

しなやかな臨床試験デザイン： 適応型デザインによる効率化

竹 林 由 武

福島県立医科大学

Improving efficiency with adaptive trial designs

Yoshitake TAKEBAYASHI

Fukushima Medical University

Clinical trials are prospective studies that evaluate the effectiveness of interventions on humans under certain circumstances. The frequentist approach in which the sample size is strictly controlled before tests is common in clinical trials. However, clinical trials that use an adaptive Bayesian method in which the trial can be flexibly stopped, based on data accumulated during the course of the trial, has been recommended to reduce costs and to meet ethical requirements in the fields of medicine and medical device development. In the field of psychological interventions, the current situation is that application based on Bayesian framework is often extremely poor. Therefore, this paper outlines the design method of clinical trials by the adaptive Bayesian method, discusses the benefits of and problems in its application to psychological intervention research, and provides examples of its virtual application.

Key words: clinical trial design, Bayesian approach, posterior probability, predictive probability, adaptive randomization, prior distribution

キーワード：臨床試験デザイン、ベイズ統計学流、事後確率、予測確率、適応的ランダム化、事前分布

1. はじめに

臨床試験は、事前に設定された特定の状況下で、人を対象に実施された介入の効果を評価する前向き研究である。医薬品の承認は、第Ⅰ相から第Ⅲ相まで、探索的な段階から検証的な段階へとステップを踏んで実施された臨床試験の結果に基づいて行われる。探索的な段階である第Ⅰ相、第Ⅱ相試験では、新規治療薬の推奨容量の確定と、短期的な安全性と有効性の検証を行う。第Ⅲ相試験では、第Ⅱ相試験で有効性の示された新規治療薬を標準治療等と比較し新薬の有効性が検証される。また第Ⅱ相試験は、単群の前後比較試験である第Ⅱ相 a、小規模の無作為化比較試験である第Ⅱ相 b に分類される。心理社会的介入に関する臨床試験は、薬剤のように処方容量の調整が困難であるため第Ⅱ相試験からスタートし、第Ⅲ相の無作為化比較試験で有効性の検証が行われるのが一般的である。

臨床試験の研究計画は、試験実施前に例数が厳密に制御される頻度主義的統計学流アプローチが主流である。例えば、先行研究から想定される新規治療群と標準治療群の効果の差（治療反応率差または平均値差）と任意の有意水準および検出力のもとで、試験に必要な例数が設定される。試験前に設定された例数は、中間解析で明らかに新規介入が無効であると判断され中止になる場合（無益性による試験の中止）を除いて、試験終了時まで変更されない。事前の想定よりも、新規治療の効果が高い場合には、事前に設定された例数よりも少ない例数でも有効性を示すことが可能になるため、試験にかかるコストが浪費される。また、そのようなケースでは、不必要に多くの人が無効性の劣る標準治療を受けることになるという倫理的な問題が生じる。従来の臨床試験の実施にかかるコストおよび倫理的な問題を改善し、試験進行中に蓄積されるデータに基づいて柔軟に試験の継続や中止を判断する適応的な臨床試験デザインの

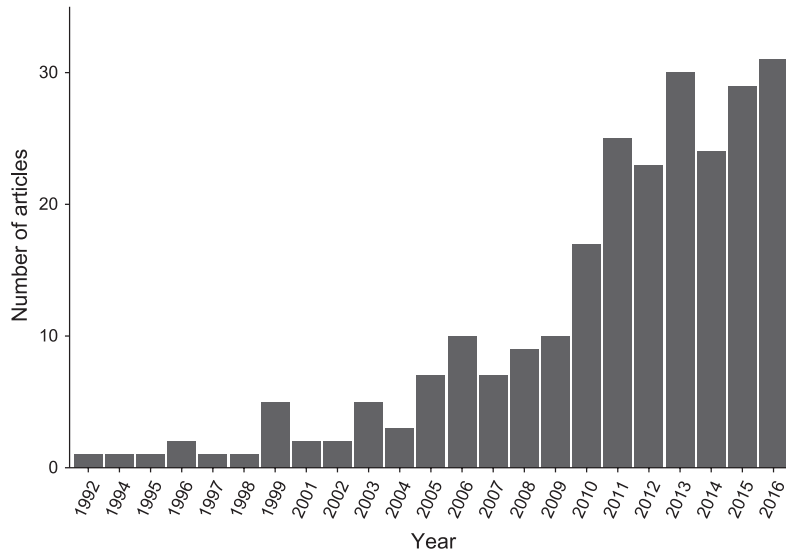


図1 ベイズ統計学流適応的デザインに関連する刊行論文数の年次推移

手法が開発されてきた。

適応的な臨床試験デザインの手法は、ベイズ統計学流のアプローチとの親和性が高く、医薬品開発の臨床試験での活用が期待されている。2018年2月28日の時点で、検索エンジンPubMedにて、「ベイズ (bayes)」「適応的 (adaptive)」「試験 (trial)」を検索ワードとし、1900年から2016年に刊行された文献を検索した結果、246件がヒットした(図1)。図1から、2000年から2016年にかけて文献数が増加しており、特に2010年以降の増加が顕著である。2008年、2009年には、ベイズ統計学流の適応的ランダム化を適用したBATTLE-StudyやI-SPYIIの試験デザインが公表され(Barker et al., 2009; Zhou et al., 2008)、その翌年2010年2月には、アメリカの食品医薬局(FDA: Food and Drug Administration)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(Guidance for the Use of Bayesian Statistics in Medical Device in Clinical Trials)を公表している(手良向, 2017; U.S. Department of Health and Human Services & Food and Drug Administration, 2010)。FDAのガイダンスでは、ベイズ統計学流のアプローチを臨床試験デザインに用いるメリットとして、1) 意思決定のためにより多くの情報を用いることが可能になること、2) 事前情報の利用によって標本サイズが減少すること、3) 適

応的試験デザインによって標本サイズが減少すること、4) 試験デザインの中間変更が可能になることが指摘されている。これらは、医薬品開発だけに関わらず、心理学的介入法の有効性を検討するデザインまたは心理学研究全般に通ずる利点と言える。しかしながら、心理学的介入の分野では、ベイズ統計学流の試験デザインの適用は極めて乏しいのが現状である。

