

医薬品開発における医療統計学の貢献

東京理科大学 創域理工学部 情報計算科学科 講師 あんどろ しゅうじ
安藤 宗司

■ はじめに

皆さんが病院で処方される医薬品は、どのようなプロセスを経て開発されているかご存じでしょうか。製薬企業などで実施される臨床試験により安全性・有効性などを確認し、厚生労働省が承認することで医薬品が誕生します。日本においては、医薬品の承認申請のための臨床試験を「治験」と称して区別してきましたが、欧米などの海外ではこのような区別はせず臨床試験 (Clinical Trial) で統一しています。本稿でも、特に区別せず臨床試験と呼ぶことにします。

臨床試験は、大きく第1相試験から第3相試験までの3段階に分けることができます。第1相試験では、初めてヒトに医薬品の候補物質 (以下、被験薬) を投与する段階です。少数の健常人男性を対象に、安全性、薬物動態 (薬物の吸収・分布・代謝・排泄) を調べます。第2相試験では、比較的少数の患者を対象に、被験薬の有効性を調べ、適切な標的集団や被験薬の用法・用量などを検討します。第3相試験では、比較的多数の患者を対象に、被験薬の安全性・有効性を検証します。第3相試験で被験薬の安全性・有効性が認められ、厚生労働省から承認を得られれば、その被験薬は医薬品となります。

ここまで臨床試験の流れを説明してきましたが、医薬品として誕生するまでにどのくらいの期間や費用を要するかご存じでしょうか。諸説がありますが、創薬から承認まで9から17年 (臨床試験だけだと3から7年) もの歳月を要し、その期間にかかる費用は約500億円といわれています。さらに、医薬品の開発成功割合は、約3万分の1とされていて、ほとんどの被験薬は途中段階で開発中止になっています。このことから、医薬品はすごく険しい道のりを経て、皆さんに届いていることをわかってもらえたでしょうか。

昔と比較すると、医薬品の開発はさらに難しくなっていて、開発に要する期間や費用は増え続けている傾向にあります。そこで、効率のかつ精度の高い臨床試験実施の一助となっているのが医療統計学です。

本稿では、医薬品開発における医療統計学の貢献に

ついて、筆者が取り組んでいる研究を題材に紹介させていただきます。医療統計学という分野に少しでも興味をもって頂けると幸いです。

■ がん第1相臨床試験

第1相試験は、通常、少数の健常人男性を対象として実施されます。ただし、抗がん剤のように、用量を増やすほど強い毒性を与えるリスクの高い場合、健常人を対象とすることは倫理的に問題があります。がん第1相臨床試験では、強い毒性のリスクよりも大きな効果を楽しむベネフィットが期待できるがん患者 (標準治療で効果が認められなかったがん患者) を対象として実施されます。したがって、健常人を対象とする第1相試験では考慮する必要のなかった治療という側面もあることに注意が必要です。

がん第1相臨床試験では、比較的重度の毒性の発現頻度のある水準以下に維持できる最大の用量を探索します。許容できない毒性の発現頻度の水準を用量制限毒性 (以下、毒性)、それに対応する用量を最大耐用量と呼びます。安全性、薬物動態 (薬物の吸収・分布・代謝・排泄) に加え、がん第1相臨床試験では、最大耐用量の同定も目的の一つとなります。そのため、用量探索試験とも呼ばれています。

がん第1相用量探索試験のデザインは、大きく「アルゴリズム基盤型デザイン」、「モデル支援型デザイン」、「モデル基盤型デザイン」に分類されます。アルゴリズム基盤型デザインは、その名の通り、事前に規定したアルゴリズムに基づき用量探索を行うデザインです。モデル支援型デザインは、二項モデルなどから導出された指標の支援を受けながら用量探索を行うデザインです。モデル基盤型デザインは、用量毒性反応関係にモデルを与えることで用量探索を行うデザインです。

