

## 論文の内容の要旨

論文題目 **A modified continual reassessment method and its applications focused on the calculation for the mean of the dose limiting toxicity occurrence rate**  
用量制限毒性発現率の期待値の計算方法に着目した  
修正 **Continual Reassessment Method** とその応用

氏名 石塚 直樹

### 1. はじめに

CRM(Continual Reassessment Method)は1990年にワシントン大学およびFred Hutchinson Cancer Research Centerに在籍していたO'Quigley *et al.*により発表された、がん等の予後不良の疾患に対する治療法についての第I相試験デザインである。本論文では、2章でがん第I相試験の要件を考え、3章では現時点で最も幅広く用いられている標準的なデザインと問題点をレビューし、その特性を計算例により検証した。4章では従来のCRMとその修正を提示し、従来法との優劣を示した。そして、5章では実例を通じてCRMを用いた第I相試験における利点と問題点を指摘した。

### 2. がん第I相試験に必要な要件

次のようながん患者の臨床試験を想定する。

- ・ 既に標準的な治療は有効ではなく、短期間で死亡するリスクが高い。
- ・ 評価する新しい療法は、高用量では毒性が高くて致命的で、低用量では薬効が期待できない。この状況では最良の用量レベルが患者に投与されるべきである。しかし、そのようながん患者を対象に最大耐用量を求める第I相試験には研究目的と患者のインセンティブにギャップがある。また、毒性を過小評価して高すぎるMTDを選択すれば、第II相試験でより多くの患者に致命的な毒性が発現する用量レベルが投与されることにもなりかねない。逆にもう少し高い用量レベルが選ばれたかもしれないのに毒性を過大評価して低すぎるMTDが選択された場合には、第II相試験では期待される有効率が得られないことにもなりかねない。

### 3. がん第 I 相試験の標準的なデザイン(3+3 デザイン)

標準的なデザインは、マウスの LD<sub>10</sub> の 1/10 を試験開始用量として 3 例に投与する。その結果、試験開始用量で用量規制毒性(DLT)が現れなければ増量して次のレベルの検討に進む。もし、試験開始用量で DLT が現れれば、さらに 3 例が同じ用量レベルで試験が続けられる。6 人中 1 人までの毒性発現なら次の用量レベルの検討に進む。増量した用量レベルで再び繰返し、2 人以上に毒性が現れたらそこで試験を止め、終了した用量を最大耐用量(MTD)としている。

ある用量レベルにおける DLT 発現率を想定すれば、二項確率から試験を中止せずに増量する確率と試験を終了する確率を計算することができる。さらに  $j$  段階目の用量レベルを考えたとき、用量レベル  $j$  が投与されるのは用量レベル 1 から用量レベル  $j-1$  まで増量と判断された条件付きとなる。低い用量レベルから順番に投与され、 $j$  番目の用量レベルで終了し MTD と判断される確率計算を導出した。そこで、試験開始用量と増量幅が MTD の選択に及ぼす影響を調べるために、いくつかの真の DLT 出現率と増量幅の設定で、MTD として各用量レベルが選択される確率関数を数え上げて計算した。さらに平均、標準偏差により、信頼性を正確さと精度の両面から評価をした。

その結果、全般的に増量幅が狭いほど、真の DLT 発現率が低い用量レベルで MTD として選択される確率が高いことがわかった。低用量レベルから始めて慎重に増量すれば、目標とする毒性レベル(例えば DLT 発現率 33%)に至るまでに MTD と判断して試験が終了してしまう可能性が高いことが示された。逆に試験開始用量レベルを高くすると目標を通過する可能性も高くなることを示していた。増量幅が狭いほど、また試験開始用量レベルが低いほど MTD として選択される用量レベルの平均 DLT 発現率が低かった。さらに、増量幅が広いほど、また試験開始用量レベルが低いほど標準偏差は大きかった。

### 4. CRM

CRM は 3+3 デザインでは得られなかった 2 つの大きな利点がある。1 つは DLT 発現率に関して定量的な解釈が可能であり、2 つめは試験前の用量反応関係に関する様々な事前情報を明示的に利用する点である。オリジナルの CRM は以下の通りとなる。