そこで、本稿では、心理学的介入法の有効性を検討する仮想事例をもとに、ベイズ統計学流の臨床試験デザインの種々の方法を概観する。まず介入による治療反応率の推測を例に、ベイズ統計学流の臨床試験デザインに関わる基礎的な枠組みを述べる。その上で、ベイズ統計学流の逐次的なモニタリング手法として、事後分布に基づく方法と予測確率に基づく方法を単群の前後比較試験の設定のもとで概観する。その後、ランダム化比較試験の設定のもとで、適応的ランダム化の方法を概観する。最後に、過去の試験情報を現行試験の推定に活用するための方法について概観する。

2. ベイズ推定の基礎

臨床試験の主な関心の一つは、ある集団に特定の治療を実施することによって、特定のアウトカム(例えば、症状、問題行動、QOL等)に有効な

改善が示されるか否かである。数名の患者にある治療を実施した際に、個々人における治療反応の有無を2値で記録（1：治療反応あり，0：治療反応なし）したデータ D が得られたとする。関心のあるパラメータ θ は、治療を実施した集団における治療反応率とする。データ D を観測したもとの、治療反応率 θ の事後分布 $P(\theta|D)$ は、

$$P(\theta|D) = \frac{P(D|\theta) \times P(\theta)}{P(D)}$$

とベイズの定理に基づいて推定される。

ここで $P(D|\theta)$ は尤度、 $P(\theta)$ は事前分布である。尤度 $P(D|\theta)$ は得られたデータ D が持つ θ に関する情報を反映しており、事前分布 $P(\theta)$ は、データを取得する前の θ に関する情報を反映する。データ取得前の情報と取得したデータの情報を併合した結果得られる $P(\theta|D)$ は、事後分布と呼ばれる。

治療反応率をベイズ統計学流の枠組みで推定する具体例を示す。そのために Ito et al. (2016) の単群前後比較臨床試験のデータを用いる。Ito et al. (2016) は、うつ病または不安症の診断基準を満たす15名の患者に、感情障害に対する診断横断的治療プロトコルに基づく認知行動療法を週1回、平均15回提供した。主要評価項目である SIGH-A（構造化面接により評価した不安症状）について、治療開始前後での有意意味な得点の変化の基準値を5点以上と設定し、その基準を超えた患者は本治療に反応したものとして算定された。結果として本試験では、15名中10名（67%）に治療反応がみられた。この臨床試験の結果を得られたデータ D とし、関心のあるパラメータである治療反応率 θ の事後分布 $P(\theta|D)$ をベイズ統計学流の枠組みで推定する。まず、得られたデータ D が2値の反応（1=反応あり，0=反応なし）である場合、データ D は二項分布に従う。治療反応率 θ の事前分布 $P(\theta)$ をベータ分布に設定すると、二項分布とベータ分布の共役関係から、治療反応率 θ の事後分布 $P(\theta|D)$ もベータ分布になる。ベータ分布は2つのパラメータ α と β から規定され、治療反応率を推定する文脈では、 α は治療反応が示された人数 n に、 β は治療が実施された対象者全体 N のうちで治療反応が示されなかった人数 $N-n$ に相当する。図2上段には、下記の設定で治療反応率のベイズ

推定を行った際の治療反応率の事前分布（図2上段左）、尤度（図2上段中央）、事後分布（図2上段右）を示している。まず事前分布に関して、国内での本介入の有効性について先行する情報はないものとして、ベータ事前分布のパラメータ α および β を1に指定する $Beta(\alpha=1, \beta=1)$ 。図2左上にベータ事前分布（ $Beta(\alpha=1, \beta=1)$ ）が示されており、この場合、治療反応率 θ は特定の範囲にピークをもたず漠然としており、これはどれくらいの反応率であるかについて情報を持っていないことを反映している。15名中10名に有効な治療反応が示されたので、尤度は治療反応率 θ を67%、治療を行った総数 N を15とする二項分布（ $Binomial(\theta=0.67, N=15)$ ）となる（図2）。治療反応率の事後分布 $P(\theta|D)$ は、ベータ事前分布の α と β に、試験結果から得られた情報（ n =治療反応を示した人数、 $N-n$ =治療反応を示さなかった人数）を加えることで求められる。つまり、治療反応率のベータ事後分布の α はベータ事前分布の $\alpha+n$ となり、ベータ事後分布の β はベータ事前分布の $\beta+N-n$ となる。したがって本データでの治療反応率のベータ事後分布は $Beta(\alpha=1+10, \beta=1+5)$ となる（図2上段）。

事後分布から、パラメータの値について任意の範囲で積分することで、その範囲に含まれるパラメータ値が得られる確率を求めることができる。例えば、図2上段右の治療反応率の事後分布について、パラメータ（治療反応率）が50%以上、60%以上になる確率をそれぞれ求めると、90%、67%となる。このように、ベイズ統計学流の枠組みでは、事後分布を任意の基準に照らし合わせることで、さまざまな意思決定を柔軟に行うことが可能となる。

図2中段および下段には、図2上段と得られたデータ D が同じ状況のもとで、異なる形状の事前分布を設定し治療反応率を推定した結果を示している。Yin (2012) は事前分布を無情報事前分布、楽観的事前分布、懐疑的事前分布、臨床的事前分布に大別している。無情報事前分布は、特定のパラメータの範囲にピークを持たない漠然とした分布であり（例えば、 $Beta(1, 1)$ ）、先に示した例のようにパラメータ θ がどの値をとるかについての情報を持っていないことを反映する。事後分布を推定する際に無情報事前分布を選択すると、