アルゴリズム基盤型デザインは、試験の実施が容易であり、最大耐用量を超える用量に投与される患者数は少ないのですが、一方で有効性が十分ではない低い用量に割り当てられる患者も多く、他の2つのデザ

インと比較して性能（最大耐用量を正しく同定する割合など）が低いことが知られています。治療という側面もあることから効果の期待できない低用量に投与される患者数は最小限に留めるべきです。モデル基盤型デザインは、性能は高いのですが、試験の実施は容易ではありません。モデル支援型デザインは、近年提案されたデザインであり、アルゴリズム基盤型デザインの簡便さとモデル基盤型と同程度の性能を兼ね備えているため注目されています。

最大耐用量の同定を目的としたがん第1相臨床試験を対象に各デザインの使用頻度を調査した研究報告¹⁾では、「モデル支援型デザイン」、「モデル基盤型デザイン」の使用頻度は年々増加していますが、割合は8%とされています。つまり、現在も試験実施の容易さから性能が低いにもかかわらず、アルゴリズム基盤型デザインの使用が主流ということになります。しかし、前述した通り、モデル支援型デザインはアルゴリズム基盤型デザインの簡便さとモデル基盤型デザインと同程度の性能を兼ね備えているため、使用頻度はさらに増加することが予想されます。

「アルゴリズム基盤型デザイン」、「モデル支援型デザイン」について、次節にて紹介していきます。

■ がん第1相用量探索試験のデザイン

被験薬の用量を $d_j (j=1, 2, \dots, J; d_1 < d_2 < \dots < d_J)$ とし、添え字は用量水準を表しています。標的毒性発現確率を ϕ と表します。用量探索試験では、1から3例の患者（以下、コホート）ごとに投与を進めます。がん第1相用量探索試験のデザインにおける最大耐用量を探索する手順を次に示します。

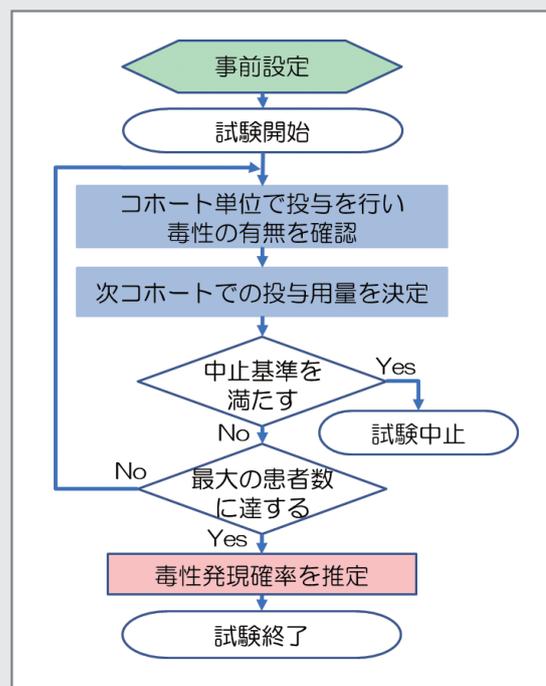
手順1 最低用量 d_1 を最初のコホートに投与する

手順2 現コホートと同じ用量を投与されたすべての患者の毒性発現データを用いて、試験開始時に規定した用量増減基準により、次のコホートの用量（増量、維持、減量）を選択する

手順3 手順2で選択した用量を次のコホートに投与する

手順4 試験開始時に定めた最大の患者数に到達した場合には手順5に進む。それ以外は、手順2に戻る

手順5 収集したすべての毒性発現データから、各用量の毒性発現確率 $p_j (j=1, 2, \dots, J)$ を推定する。毒性発現確率の推定値が標的毒性発現確率 ϕ に最も近い用量を最大耐用量とする。



