

博士論文（要約）

An efficient adaptive sample size determination  
in confirmatory randomized clinical trials  
with slow patient accrual rate

（症例登録速度が緩やかな検証的ランダム化試験における  
効率的な適応的症例数設計法の提案）

柏原康佑

## 論文の内容の要旨

論文題目     An efficient adaptive sample size determination in confirmatory randomized clinical trials with slow patient accrual rate  
(症例登録速度が緩やかな検証的ランダム化試験における効率的な適応的症例数設計法の提案)

氏       名     柏   原     康   佑

試験治療が新たな標準治療として認められるためには、その安全性と有効性が適切に計画・実施された一連の臨床試験によって探索・検証されなければならない。最終段階にあたる臨床試験は検証的試験と呼ばれ、試験治療の有効性および安全性を検証することがその目的である。検証的試験の試験デザインは、試験治療を対照治療と比較するランダム化比較試験が最も標準的である。

検証的試験における目標症例数は、試験治療の有効性を有意水準  $\alpha$  の検定によって判断するものとして、臨床的に意味のある治療効果の大きさ  $\mu_1$  に対する検出力  $1 - \beta$  が十分な大きさ（通常 0.8~0.9）になるように決定される。このとき、 $\mu_1$  に加えて臨床的には直接興味のないその他のパラメータ（局外パラメータ）も決定しなければならない。

局外パラメータについては、当該疾患に対する過去の経験があればある程度の精度をもって決定できる可能性がある。一方で、試験治療の対照治療に対する効果の大きさについては、それを調べるために検証的試験を実施するのだから、原理的に十分な情報は存在しない。このように、目標症例数の決定は不確実性を伴うことを避けられない。特に、希少疾患のように当該疾患に対する先行研究が豊富でない場合には、不確実性の程度が甚大になりえる。これは、検証的試験の計画・実施の大きな障害になる。

不確実性の原因は情報の不足であるから、その対処法は新たな情報の収集以外にない。対処法には、パイロット試験の実施、逐次デザイン・群逐次デザインの利用、適応的デザインの利用の3つがある。この中であらゆる不確実性に柔軟かつ科学的に対応できる方法は適応的デザインのみである。しかし、適応的デザインの基礎理論はすでに確立されているものの、中間解析時に症例数設計を大幅に更新した場合に検定効率が大きく損なわれることが知られており、大きな不確実性に対する対応能は十分ではない。そこで、本研究で

は検証的試験のデザインとして、検定効率を大きく損なうことなく、当該試験の蓄積データを活用しながら目標症例数を柔軟に決定可能な適応的症例数設計法を提案する。

まず、試験実施の枠組みとして「目標症例数はあらかじめ決定せず、試験開始時には少なくとも1回の中間解析を計画し、中間開始時には試験の中止・継続、さらに継続の場合には次の解析が最終解析か中間解析か否かのみを決定する」デザインを提案し、これを「加法的適応的デザイン」と定義する。加法的適応的デザインは、実際には適用しない作業デザインを準備し、条件付（第1種の）過誤確率保存則を通してそれを利用することによって実現可能である。本研究では、最も古典的な逐次デザインである sequential probability ratio test (SPRT) を作業デザインとした SPRT に基づく加法的適応的デザインについて整理し、その性質を明らかにする。

当該試験における検定統計量を  $X = \{X_t\}_{t \geq 0}$  と表す。添え字  $t$  は当該試験の開始からの情報量を表す。この  $t$  の上限  $T$  を決めることを症例数設計と定義する。第  $k$  回中間解析を実施する時点  $t_k$  としたとき、時点  $t_{k-1}$  から時点  $t_k$  にかけての検定統計量の増分は当該区間外の検定統計量と独立でありかつ期待値  $\mu(t_k - t_{k-1})$ 、分散  $(t_k - t_{k-1})$  の正規分布に従うと仮定する。

SPRT に基づく加法的適応的デザインの実施手順は次の通りである。まず試験開始時において第0次作業デザインとして第0次 SPRT を一意に定め、試験を開始する。次に、症例登録速度など試験の進捗を勘案しながら第1回中間解析を実施する。このとき、まず現在の時点  $t_1$  を求め、 $t_1$  まで解析を実施しなかったことを考慮するために作業デザインを