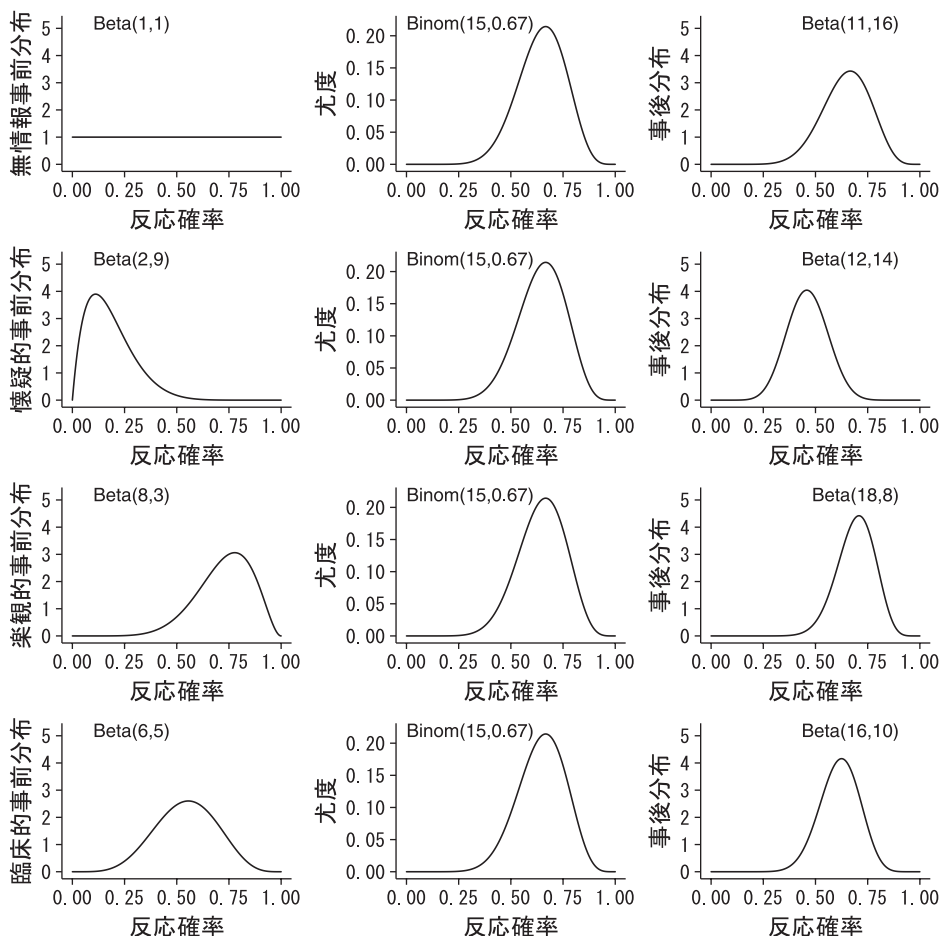


図2 事前分布が事後分布の推定に与える影響

事後分布は尤度のみの情報を反映することになる。楽観的事前分布では、治療反応率の分布のピークを高め、懐疑的事前分布では低めに設定する。臨床的事前分布は、複数の臨床家の経験に基づいたり、種々の先行情報から臨床的に妥当な、ある程度の客観性を持った分布として設定される。図2から、事前分布の設定が事後分布の推定結果に影響を与えることが明らかである。また、標本サイズが小さい状況ほど、事前分布が事後分布の推定に与える影響は大きくなる。そのため、ベイズ統計学流の枠組みでは、適切な事前分布を選択することが妥当な推定を行うために極めて重要な要件となる。

3. 単群の臨床試験

3.1 事後分布に基づくモニタリング

Thall and Simon (1994) は、治療反応率の事後分布を用いて試験の結果を逐次的にモニタリングする方法を提案している。ここでは、新規治療の反応率 p_E が、標準治療の反応率 p_S と比較される。 p_E および p_S の事前分布にはベータ事前分布を設定する。標準治療の反応率は先行研究のデータをもとに設定されるが、先行研究のデータが当該研究の推定結果に与える影響を割り引くために分散を大きく設定する場合がある。例えば、標準治療に関して、40名治療して20名の反応があった場合には、 $Beta(11, 11)$ として、20名分のデータとして割り引いた設定とする。ここでは、任意に半分のデータを割り引いたが、先行する類似の試験に

表1 事後確率に基づく無益性の中止境界

| | | | | | | | | | | | | | | | | |
|-----|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|
| 治療数 | 3 | 5 | 7 | 9 | 11 | 13 | 14 | 16 | 18 | 20 | 21 | 23 | 25 | 27 | 29 | 30 |
| 反応数 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |

表2 事後確率に基づく優越性の中止境界

| | | | | | | | | | | | | | | | | | | | |
|-----|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 治療数 | 3 | 4 | 5 | 7 | 8 | 10 | 11 | 13 | 14 | 16 | 17 | 19 | 20 | 22 | 24 | 25 | 27 | 28 | 30 |
| 反応数 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |

関する情報を当該試験の推定に統合する際の事前分布の設定方法は後の節で詳細を述べる。標準治療の事前分布は、標準治療と比べて情報の蓄積が乏しいため大きく設定される（例えば、 $Beta(2, 2)$ ）。得られたデータと事前分布から新規治療の反応率の事後分布が得られ、標準治療の治療反応率について設定された分布と比較される。ここで、現状の推定結果に基づいて、試験を継続するか中止するかを判断するための二つの基準を設定する。一つは、新規治療が標準治療よりも明らかに優れる見込みが高い場合、つまり新規治療の優越性により試験を中止する基準である。新規治療の反応率の事後分布から、新規治療の反応率が標準治療の反応率よりも高い確率 $P(p_E > p_S|y)$ が、事前に設定した任意の値 θ_U 以上になる最小の人数 U_n を求める。 θ_U は、一般的に 0.95～0.99 の高い確率が設定される（手良向, 2017; Yin, 2012）。 U_n 以上の治療反応が観察された場合、優越性により試験は中止され、 U_n 未満の場合には試験を継続する。もう一つの基準は、新規治療が標準治療よりも明らかに優れるという見込みが乏しい場合、つまり新規治療の無益性により試験を中止する基準である。この基準では、新規治療を標準治療と比較した場合の新規治療の最低限許容可能な反応率の増分 δ を設定し、新規治療の反応率の事後分布から新規治療の反応率がその増分 δ 以上である確率 $P(p_E > p_S + \delta|y)$ を求める。そして、事前に設定した任意の値 θ_L 以下となる最大の人数 L_n を求める。一般的には、 θ_L は 0.01～0.05 の範囲で低い確率が設定される（手良向, 2017; Yin, 2012）。 L_n 以下の治療反応が観察された場合、無益性により試験は中止され、 L_n 以上の場合には試験を継続する。つまり、この基準のもとでは新規治療が標準治療よりも事前に想定した増分以上に効果を示す見込みが極めて低いため、試験の継続が無益であると