---

用量レベル :	$x_i \ (i=1, \dots, k)$
$j$ 番目の患者の応答:	$Y_j \begin{cases} 1 & \text{if toxic response} \\ 0 & \text{if no toxic response} \end{cases} \ (j=1, \dots, n)$
用量反応 (毒性) モデル:	$\psi(x_i, a)$ , 例えば、 $E[Y_j] = \tilde{\theta} = \psi(x_i, a) = \frac{\exp(3+ax_i)}{1 + \exp(3+ax_i)}$ (4.1)
$j$ 番目の患者のパラメータ $a$ についての事前分布 :	$f(a, \Omega_j)$ ただし、 $\Omega_j = \{y_1, \dots, y_{j-1}\}$ かつ $\int_0^\infty f(a, \Omega_j) da = 1 \ (j=1, \dots, n)$
$j$ 番目の患者が $i$ 番目の用量レベルにおける応答確率の事前平均:	$\theta_{ij} = \int_0^\infty \psi(x_i, a) f(a, \Omega_j) da \ (i=1, \dots, k)$ (4.2)
パラメータ $a$ の事前平均を用いた応答確率の近似的な事前平均, $\mu(j)$ :	$\theta'_{ij} = \psi\{x_i, \mu(j)\} \ (i=1, \dots, k), \ \mu(j) = \int_0^\infty af(a, \Omega_j) da$ (4.3)

---

そして、次のように始まる:

- (i) 用量反応 (毒性) モデル  $\psi(x_i, a)$  を仮定する。ただし、 $x_i, i=1, \dots, k$  は試験開始前に定められた用量レベルで、 $a$  はモデルのパラメータ。
- (ii)  $a$  の事前分布を仮定する。

(iii) 目標とする DLT の発現する確率  $\theta$  を定める。

(iv)  $j$  番目の患者に目標とする毒性レベルと用量レベルの距離  $\Delta(\theta'_j, \theta)$  に最も近い、用量レベル  $x(j)$  を割付ける。毒性の有無である二値応答  $y_j$  が観察されれば、パラメータ  $a$  の事後分布がベイズの定理に従った計算で更新される。この過程を事前に決めたサンプルサイズ、例えば 25 例、あるいは他の条件を満たすまで繰り返す。

そこで、DLT 発現率を  $\theta'_j$  の近似ではなく、事前分布の分散が大きく、かつ歪んでいることを考慮に入れている平均確率（期待値） $\theta_j$  を用いることを提案した。この提案により投与開始用量レベルと増量スピードに差があり、試験開始直後に危険な増量を防ぐことが可能である。また、モデルのパラメータ  $a$  に代えて DLT 発現率  $\tilde{\theta}$  の分布をモニターすることを提案し、 $\tilde{\theta}$  の分布は  $a$  から単純な確率変数の変数変換として導出することを示した。

さらに、この提案を 1000 回のシミュレーションにより評価した。評価項目は 25 例により最終的に選択された用量レベル、真の目標となる用量レベルの事後確率、各用量レベルで試験された例数（割付けられた患者数）、DLT 発現の観察された例数とした。結果は、最終的に選択された用量レベルにおいて  $\theta_j$  に基づいた割付けルールと  $\theta'_j$  に基づいた割付けルールで変わらず、高用量レベルが投与される症例数を  $\theta_j$  に基づいた割付けルールが  $\theta'_j$  に基づいた割付けルールに比べて減らしていた。

## 5. CRM を用いた事例

シミュレーションに加えて、事前情報を活用することを個別の事例により考察した。

### 5.1 JCOG9512 試験

がん治療では、2 剤以上の薬剤による併用化学療法が一般的に用いられている。各薬剤の推奨用量で毒性と有効性について既に第 II 相試験で確認されているが、併用という状況では毒性が増強され単剤の推奨用量を投与することが困難になる可能性がある。そのため、併用の MTD あるいは推奨用量を決定するために第 I / II 相試験と呼ばれる新規の臨床試験が必要となる。日本臨床腫瘍研究グループの JCOG9512 試験は、進行非小細胞肺癌を対象として 3 剤併用の化学療法を CDDP の用量レベルを固定し、CPT-11 と VP-16 の 2 薬剤で適当な用量レベルを探索するために計画された。単剤の事前情報を反映した用量反応モデル、事前分布を設定し、目標毒性レベルは 33% であった。増量は 5 段階の用量レベルのうち最低レベルから開始され、先に説明した修正 CRM に基づき各用量で少なくとも 3 例が投与された。また、用量レベル  $\Pr[\theta > 50\%] > 1/3$  の時には増量しなかった。この試験では中止規準として、