【図1】最大耐用量を探索する手順のフローチャート

最大耐用量を探索する手順のフローチャートを【図1】に示しています。

「アルゴリズム基盤型デザイン」、「モデル支援型デザイン」、「モデル基盤型デザイン」で異なるのは、手順2の用量増減基準のみであることに注意します。手順1から5に加えて、安全性の観点から適用する手順もありますが本稿では簡単のため説明を省略します。

アルゴリズム基盤型デザインの代表例として、3+3デザインを取り上げます。3+3デザインでは、手順2における用量増減基準を次のように定めています。一つのコホートを3例として、毒性を発現した人数が0例であれば次のコホートの投与用量を増量する。1例であれば次のコホートの投与用量を維持する。2例もしくは3例であれば、その用量は最大耐用量を上回ったと判断し、そこで試験を中止する。同じ用量で合計6例が投与され1例だけに毒性が発現している場合、次のコホートの用量を増量する。このように、次のコホートの用量選択をアルゴリズムにのみ基づき決定するために、試験実施が容易になっています。

モデル支援型デザインの代表例として、Bayesian optimal interval (BOIN) デザイン²⁾を取り上げます。BOINでは、手順2における用量増減基準を毒性許容区間に基づく増減表により定めています。毒性許容区間は、用量 $d_j (j=1, 2, \dots, J; d_1 < d_2 < \dots < d_J)$ に対し【表1】に示した3つの仮説 $H_{0j}: p_j = \phi, H_{1j}:$

【表 1】BOIN デザインにおける 3 つの仮説

$H_{0j}: p_j = \phi$	$H_{1j}: p_j = \phi_1$	$H_{2j}: p_j = \phi_2$
S: 維持	E: 増量	D: 減量
\bar{S} : 増量, 減量	\bar{E} : 維持, 減量	\bar{D} : 維持, 増量

【表 2】毒性許容区間

	ϕ		
	0.26	0.3	0.34
$\phi - \Delta_L$	0.205	0.236	0.268
$\phi - \Delta_U$	0.310	0.358	0.407

【表 3】増減表

	3	6	9	...	33	36
下限	0	1	1	...	6	7
上限	1	2	3	...	11	12

$p_j = \phi_1, H_{2j}: p_j = \phi_2$ を設定し、誤った用量を選択する確率を最小にするように求めます。

各仮説において、S, E, D が適切な用量を選択する事象であり、 $\bar{S}, \bar{E}, \bar{D}$ が誤った用量を選択する事象を表しています。このとき、誤った用量を選択する確率を最小にするように毒性許容区間を決定します。最大耐用量と考える毒性発現確率の区間を $[\phi_1, \phi_2]$ とします。一般的な使用では、 $\phi_1 = 0.6\phi, \phi_2 = 1.4\phi$ と設定することが推奨されています。

毒性許容区間を導出する過程の詳細は省略しますが、毒性発現確率の下限 $\phi - \Delta_L$ と上限 $\phi - \Delta_U$ は次式に求めることができます²⁾。

$$\phi - \Delta_L = \frac{\log\left(\frac{1 - \phi_1}{1 - \phi}\right)}{\log\left(\frac{\phi(1 - \phi_1)}{\phi_1(1 - \phi)}\right)}$$

$$\phi - \Delta_U = \frac{\log\left(\frac{1 - \phi}{1 - \phi_2}\right)}{\log\left(\frac{\phi_2(1 - \phi)}{\phi(1 - \phi_2)}\right)}$$

したがって、毒性許容区間は、 ϕ, ϕ_1, ϕ_2 のみに依存することから、標的毒性発現確率 ϕ を設定することで毒性許容区間を設定することが可能になっています。標的毒性発現確率 ϕ に対する毒性許容区間を【表 2】に示します。

BOIN を使用するという立場になれば、標的毒性発現確率 ϕ を設定すれば、手順 2 の用量増減基準を定めることができるので、アルゴリズム基盤型デザイン

と同様の簡便さがあることを理解して頂けるのではないのでしょうか。

では、これから増減表をどのように作成するのか、具体例を用いて考えていきます。標的毒性発現確率 $\phi = 0.26$ 、1 コホート 3 例、コホート数 12 (総患者数 36 例) の場合を考えます。このときの増減表は【表 3】のようになります。