判断され試験が中止される。

ここでは、治療反応率を事後分布に基づいて逐次的にモニタリングする上述の方法を、仮想的な心理学的介入法の単群前後比較試験に適用し例を示す。全般的な不安症に関して、有効性が示されている標準治療があり、先行研究から 100 名に治療を実施した際の治療反応率が 50% であることが示されているとする。新規治療として別の心理学的介入があり、試験の実施者は新規治療は標準治療よりも 20% 高い治療反応率を期待しているとする。標準治療と新規治療のそれぞれの反応率にベータ事前分布を設定する。標準治療の事前分布は、先行研究から情報を半分引き引いた事前分布 $Beta(25, 25)$ とする。新規治療の事前分布は $Beta(1.4, 0.6)$ とする。ここでは、ベータ事前分布の α と β の値は 70% の治療反応率 $(1.4/(1.4+0.6) = 0.70)$ を想定するとともに、値を小さく設定することで分散が大きくなるよう設定している。優越性による中止に関して $\theta_U = 0.95$ とし、無益性による中止に関して、 $\theta_L = 0.05$ かつ $\delta = .20$ として、最大標本サイズ $N = 30$ を採用し、試験を患者 1 名ずつモニタリングする。この設定のもとでの無益性による中止境界を表 1 に、優越性による中止境界を表 2 に示した。例えば、3 名中 1 名も治療反応が示されない場合、5 名中 1 名しか治療反応が示されない場合、7 名中 2 名以下しか治療反応が示されない場合、無益性による試験中止が可能になる。一方、5 名までは全員に治療反応が示される場合、7 名中 6 名に治療反応が示される場合、8 名中 7 名に治療反応が示される場合に優越性による試験の中止が可能になる。

事後分布に基づく逐次的なモニタリングは標本サイズが結果的に最大 N より大きくなった場合に「有効かつ無効」という矛盾した判断に直面する可能性があり、その意味であまり好ましい方法

とは言えない（手良向，2017）。

3.2 予測確率に基づくモニタリング

ベイズ統計学流の枠組みでは、観測したデータを所与としたときに将来得られるデータの分布である事後予測分布を求めることが可能である。Lee and Liu（2008）は、治療反応率の事後分布から、試験終了までに追加される参加者のデータについての事後予測分布を求め、臨床試験が成功する確率である予測確率（PP: Predictive Probability）に基づいて試験結果を逐次的にモニタリングする方法を提案している。例えば、最大例数を N とする臨床試験で n 名に新規治療をした時点で中間解析を行ったとする。その場合の治療反応数 X は、二項分布 $X \sim \text{Binomial}(n, p_E)$ に従う。新規治療の治療反応率 p_E の事前分布をベータ分布 $p_E \sim \text{Beta}(\alpha_0, \beta_0)$ とすると、治療反応率の事後分布は共役関係からベータ分布、 $\text{Beta}(\alpha_0 + x, \beta_0 + n - x)$ となる。中間解析後に募集される患者数 $N - n$ のうち、新規治療に治療反応を示す患者数を Y とする。新規治療に治療反応を示す患者数 Y はベータ二項分布、 $\text{Beta-Binomial}(N - n, \alpha_0 + x, \beta_0 + n - x)$ に従う。そして、将来のありうる治療反応数 $Y = i$ の事後分布 $P(p_E | X = x, Y = i)$ はベータ分布、 $\text{Beta}(\alpha_0 + x + i, \beta_0 + N - x - i)$ に従う。この試験では、「新規治療の治療反応率 p_E がある標準治療の反応率 p_S よりも高い確率が特定の閾値 θ_T よりも高い」という仮説を持っているとする。例えば、 $\theta_T = .95$ であれば、「新規治療の反応率が標準治療の反応率よりも高い確率が95%以上である」という想定をもつことになる。中間解析時の治療反応数 x と中間解析後に追加された参加者において観測された特定の反応数 $Y = i$ が観測されたもとで、新規治療の反応率（の事後分布） p_E が標準治療 p_S よりも高い確率 $P(p_E > p_S | x, Y = i)$ を Bi とする。予測確率 PP は、 Bi が特定の閾値 θ_T を超える Y_i について、事後予測分布から求められる治療反応数 $Y = i$ の生起確率を合計することで求められる。試験の中止規則は、 PP について、任意の上側、下側カットオフ θ_U と θ_L を設定し、 $PP > \theta_U$ の場合、新規治療が有望であるとして試験を中止、 $PP < \theta_L$ の場合、新規治療が有望ではないとして試験を中止、その他の場合には、試験を最大標本サイズ N に達するまで継続する。

先述した「治療反応率の事後分布に基づくモニタリング」で用いた状況と同様の設定で中間解析を行い得られた事後分布から予測確率 PP を算出する例を示す。最大例数 N が40名であり、中間解析時の人数 n は20名、治療反応数 X は14名であったとする。標準治療の反応率 p_S は.50と設定する。そして、治療の成功確率に関して、「95%以上の確率（ $\theta_T = .95$ ）で新規治療の治療反応率 p_E が標準治療の治療反応率 p_S よりも高い」ことを条件として予測確率を求める。無益性による中止に関して、 $PP < \theta_L = 0.05$ を採用する。先ほどと同様に、治療反応率の事前分布をベータ事前分布 $\text{Beta}(\alpha_0 = 1.4, \beta_0 = 0.6)$ とすると、治療反応率の事後分布は、 $\text{Beta}(1.4 + 14, 0.6 + 20 - 14)$ となる。中間解析後は試験終了までにあと20名のデータが得られ、その事後予測分布は、 $\text{Beta-Binomial}(20, 1.4 + 14, 0.6 + 20 - 14)$ となる。事後予測分布から、中間解析後に得られたデータのうちの治療反応を示す人数の割合が、それぞれの治療反応数について求められる。また、個々の治療反応数について、 $Bi = P(p_E > 0.50 | x, Y = i)$ が求められる。

上記の設定で算出した事後予測分布に基づく治療反応数 $Y = i$ の生起確率、 Bi を表3に示した。中間解析後に観測した20名のうち治療反応を示した人が11名以上である場合に、新規治療の反応率が標準治療の反応率.50を超える確率 Bi が90%以上となる。そこで、治療反応数が11以上の場合の治療反応数の生起確率を合計すると予測確率 PP が求められ、0.93となる。 $PP > 0.05$ であるため試験は継続される。

