が多少みられることがわかった。類似度指標や情報借用を行うか否かの判断基準を工夫することで, FWERの制御を行う必要があるであろう。

サンプルサイズ設計においては, 各部分試験の power のほか, いずれか 1つ以上の部分試験で有効と判断する割合である experiment wise power (EWP) も決定指標のひとつとなる。すべての部分試験において期待される反応割合である場合のほか, いくつかの部分試験でのみ期待反応割合であり, 残りの部分試験では閾値反応割合である状況で power や EWP を求めることが, 試験デザインの決定やサンプルサイズ設計を議論するうえで肝要であろう。

おわりに

本稿ではマスタープロトコル試験のうち, 事後分布の類似度による情報借用を行うベイズ流バスケット試験デザインを利用した実例を踏まえ, デザインの紹介や適用上の注意点を述べた。令和2年8月31日に治験届の取り扱いについて, マスタープロトコル試験を念頭に置いたものとして, 厚生労働省より局長通知が発出された (薬生薬審発 0831 第 10 号)。柔軟な試験デザインを採用することにより, ますます効率的な開発が進むであろう。本稿で述べたような経験の共有がその一助となれば幸いである。

【資金源】

本稿は、国立研究開発法人日本医療研究開発機構 (AMED) 臨床研究・治験推進研究事業「SCRUM-Japan の基盤を活用した血液循環腫瘍 DNA ゲノムスクリーニングに基づく BRCA 変異を有する胆道がんに対する医師主導治験」(代表: 川本泰之) の助成を受けた。

【利益相反】

本稿は個人的見解に基づくものであり、(独) 医薬品医療機器総合機構の公式見解ではない。横田勲は本稿とは直接関係しないが、日本学術振興会、日本医療研究開発機構、厚生労働省からの研究費、中外製薬、アストラゼネカ、日本たばこ、日本新薬からの講演料を受け取った。その他の著者について開示する利益相反はない。

文 献

1) Enomoto T, Aoki D, Hattori K, et al. The first Japanese nation-

wide multicenter study of BRCA mutation testing in ovarian cancer: characterizing the cross-sectional approach to ovarian cancer genetic testing of BRCA (CHARLOTTE). Int J Gynecol Cancer 2019; 29: 1043-9.

- 2) Pohl M, Krisam J, Kieser M. Categories, components, and techniques in a modular construction of basket trials for application and further research. Biom J 2021 [Epub ahead of print]
- 3) 平川晃弘, 浅野淳一, 佐藤宏征ほか. がん臨床試験におけるベイズ流バスケットデザインの理論と実装. 計量生物学 2018; 39: 103-22.
- 4) Fujikawa K, Teramukai S, Yokota I, Daimon T. A Bayesian basket trial design that borrows information across strata based on the similarity between the posterior distributions of the response probability. Biom J 2020; 62: 330-8.