増減表に従い、次コホートの投与用量を決定します。ある用量での累積投与人数に対して、毒性が発現した累積人数が下限以下であれば次コホートにおける投与用量を増量する。上限以上であれば次コホートにおける投与用量を減量する。いずれにも該当しない場合は、現在の投与用量を維持することになります。

次節では、手順 5 の毒性発現確率の推定と最大耐用量の選択方法に注目して、筆者が取り組んでいる研究³⁾を紹介したいと思います。

■ 毒性発現確率の推定と最大耐用量の選択

モデル支援型デザインでは、収集したすべての毒性発現データから、各用量の毒性発現確率 p_j ($j = 1, 2, \dots, J$) を推定します。毒性発現確率の推定値と用量の間に単調性が認められない場合、Pool-Adjacent-Violators Algorithm⁴⁾ (PAVA) により毒性発現確率の推定値を補正して最大耐用量を選択します。PAVA による毒性発現確率の補正及び最大耐用量の選択手順を次に示します。

手順 1 1 例も投与されていない、または安全性の観点から適用する除外基準を満たす用量水準を最大耐用量の候補から除外する

手順 2 最大耐用量の候補である用量水準 j について、用量水準 j での投与人数 n_j と毒性発現人数 m_j を用いて、毒性発現確率の推定値 \hat{y}_j を次のように求める

$$\hat{y}_j = \frac{m_j + 0.05}{n_j + 0.1}$$

手順 3 単調性を満たす場合、毒性発現確率の推定値 \hat{y}_j を最終的な毒性発現確率の推定値 \hat{p}_j とする。単調性を満たさない場合、違反する用量水準間の毒性発現確率の推定値をベータ分布の分散の逆数を用いて、毒性発現確率の推定値を補正して、最終的な毒性発現確率の推定値 \hat{p}_j とする

$$\hat{p}_j = \hat{p}_{j+1} = \frac{\frac{1}{V(\hat{y}_j)} \hat{y}_j + \frac{1}{V(\hat{y}_{j+1})} \hat{y}_{j+1}}{\frac{1}{V(\hat{y}_j)} + \frac{1}{V(\hat{y}_{j+1})}}$$

$$V(\hat{y}_j) = \frac{(m_j + 0.05)(n_j - m_j + 0.05)}{(n_j + 0.1)^2 (n_j + 0.1 + 1)}$$

手順4 標的毒性発現確率 ϕ に最も近い毒性発現確率の推定値を持つ用量水準を最大耐用量と選択する。複数の用量水準が該当する場合、その中で最も低い用量水準を最大耐用量と選択する

手順2 では、通常よく用いられる最尤推定値 m_j/n_j とは異なり、ベータ分布 Beta (0.05, 0.05) を事前分布とした事後分布の平均を用いていることに注意します。最尤推定値 m_j/n_j を用いない理由としては、がん第1相臨床試験は、試験の参加者数が少ないため、各用量水準の投与人数 n_j は非常に少なくなることが想定されます。そのような場合には、最尤推定よりもベイズ流の推定の方が優れていることが知られているためです。

手順3 からわかるように PAVA による毒性発現確率の補正では、毒性発現確率の推定値と用量の間に「単調増加性」は保証されず、「単調非減少」を担保するにとどまっています。また、毒性発現確率の補正では $\hat{p}_j = \hat{p}_{j+1}$ となることから、補正した用量水準が標的毒性発現確率 ϕ に最も近い毒性発現確率となる場合、複数の用量水準が該当することになります。以上のことから、PAVA による補正はあまり精度が高くないことがわかるかと思えます。

そこで、筆者らは PAVA と比べて毒性発現確率の推定精度の向上が期待される Centered Isotonic Regression Algorithm⁴⁾ (CIRA) を用いて最大耐用量を選択する手順を提案しています。