上述の例では、中間解析時での予測確率のみを求めたが、事後分布に基づくモニタリング手法と同様に、事前に最大例数を設定し、取得するデータ数と反応数の全ての場合について求めた PP に基づいて、 $PP > \theta_U$ となる最小の人数 U_n または $PP < \theta_L$ となる最大の人数を算定することで、無益性による中止境界と優越性による中止境界を定めることが可能となる。ここでは、事後予測分布を用いた逐次的なモニタリングの例で用いた設定と同じく、最大例数 N を30名、新規治療の事前分布を、 $\text{Beta}(1.4, 0.6)$ 、標準治療の反応率 p_S を50%、無益性のカットオフ確率 θ_U を0.95、優越性のカットオフ確率 θ_L を0.05に設定し、予測確率に基づくモニタリングの例を示す。上記の設定で、新規

表3 試験成功の予測確率の計算例

| Y=i | 事後予測分布 | Bi |
|-----|--------|--------|
| 1 | 0.0010 | 0.0400 |
| 2 | 0.0000 | 0.0749 |
| 3 | 0.0000 | 0.1309 |
| 4 | 0.0000 | 0.2088 |
| 5 | 0.0000 | 0.3087 |
| 6 | 0.0040 | 0.4256 |
| 7 | 0.0010 | 0.5495 |
| 8 | 0.0060 | 0.6683 |
| 9 | 0.0170 | 0.7732 |
| 10 | 0.0290 | 0.8551 |
| 11 | 0.0510 | 0.9151 |
| 12 | 0.0649 | 0.9540 |
| 13 | 0.0989 | 0.9770 |
| 14 | 0.1119 | 0.9900 |
| 15 | 0.1488 | 0.9960 |
| 16 | 0.1359 | 0.9980 |
| 17 | 0.1339 | 0.9990 |
| 18 | 0.1119 | 0.9990 |
| 19 | 0.0569 | 0.9990 |
| 20 | 0.0240 | 0.9990 |

治療の事後分布が標準治療よりも高い確率 B_i の閾値 θ_r を .90 とした場合の優越性および無益性による中止境界を表4, 表5に示した。例えば, 3名中1名も治療反応が示されない場合, 5名中1名しか治療反応が示されない場合, 7名中2名以下しか治療反応が示されない場合, 無益性による試験中止が可能になる。一方, 6名までは全員に治療反応が示される場合, 8名中7名に治療反応が示される場合, 9名中8名に治療反応が示される場合に優越性による試験の中止が可能になる。

ここまで述べた事後分布に基づく方法と予測分布に基づく方法は, 2群間の比較を行うランダム化比較試験にも拡張可能である (Yin, 2012)。その際, 単群試験で, 既存研究から値を設定して

いた標準治療群について実際の対照群のデータの事後分布を当てはめれば, 逐次的に試験結果を同じ枠組みでモニタリングすることが可能である。なお, 事後分布と予測分布に基づく逐次的な試験結果のモニタリングは, 2値のアウトカムから治療反応率を検討する単群前後比較試験において, データ解析環境 R で実行可能な ph2bayes パッケージを用いて容易に実施できる (Nagashima, 2016)。

4. 適応的ランダム化

第II相b試験や第III相試験では, 2群以上の群間無作為化比較試験が用いられる。無作為化比較試験では, 一般的に群への割付比率は群間で等しくなるよう設定される (2群間比較であれば, 50%でいずれかの群に割付けられる)。頻度主義統計学流の枠組みでは, 割付比は事前に固定され, 試験終了時まで不変となる。一方, ベイズ統計学流の枠組みでは試験結果が随時モニタリングされ, それまでに治療した患者の反応から治療反応率の事後分布を推定し, その事後分布は次の1名の試験結果を推定する際の事前分布として利用可能となる。このベイズ統計学流の枠組みを用いて, 蓄積される試験の推定結果に応じて割付比を調整する方法が提案されている。これらの方法は, ベイズ統計学流適応的ランダム化 (Bayesian adaptive randomization) と呼ばれ, BATTLE 研究や I-SPY 研究といった著名な研究の中でも用いられている (Barker et al., 2009; Zhou et al., 2008)。群間比較試験において, 適応的ランダム化を用いるメリットの一つは, 新規治療が有効である場合に, 試験の進行と共に参加者が有効性の劣る対照群の治療に割付られる確率が低くなるという性質にある。この性質は, 多くの参加者が有効性の劣る治療を受けなければいけないという倫理的な問

表4 予測確率に基づく無益性の中止境界

| 治療数 | 3 | 5 | 7 | 9 | 11 | 12 | 14 | 16 | 17 | 19 | 20 | 22 | 23 | 24 | 26 | 27 | 28 | 29 | 30 |
|-----|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 反応数 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |

表5 予測確率に基づく優越性の中止境界

| 治療数 | 3 | 4 | 5 | 6 | 8 | 9 | 10 | 12 | 13 | 15 | 17 | 18 | 20 | 22 | 24 | 25 | 28 |
|-----|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|
| 反応数 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |

題を軽減する。本節では、ベイズ統計学流の適応的ランダム化を用いた無作為化比較試験のデザイン方法について概要を述べる。

4.1 割付比率

2群の無作為化比較試験にベイズ統計学流の適応的ランダム化による群割付を行う場合、最大標本サイズ N に達するまでの間に任意の回数、それまでに蓄積されたデータからどちらの群が優勢であるかについて確率を求め、その確率に基づいて割付を行う。これが任意の回数繰り返されることで、試験の対象者は、試験の進行と共に高い確率で有望な治療へと割付けられるようになる。具体的には、新規治療群および標準治療群の治療反応率の事後分布を比較し、新規治療群の反応率 p_E が標準治療の反応率 p_S よりも高い確率 $\lambda = P(p_E > p_S \mid \text{解析時点で蓄積されているデータ})$ を求める。ただし、 λ の値はそのまま割付確率として用いられず、べき乗変換、

$$\pi(\lambda, \gamma) = \frac{\lambda^\gamma}{\lambda^\gamma + (1-\lambda)^\gamma}$$