（当該時点の検定統計量  $x_{t_1}$  とは独立に）第1次 SPRT に更新する。その下で、試験の検定統計量の確率過程に対して作業モデルとして Wiener 過程を導入すれば、実現された検定統計量  $x_{t_1}$  に条件付けた第1次 SPRT の下での正確な条件付過誤確率  $\varepsilon_1$  を算出可能である。もし  $\varepsilon_1$  が1であれば当該中間解析において試験を中止し試験治療の有効性を宣言する。一方で、 $\varepsilon_1$  が1より小さければ試験を継続する。継続の場合は、条件付過誤確率保存則により第1回中間解析以降を有意水準  $\varepsilon_1$  の新たな試験とみなせる。よって、上記手順を繰り返して同様に試験を進めることができる。ある中間解析において試験が継続されたときに、次の第  $K$  回目の解析を最終解析に決定したとする。このとき最終中間解析における条件付過誤確率  $\varepsilon_{(K-1)}$  を用いて、最終解析では有意水準  $\varepsilon_{(K-1)}$  の検定を実行する。

第  $(K-1)$  回中間解析までに蓄積された当該試験自身のデータから、試験治療の治療効果の大きさについて試験開始前よりも精度の高い情報を得ることができる。これに基づいて、当該試験において検出力を確保すべき治療効果の大きさ  $\mu_1$  を定める。第0次から第  $(K$

- 1)次までの作業デザインについて SPRT の切片以外をすべて共通にして導かれるデザインを、「一致したデザイン」と定義する。このとき、本研究にて示されるとおり、一致したデザインの全体検出力の最終中間解析時点  $t_{k-1}$  に条件付けた上での下限を、第 0 次作業デザインから近似的に求めることができる。そこで、この下限が治療効果の大きさ  $\mu_1$  に対して 0.8~0.9 など希望の値になるように目標症例数  $T$  を定めることを提案する。

本研究では、全体検出力の下限の近似性能の評価と、提案したデザインの動作特性の評価を数値計算により行い、さらにすでに終了している実際の検証的試験に対して本方法の適用を試みた。近似性能の評価について、有意水準  $\alpha = 0.000001 \sim 0.999999$ 、目標症例数  $T = 0.2 \sim 60$ 、最終中間解析時点を  $T^*$  としたときそのタイミング  $R = T^* / T = 0.000001 \sim 0.999999$ 、真の治療効果の大きさ  $\mu = 0.5 \sim 5$  の各組み合わせについて調べた。評価は、最も粗い ( $K=2$ ) SPRT に基づく加法的適応的デザインの全体検出力と最も細かい ( $K=\infty$ ) 同デザインの全体検出力の比によって行った。動作特性の評価について、有意水準片側  $\alpha = 0.025$ 、全体検出力  $1 - \beta = 0.4 \sim 0.9$ 、真の治療効果の大きさ  $\mu = 0.5 \sim 5$  の各組み合わせについて調べた。評価は、提案法によって定めた目標症例数  $T$  の、試験開始時に  $\mu_1$  が既知であれば計算できたであろう固定症例数デザインによる目標症例数からの増分（差および比）によって評価した。

近似性能の評価では、最終中間解析のタイミング  $R$  が 0.99 より前であれば、最も粗いデザインの全体検出力の最も細かいデザインのそれに対する比が 99.4% を下回ることはなかった。動作特性の評価の結果を図に示した。図(A)より、真の治療効果の大きさと作業デザインである SPRT の傾き  $\rho_0$  がある程度等しい場合に目標症例数の増分が少なく提案法が非常に効率的であること、および真の効果の大きさが  $\rho_0$  から乖離するほど効率が低下することが示された。しかし、図(B)より真の効果の大きさが  $\rho_0$  から増加する方向に乖離している場合には、絶対値としての増分はほぼ変わらないことが示された。