CIRA による毒性発現確率の補正及び最大耐用量の選択手順を次に示します。手順1と2は、PAVA による毒性発現確率の補正及び最大耐用量の選択手順と同じであるため省略します。

手順3 単調性を満たす場合、毒性発現確率の推定値 \hat{y}_j を最終的な毒性発現確率の推定値 \hat{p}_j とする。単調性を満たさない場合、違反する用量水準間の毒性発現確率の推定値をベータ分布の分散の逆数を用いて、毒性発現確率の推定値を補正して、最終的な毒性発現確率の推定値 \hat{p}_j とする。さらに、用量についても同じ重みを用いて補正する。このとき、用量水準 $j+1$ は削除する

$$\hat{p}_j = \hat{p}_{j+1} = \frac{\frac{1}{V(\hat{y}_j)} \hat{y}_j + \frac{1}{V(\hat{y}_{j+1})} \hat{y}_{j+1}}{\frac{1}{V(\hat{y}_j)} + \frac{1}{V(\hat{y}_{j+1})}}$$

$$\hat{d}_j = \frac{\frac{1}{V(\hat{y}_j)} d_j + \frac{1}{V(\hat{y}_{j+1})} d_{j+1}}{\frac{1}{V(\hat{y}_j)} + \frac{1}{V(\hat{y}_{j+1})}}$$

$$V(\hat{y}_j) = \frac{(m_j + 0.05)(n_j - m_j + 0.05)}{(n_j + 0.1)^2 (n_j + 0.1 + 1)}$$

手順4 補正後の投与用量の定義域が試験計画時の定義域よりも小さくなっている場合、試験計画時の定義域と同じになるように補正する

手順5 標的毒性発現確率 ϕ に最も近い毒性発現確率の推定値を探索する。それに対応する \hat{d}_j を \bar{d} とする。各用量 d_j の中から \bar{d} に最も近い用量水準を最大耐用量と選択する

CIRA では、毒性発現確率の推定値と用量の間に「単調増加性」が保証され、さらに PAVA とは異なり用量に関する情報も使用するため、最大耐用量を正しく選択する可能性が高くなることが期待されます。

具体例を用いて、PAVA と CIRA にどのような違いが生じるのかを説明したいと思います。標的毒性発現確率を 0.3、1 コホート 3 例、コホート数 10 (総患者数 30 例)、投与用量を $(d_1, d_2, d_3, d_4, d_5) = (5, 7, 12, 15, 18)$ とします。この設定で試験を実施した結果、各投与用量に対応する累積投与人数 $(n_1, n_2, n_3, n_4, n_5) = (3, 9, 12, 6, 0)$ 、累積毒性発現人数 $(m_1, m_2, m_3, m_4, m_5) = (0, 3, 4, 3, 0)$ が得られたとします。PAVA では、毒性発現確率の推定値は $(\hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{y}_4) = (0.02, 0.335, 0.334, 0.50)$ 、最終的な毒性発現確率の推定値 \hat{p}_j は $(\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4) = (0.02, \mathbf{0.33}, \mathbf{0.33}, 0.50)$ となり、最大耐用量は 7 を選択します。CIRA では、毒性発現確率の推定値は $(\hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{y}_4) = (0.02, 0.335, 0.334, 0.50)$ 、最終的な毒性発現確率の推定値は $(\hat{p}_1, \hat{p}_2, \hat{p}_4) = (0.02, \mathbf{0.33}, 0.50)$ 、補正した用量は $(d_1, \bar{d}, d_4) = (5, \mathbf{9.82}, 15)$ となり、最大耐用量は 12 を選択します。このように、PAVA と CIRA では、最大耐用量の選択結果が異なることがあることに注意します。