を通じて、次の中間解析までに治療が行われる患者に適用される割付確率として利用される。べき乗パラメータである γ は調整パラメータであり、0 から 1 の範囲をとる。 γ の設定によって、割付確率がどのように変化するか、その挙動を図 3 に示した。 γ が 0 の場合、つまり γ による調整が行われない場合、 λ はその値に依らず 0.5 となる。すなわち、その時点でいずれかの群が優勢であっても、群の割付比率は 0.5 と一定に保たれる。 γ が 1 の場合、割付確率は λ に正比例する。これは、 λ の値がそのまま割付確率に反映されることを意味する。 γ を 1 から 0 に小さくしていくほど割付確率は 0.5 に近づいていく。また $\lambda = 0.5$ の場合、 γ の値に関わらず、割付確率は 0.5 となる。Thall and Wathen (2007) はべき乗パラメータ λ を、試験で計画された総標本サイズである N と解析時点で蓄積された標本サイズ n に依存させ

$$\gamma_n = \frac{n}{2N}$$

とする方法を推奨している。この方法に基づくと、試験の序盤では蓄積された標本サイズは小さ

いため、 γ_n は 0 に近づき、割付確率は 0.5 に近づく方向で調整される。一方、試験中により多くのデータが収集されるにつれて γ は 1 に近づくため、 λ が割付確率に反映される程度が大きくなる。つまり、試験序盤の標本サイズが小さな不確実性の高い状況では、割付比が群間で均等になる方向で調整され、標本サイズが蓄積してくるにしたがって、試験結果を反映した割付比になる用に調整されることを意味する。適応的ランダム化によって割付比が試験の進行と共にどのように変化するかを例に示す。例えば、最大標本サイズを 100 とし、20 名ごとに中間解析をするとする。最初の 20 名までは等確率で 2 群に割付を行い、その結果新規治療群では 10 名中 6 名、標準治療群では 10 名中 3 名に治療反応が示されたとする。各群に無情報ベータ事前分布 $Beta(1, 1)$ を設定すると、各群の治療反応率 (p_E, p_S) の事後分布は $Beta_{p_E}(1+6, 1+4)$, $Beta_{p_S}(1+3, 1+7)$ となり、 λ を求めると 0.44 となる。この時点での標本サイズ n が 20、最大標本サイズ N が 100 であるため、

$$\gamma_n = \frac{20}{2 * 100} = 0.1$$

となり、割付確率は

$$\frac{0.44^{0.1}}{0.44^{0.1} + (1-0.44)^{0.1}} = 0.494$$

となる。そこで、次の 20 名については割付確率 0.504 で新規治療に割付を行う。次の 20 名に関しては、先の 20 名から算出した事後分布が各々事前分布として設定される。割付確率は次の 20 名でもほぼ 50% と等確率であるため、各群には 10 名ずつ割付が行われたものとする。そして先ほどと同様に、新規治療では 10 名中 6 名、標準治療では 10 名中 3 名に治療反応が得られたとする。この状況で、各群の事後分布は $Beta_{p_E}(1+6+6, 1+4+4)$, $Beta_{p_S}(1+3+3, 1+7+7)$ となり、 λ を求めると 0.85 となる。この時点での標本サイズ n が 40、最大標本サイズ N が 100 であるため、

$$\gamma_n = \frac{20}{2 * 100} = 0.2$$

となり、割付確率は

$$\frac{0.85^{0.2}}{0.85^{0.2} + (1-0.85)^{0.2}} = 0.586$$

となる。次の20名は割付確率0.586で新規治療に割付を行う。今回の割付確率は新規治療への割付がやや高くなるため、新規治療には12名、標準治療には8名割付けられたとする。今回は、新規治療では12名中7名、標準治療では8名中3名に治療反応が得られたとする。この状況で、各群の事後分布は $Beta_{p_k}(1+6+6+7, 1+4+4+5)$, $Beta_{p_k}(1+3+3+3, 1+7+7+5)$ となり、 λ を求めると0.97となる。この時点での標本サイズ n が60, 最大標本サイズ N が100であるため、

$$\gamma_n = \frac{20}{2 * 100} = 0.3$$

となり、割付確率は

$$\frac{0.97^{0.3}}{0.97^{0.3} + (1-0.97)^{0.3}} = 0.627$$

となる。そこで、次の20名については割付確率0.627で新規治療に割付を行う。今回の割付確率も新規治療への割付がやや高くなるため、新規治療には13名、標準治療には7名割付けられたとする。今回は、新規治療では13名中8名、標準治療では7名中3名に治療反応が得られたとする。この状況で、各群の事後分布は $Beta_{p_k}(1+6+6+7+8, 1+4+4+5+5)$, $Beta_{p_k}(1+3+3+3+3, 1+7+7+5+4)$ となり、 λ を求めると0.99となる。この時点での標本サイズ n が80, 最大標本サイズ N が100であるため、

$$\gamma_n = \frac{20}{2 * 100} = 0.4$$

となり、割付確率は

$$\frac{0.99^{0.4}}{0.99^{0.4} + (1-0.99)^{0.4}} = 0.86$$

となる。そこで、次の20名については割付確率0.86で新規治療に割付を行う。今回の割付確率は新規治療への割付が極めて高くなり、新規治療には17名、標準治療には3名割付けられたとする。今回は、新規治療では17名中10名、標準治療では3名中1名に治療反応が得られたとする。この状

況で、各群の事後分布は $Beta_{p_k}(1+6+6+7+8+10, 1+4+4+5+5+7)$, $Beta_{p_k}(1+3+3+3+3+1, 1+7+7+5+4+2)$ となり、 λ を求めると0.99となる。この時点での標本サイズ n が100, 最大標本サイズ N が100であるため、

$$\gamma_n = \frac{20}{2 * 100} = 0.5$$

となり、割付確率は

$$\frac{0.99^{0.4}}{0.99^{0.4} + (1-0.99)^{0.4}} = 0.90$$

となる。このように、ベイズ統計学流の枠組みを利用することで、割付確率は試験結果に応じて適応的に変化していく。Yin (2012) は同様の枠組みで、事後予測分布から算出される予測確率を用いて適応的ランダム化を行う方法も提案している。

4.2 多群の適応的ランダム化

3群以上の多群比較の場合には、比較対象となる治療反応率を試験終了時まで固定する固定参照型適応的ランダム化と比較対象となる群を結果に応じて一つずつ除外していく移動参照型適応的ランダム化がある。固定参照型適応的ランダム化では、任意の固定された治療反応率または、試験に含まれるいずれかの群の治療反応率の事後分布を参照として、各群の治療反応率が参照以上になる