(1) 用量レベル間で DLT 発現率の事後密度関数がきれいに分離できれば試験を終了できる。あるいは、分離ははっきりしていてもその範囲に目標とするレベルが存在しなければ新しい用量レベルを定めるべきである。

(2) 目標とする DLT 発現率のベイズ流信頼区間が十分に狭ければ試験を終了できる。  
を提案し、20 例で終了した。

### 5.2 TOP-53 試験

この試験は、CRM と Pharmacokinetic Guided Dose Escalation (PGDE) を組合せたハイブリッドなデザインであった。TOP-53 は、がん治療薬として一般的な VP-16 のアナログ化合物で、TOP-53 の薬物動態はマウスと犬で、線型であった。動物では骨髄抑制が DLT で、AUC と毒性の強い相

関が観察された。この現象はヒトでも同様であると期待された。毒性プロファイルも VP-16 の動物実験と同様であり、VP-16 の第 I 相試験のデータを事前情報として用いることができた。

用量反応モデルには 1 パラメータのロジスティック・モデルを仮定し、傾きのパラメータにガンマ事前分布を採用した。試験開始用量はマウスの LD<sub>10</sub> レベルの 1/10 (5.7mg/m<sup>2</sup> = 1n) に決定された。マウスの LD<sub>10</sub> の 40% に達する用量レベルまでは PGDE を用いて 100% の増量を行い、それ以後は 33% の増量を計画した。AUC と用量の関係を線型モデル (ただし切片はなし) と傾きのパラメータに曖昧な事前分布を仮定し、次用量レベルの AUC を事後 t 分布を用いて推定した。10 例目以降で CRM の計算を始め、減量の可能性も考慮に入れた。MTD の目標毒性レベルは 33% とした。試験期間中に Pr[ $\theta > 0.33$ ] がモニターされた。

Modified Fibonacci を用いた 3+3 デザインに比べて少ない患者数の試験になり、試験は 25 例で終了した。ただし、1 例は試験薬の投与を受けず、4 例に DLT が観察された。DLT 発現率に関して、Pr[ $\theta > 0.33$ ] は 14.1n と 18.8n で、それぞれ 7%、70% であり、確率密度関数間で十分に分離していた。

## 6. まとめ

オリジナルの CRM は、目標より高い用量レベルを割付ける傾向がある。ここで提案した修正バージョンは、割付けルールを各用量レベルの DLT 発現率の正確な期待値 (平均)  $\theta_j$  に基づいている。シミュレーション結果から、最終的に選択される用量レベルがオリジナル CRM と同等であり、一方で高用量レベルが投与される症例数を減らすことが確認された。また、DLT 発現率が特定の毒性レベルを超える確率の事後確率を用いることを提案した。さらに安全性モニタリングと、用量レベル間における密度関数の分離の判断にこれらが有用な尺度であることを示した。

実際の第 I 相試験では何らかの事前情報が存在することが一般的であり、単剤の治療のデータを事前情報として活用することができる併用療法を対象にした第 I/II 相試験、アナログ化合物の第 I 相試験に適用した実例を提示した。このような実例を通じて、事前情報をどのように定式化するのか、そのうち一般化できることは何なのかを経験的に知識を集積することが重要と考える。CRM は、増量・減量や試験の中止といった複雑な意思決定を必要とするような場面で、早期がん臨床試験におけるモニタリング・ツールとして、あるいは臨床家と統計家のコミュニケーションのツールとしてフレキシブルに用いることが可能である。

要約すると、CRM の適用にはモデルの設定と事前分布の選択に関心を払うべきである。そのような努力は臨床家と科学的なデータに基づくべきである。そして、このような姿勢こそが、研究と倫理のジレンマの存在する困難な分野で、新規のアプローチに必要であると確信している。