実際に、PAVA と CIRA の性能を比較した結果を次節にて紹介します。

【表4】数値実験のシナリオ

シナリオ	d_1	d_2	d_3	d_4	d_5	d_6
シナリオ1	20	22	25	28	34	40
用量	0.26	0.30	0.36	0.41	0.50	0.60
毒性発現確率	0.11	0.26	0.32	0.41	0.50	0.60
シナリオ2	12	20	23	28	34	40
用量	0.11	0.26	0.32	0.41	0.50	0.60
毒性発現確率	0.06	0.11	0.26	0.32	0.50	0.60
シナリオ3	7	12	20	23	24	28
用量	0.26	0.27	0.32	0.36	0.39	0.40
毒性発現確率	0.11	0.26	0.29	0.32	0.35	0.40
シナリオ4	12	16	20	23	25	28
用量	0.11	0.18	0.26	0.30	0.35	0.40
毒性発現確率	0.11	0.18	0.26	0.30	0.35	0.40

過毒性な用量

■ PAVAとCIRAの比較

数値実験によりPAVAとCIRAを比較した結果を紹介します。総患者数を36例、1コホート3例、コホート数12として、用量水準数を6、標的毒性発現確率を $\phi=0.26$ とします。数値実験で用いた用量水準および毒性発現確率のシナリオを【表4】に示します。

各シナリオにおいて、マゼンタ色で示した用量は、真の最大耐用量を意味します。1つのシナリオに対して、10000回試行することになります。つまり、乱数により10000回の臨床試験データを生成して、PAVAとCIRAにより最大耐用量を選択した結果を比較することになります。

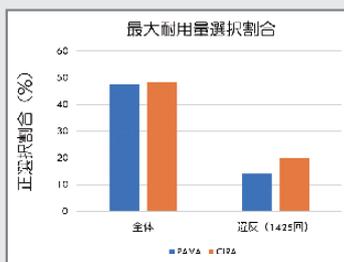
評価指標は、「真の最大耐用量を正しく選択した割合（以下、正選択割合）」と「真の最大耐用量よりも高い用量を選択した割合（以下、過毒性選択割合）」とします。ここで、用量水準 d_j を選択した割合を次のように定義します。

$$\frac{\text{用量水準 } d_j \text{ を選択した回数}}{\text{試行回数}} \times 100(\%)$$

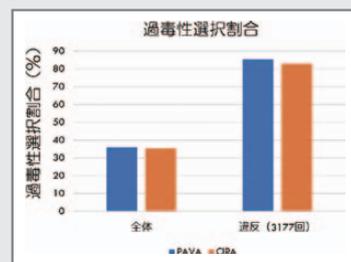
正選択割合は値が高いほど精度が高い、過毒性選択割合は値が低いほど精度が高いことを意味します。

シナリオ1から6に対する正選択割合の結果を【図2】から【図7】、過毒性選択割合の結果を【図8】から【図13】に示します。

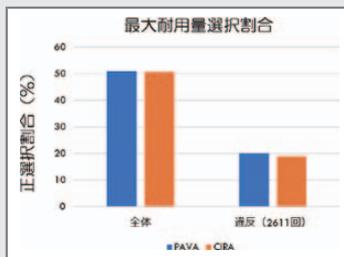
10000回の数値実験の結果を「全体」、単調性を満たさなかった場合の結果を「違反」としています。つまり、単調性を満たす場合は、PAVAとCIRAの結果は一致するので、「違反」の結果の違いが、PAVAとCIRAの性能の違いを表して