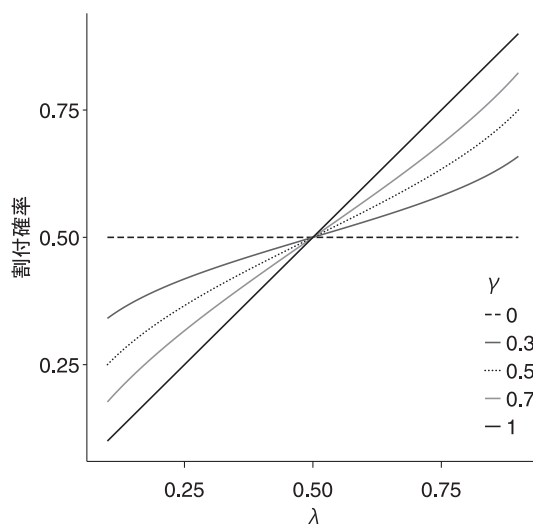


図3 適応的ランダム化のパラメータ特性

確率に基づいて、割付比が求められる。移動参照型適応的ランダム化では、多群の平均治療反応率の事後分布 \bar{p} を求め、それを参照とし各群の反応率の事後分布が \bar{p} 以上である確率を求め、その確率が最も低い群が試験から除かれる。このプロセスを繰り返し、試験の進行とともに無効な群が除外され、最終的に有効な治療へと患者が多く割り当てられるようになっていく。多群の適応的ランダム化の詳細については Yin (2012) を参照されたい。

5. 事前情報の活用

ここまでは、試験結果を逐次的にモニタリングし、推定結果に応じて柔軟に試験結果を判断する方法について述べてきた。ここからは、ベイズ統計学流の事前分布を用いた推論の枠組みを活用し、過去の臨床試験データを現行の臨床試験結果の推定に活用する方法について述べる。新規治療の検討を行う際に、対照群が設定されるが、その対照群が過去の臨床試験で非常に類似した集団で複数実施されている場合がある。そのような場合に、過去の臨床試験の統制群の情報を現在の試験の対照群の事前情報として事前分布に組み込むことで、推定の精度が高くなることが期待される。武田 (2015) はベイズ統計学流の枠組みを用いる利点の一つに、複数の過去の信頼性やバラツキを考慮しながら、それらの情報を事前分布として柔軟に取り込むことができる点を指摘している。ベイズ統計学流の枠組みでは、過去の臨床試験の信頼性が低ければ、その程度に応じて情報を割り引くことが可能になる。過去の試験データと新規試験データの間には、不均一性が存在する可能性があるため、過去の試験データと新規試験の類似性の程度に応じて連続的に過去の臨床試験データから利用する情報量を決定する必要がある (武田, 2015)。過去の臨床試験の対照群のデータを事前分布に用いる方法は、パワー事前分布に基づく方法と階層モデルに基づく方法に大別される (武田, 2015)。

5.1 パワー事前分布

パワー事前分布は、過去の試験データを得る前

の事前分布である初期事前分布と過去の試験データの尤度の積で求められる事後分布のことを指す。ただし、尤度は直接用いられるのではなく、ある程度情報量が割り引かれた形で用いられる。過去の試験データの尤度の情報を割り引くために、尤度はべき乗パラメータ (またはパワーパラメータ) a_0 でべき乗される。パワー事前分布は、新規試験データを解析する際に過去の試験の情報を有した事前分布として用いられる。べき乗パラメータ a_0 の設定方法は、3つに大別される (武田, 2015)。まず、 a_0 について分布を仮定せず既知の定数を与えるものと分布を想定するもの2つに大別され、前者が条件付きパワー事前分布 (Conditional Power Prior: CPP) (Ibrahim & Chen, 2000) と呼ばれる。そして、 a_0 について分布を仮定するものは、ジョイントパワー事前分布 (joint power prior: JPP) (Ibrahim & Chen, 2000) と修正パワー事前分布 (modified power prior: MPP) (Neuenschwander, Branson, & Spiegelhalter, 2009) に大別される。

条件付きパワー事前分布のべき乗パラメータ a_0 は、0～1の値をとり、 $a_0=0$ の場合には、尤度の情報が打ち消され、初期事前分布がそのまま新規試験の事前分布として利用される。 $a_0=1$ の場合には、尤度の情報が割り引かれることなく新規試験の解析に統合される。例えば、初期事前分布を $Beta(1, 1)$ 、既存試験のデータが $Binomial(20, 0.5)$ である場合を想定すると、 $a_0=0$ の場合の条件付きパワー事前分布は $Beta(1, 1)$ と初期事前分布に等しくなる。 $a_0=0$ の場合の条件付きパワー事前分布は $Beta(1+10, 1+10)$ となる。ジョイントパワー事前分布では、べき乗パラメータ a_0 に事前分布を想定する。 a_0 は0～1の値をとるため、通常ベータ分布が用いられる (Ibrahim & Chen, 2000)。ジョイントパワー事前分布は、 a_0 の事後分布が0に集中しやすく、過去の試験データの情報がほとんど新規試験に活用されていないことが指摘されている (Neuenschwander et al., 2009)。その問題を克服するために、修正パワー事前分布が提案され、基準化された条件付きパワー事前分布とべき乗パラメータの分布との積から構成される。修正パワー事前分布は、過去の試験と新規試験のデータの性質が異なる場合に、べき乗パラメータの事後平均が適切に0の方向に変化することが示されている (武田, 2015)。また、ジョイントパワー

事前分布や修正パワー事前分布は各々、過去の試験データが複数ある場合にも拡張されている (Ibrahim & Chen, 2000)。しかしながら、ジョイントパワー事前分布と修正パワー事前分布はいずれも直接的には、過去の試験データと新規試験データの類似性をパラメータ化していないことが問題となる。類似性をパラメータ化したものには commensurate 事前分布があり、過去の試験データから得られる情報量を類似性の程度に応じて調整可能となる。べき乗パラメータ a_0 の決定方法として、専門家の意見に基づき設定する方法 (Neelon & O'malley, 2009) や、複数の異なる a_0 の下で算出された条件付きパワー事前分布の中で事後分布に関する偏差情報量基準 (Deviance Information Criterion: DIC) が最も小さいものを選ぶ方法 (Ibrahim, Chen, & Chu, 2012)、事前有効サンプルサイズ (prior effective sample size) を用いて、条件付きパワー事前分布の情報量を人数に換算して評価する方法 (Morita, Thall, & Müller, 2012) などが提案されている。