提案した方法は、中間解析時点において制約なくデザインを修正できるクラスの適応的デザインにおいて、全体検出力を保証できる初めての方法である。提案した全体検出力の下限は、近似性能の評価の結果から正確ではないものの近似として十分であることが示された。また動作特性の評価から、治療効果の大きさについてあらかじめ真値が既知であれば  $\rho_0$  をおよそそれと等しくすることでより効率のよい一致したデザインを構築できるものの、真値が極めて不確実であっても症例数の増分は一般的な規模のパイロット試験よりも少ない程度であることが示された。

理論的には、提案法はすべての検証的試験に適用することができる。しかし、その性能は試験における症例登録速度が比較的遅い試験において最もよく発揮される。よって、希

少疾患などこれまで臨床試験の計画が困難であった領域への適用が期待される。

一方で、適応的デザインが有する欠点を提案法はすべて継承している。1つは、実施組織体制の複雑化である。適応的デザインでは、一般的な中間解析が必要とされる独立データモニタリング委員会に加えて、有効性の中間評価および適応の要否の判断を行うための第三者機関がさらに必要であるとの意見もある。また、臨床試験の原則として試験計画時には症例数を含め入念な検討を行うべきである点是不変であり、適応的デザインはその簡略化を促すものではない。

提案法の動作特性についての解析的裏づけ、作業デザインとして SPRT 以外の逐次デザインの応用、多変量正規分布以外への拡張、症例数以外の適応的デザインへの応用は今後の課題である。

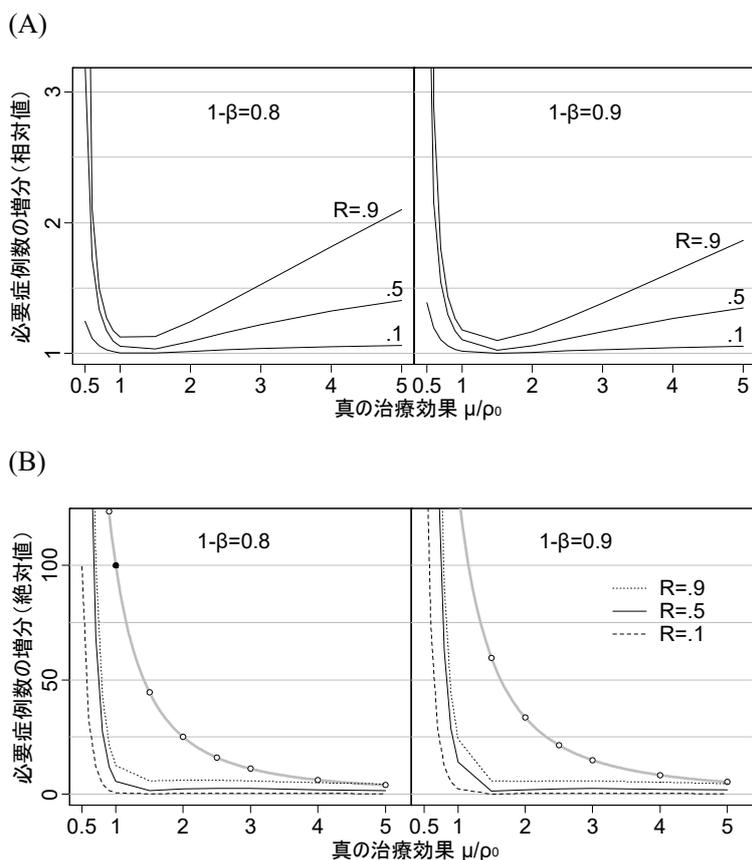


図. 提案法による目標症例数と固定症例数デザインによる目標症例数の比較 ((A)は比、(B)は差)。 $\rho_0$ は作業デザインである SPRT の傾きを表す。検出力  $1-\beta$  と最終中間解析のタイミング  $R$  についていくつかの組み合わせを示した。(B)の灰色線は固定症例数デザインにおける目標症例数を表し、縦軸はそれが  $\mu=1$  および  $1-\beta=0.8$  のとき 100 になる (●にて示した) 尺度とした。