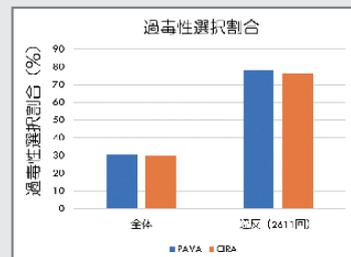
【図2】シナリオ1の正選択割合



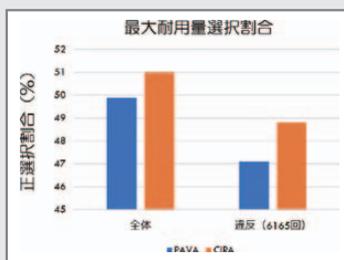
【図8】シナリオ1の過毒性選択割合



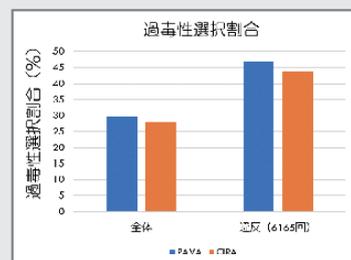
【図3】シナリオ2の正選択割合



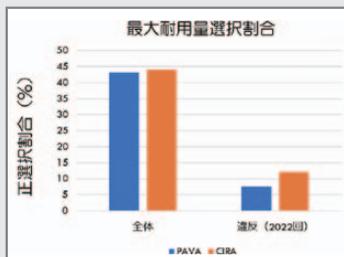
【図9】シナリオ2の過毒性選択割合



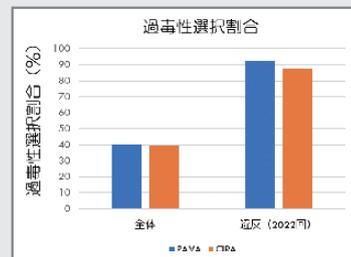
【図4】シナリオ3の正選択割合



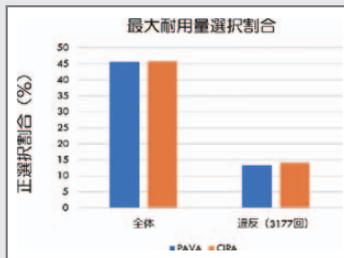
【図10】シナリオ3の過毒性選択割合



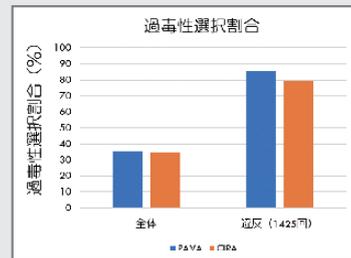
【図5】シナリオ4の正選択割合



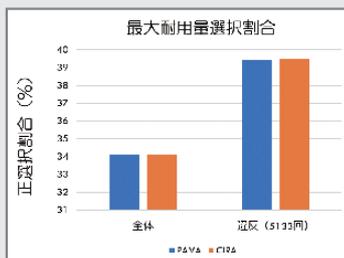
【図11】シナリオ4の過毒性選択割合



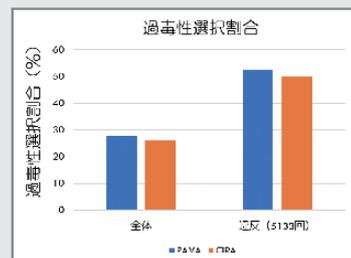
【図6】シナリオ5の正選択割合



【図12】シナリオ5の過毒性選択割合



【図7】シナリオ6の正選択割合



【図13】シナリオ6の過毒性選択割合

いることに注意します。「全体」では、正選択割合は PAVA よりも CIRA の方が約 0.5% 高く、過毒性選択は PAVA よりも CIRA の方が約 1.1% 低いという結果でした。一方、「違反」では、正選択割合は PAVA よりも CIRA の方が約 2% 高く、過毒性選択は PAVA よりも CIRA の方が約 3.5% 低いという結果でした。このように、CIRA では、毒性発現確率の推定値と用量の間に「単調増加性」が保証され、さらに用量に関する情報も使用するため、PAVA よりも性能が良くなるという結果になりました。僅かしか改善してないと思う方もいると思いますが、この分野ではこの数%の改善が大きな意味を持っています。

シナリオ 2 では、正選択割合は PAVA と比較して CIRA の改善は認められませんでした。このように、CIRA は PAVA の上位互換にはなっていないことが分かります。この原因としては、用量を補正する際に、毒性発現確率の補正と同じ重みを使用していることが挙げられるかと思えます。(他にも色々な原因がある可能性もあるので、現在色々な観点から検証中です。)

そのため、今後のさらなる展開としては、CIRA を改善していく予定です。

■ おわりに

新型コロナウイルス感染症により、この数年皆さんも医療に関心が高まっているのではないのでしょうか。ニュースなどでも新型コロナウイルス感染症のワクチン、治療薬に関する話題、臨床試験でどのような結果が出たのかが取り上げられたりしていたかと思えます。臨床試験において、被験薬の安全性・有効性を検証するために医療統計学が活用されています。

医学の発展により、医療統計学の方法論が発展することも多々あります。例えば、近年、がんの治療法としてがん免疫療法が注目されています。私たちの体は免疫の力によって、発生したがん細胞を排除していきますが、がん細胞は免疫の機能を弱める作用があります。がん免疫療法では、免疫ががん細胞を攻撃する力を保つようにし、免疫本来の力を利用してがんを攻撃する治療法です。このように、従来の治療法より副作用なども少なく、有効性も高いという結果が得られています。2018年にノーベル医学・生理学賞を受賞した京都大学の本庶佑先生の研究成果を基に、がん免疫療法の代表例である免疫チェックポイント阻害薬「オプジーボ」が小野薬品工業株式会社により開発されています。このように、がん免疫療法の分野では日本の

貢献が大きいです。

がん第 2 相、第 3 相臨床試験において、被験薬の有効性を評価する際、生存時間解析がよく用いられます。生存時間解析は、関心のある事象(例えば、死亡やがんの進行など)が発現するまでの時間を対象とした方法です。抗がん剤では、どのくらい長く生きられるかが有効性の評価において重要な観点となるためです。

医療統計学における生存時間解析では、3種の神器と呼ばれる「カプラン・マイヤー法」、「ログラंक検定」、「Cox 比例ハザードモデル」の3つの手法を用いて解析することが一般的です。しかし、免疫チェックポイント阻害薬では、「ログラंक検定」、「Cox 比例ハザードモデル」が前提としている仮定を満たさないことが知られています⁵⁾。そのため、免疫チェックポイント阻害薬の有効性を評価するために、「ログラंक検定」、「Cox 比例ハザードモデル」を用いることが適切ではないため、適切な医療統計学の方法論に関する研究がここ数年で急速に発展しています。これからも、医学の発展により、既存の方法では上手くいかない事例も増えていくことが予想されます。

本稿では、がん第 1 相臨床試験における用量探索法を題材として、医薬品開発における医療統計学の貢献について紹介しました。医薬品開発に掛かる費用も年々高騰しているため、効率的かつ精度の高い臨床試験の実施が求められています。医薬品開発、さらには医療に対して、医療統計学がどのように貢献しているのか少しでも理解して頂けたのであれば幸いです。

参考文献

- 1) Araujo, D.V., et al. (2021), Contemporary dose-escalation methods for early phase studies in the immunotherapeutics era. *European Journal of Cancer*, 158, 85-98, doi: 10.1016/j.ejca.2021.09.016
- 2) Liu, S. and Yuan, Y. (2015), Bayesian optimal interval designs for phase I clinical trials. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64, 507-523, doi: 10.1111/rssc.12089
- 3) 若山, 安藤, 桃崎. (2023), がん第 1 相試験におけるモデル支援型デザインに対する最大耐用量の選択方法の提案. *応用統計学会 2023 年年会講演予稿集*, 21-26.
- 4) Oron, A. P. and Flournoy, N. (2017), Centered isotonic regression: point and interval estimation for dose-response studies. *Statistics in Biopharmaceutical Research*, 9, 258-267, doi: 10.1080/19466315.2017.1286256
- 5) Trinquart L., et al. (2016), Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials. *Journal of Clinical Oncology*, 34, 1813-1819, doi: 10.1200/JCO.2015.64.2488