5.2 階層モデル

新規試験と比較される対照群について、複数の過去の試験情報が存在する場合、階層モデルを用いて対照群の事前分布を設定することが可能である。階層モデルでは、新規試験の対照群と過去の試験のパラメータは、同じ分布からのランダムサンプルであるとみなす。この仮定によって、新規試験のパラメータと過去の試験のパラメータが完全に同一でも独立でもないという柔軟性が表現される (武田, 2015)。

Schmidli et al. (2014a) は階層モデルを用いて、複数の臨床試験の対照群のデータをメタ分析的に統合して新規試験の対照群の事前分布に用いる方法を提唱している (Schmidli et al., 2014a)。この方法で推定される事前分布は、メタ分析的予測事前分布 (Meta-Analytic Predictive prior; MAP 事前分布) と呼ばれる。まず、利用可能な過去の臨床試験の対照群データが複数 H 存在する場合を考える。既存試験の対照治療群の要約統計量およびパラメータを $Y_h (h=1, \dots, \theta_h)$ とし、新規試験の対照治療群の要約統計量およびパラメータを Y^*, θ^* とする。過去の臨床試験のデータと初期事前分布が共に正規分布に従う場合、データは、

$$Y_h | \theta_h, \sigma_h^2 \sim N(\theta_h, \sigma_h^2)$$

と書くことができ、初期事前分布は、

$$\theta, \dots, \theta_H, \theta^* \sim N(\mu, \tau^2)$$

と書くことができる。

初期事前分布はすべてのパラメータが共通の分布に従うことを表しており、 τ^2 が試験間分散、 σ_h^2 が試験内分散を示している。またパラメータ μ, τ^2 にも事前分布を仮定し、各パラメータについて不確実性を考慮することも可能である。

階層モデルにおいては、過去の臨床試験の利用の程度と関連するパラメータは、試験間のばらつきを表す τ である。 τ がゼロに近い場合、パラメータ θ が試験間でほぼ同じであることを意味し、過去の臨床試験を積極的に利用して関心のあるパラメータを推定する。 τ が大きければ試験結果が互いに大きく異なることを意味し、過去の臨床試験を利用せず、新規試験のみを使った解析となる。そして、新規試験のデザイン段階で用意する事前分布として予測分布 $p(\theta^* | Y_1, \dots, Y_H)$ を利用することができる (Gsteiger et al., 2013; Neuenschwander et al., 2010)。これは、複数の過去の試験データを所与とした場合の新規試験のパラメータ θ^* の分布である。 τ^2 と σ_h^2 を既知とし、 μ の超事前分布に一樣分布を仮定すると、予測分布は、

$$\theta | Y_1, \dots, Y_H, \tau \sim N\left(\frac{\sum w_h Y_h}{\sum w_h}, \frac{1}{\sum w_h} + \tau^2\right)$$

と書くことができる。ただし、 $w_h = (\sigma_h^2 + \tau^2)^{-1}$ である。

$$\frac{\sum w_h Y_h}{\sum w_h}$$

は既存試験の重み付け平均であり、予測分布の分散

$$\frac{1}{\sum w_h} + \tau^2$$

は、全体平均 μ の事後分散に新規試験のパラメータ θ^* の不確実性を追加した形で表現される。ここから、既存試験間の分散の影響が大きいことがわかる。

試験間のばらつきの指標 τ は既存試験の結果に基づき推定されるが、利用可能な既存試験は少ないことが多いため、正確な推定が難しく、超事前

表6 MAP 事前分布の推定に使用するデータ

| study | n | y | y.se |
|------------|-----|-----|-------|
| 1 Gastr06 | 74 | -51 | 10.23 |
| 2 AIMed07 | 166 | -49 | 6.83 |
| 3 NEJM07 | 328 | -36 | 4.86 |
| 4 Gastr01a | 20 | -47 | 19.68 |
| 5 APhTh04 | 25 | -90 | 17.60 |
| 6 Gastr01b | 58 | -54 | 11.55 |

分布の影響を受けやすいことが指摘されている (Gelman, 2006; Lambert et al., 2005)。一般的には、分散に対して共役分布である逆ガンマ分布、あるいは、分散の対数または標準偏差に対して一定の範囲内の値が等確率であることを意味する一様分布などが用いられることが多い (Gelman, 2006)。しかし、試験間のばらつきに関する事前情報が少ない場合や試験数が少ない場合には、複数の異なる超事前分布を用いた感度解析が必要である。多くは無情報事前分布になるように超事前パラメータを設定することが多い (Gelman, 2006)。

階層モデルでは、過去の臨床試験からどれくらいの情報量を利用するかは試験内分散 σ_h^2 、試験間分散 τ^2 および試験間のばらつきの超事前分布に依存する。多重積分が必要なためパワー事前分布のような明示的な表現は困難だが、分散比

(Neuenschwander et al., 2010) や 事前有効サンプルサイズ (Morita et al., 2012) を用いて、事前分布の情報量を評価する方法が提案されている。

ここでは、サンプルデータを用いて、MAP 事前分布を活用した分析例を紹介する。表6には、各研究の略名、対照群の人数 n 、治療前後のアウトカム (Crohn Disease Activity index: CDAI) の変化 y とその標準誤差 $y.se$ が示されている。 y はマイナス方向に大きいほど改善が大きいことを意味する。

試験間の分散 τ によって、過去の試験データをどの程度、新規試験に反映させるかが決定される。過去の情報から、標準偏差 σ は既知とし、過去の試験情報を借りる程度は、分散比 τ/σ に依存する。 τ の事前分布には切断正規分布である半正規分布 Half-normal(θ, σ^2) の使用が推奨されている (Spiegelhalter, Abrams, & Myles, 2004)。この設定の元で MCMC 推定法を用いた推定を行うと、MAP 事前分布として $N(-49.8, 19.3)$ が推定される。Schmidli et al. (2014b) は MCMC サンプルについて、混合分布を利用した変換を通じてより頑健な事前分布のパラメータを推定する方法を推奨している。Schmidli et al. (2014b) の方法を通じて得られた頑健化された MAP 事前分布は、今回のケースでは $N(-50, 88)$ と推定された (図4)。

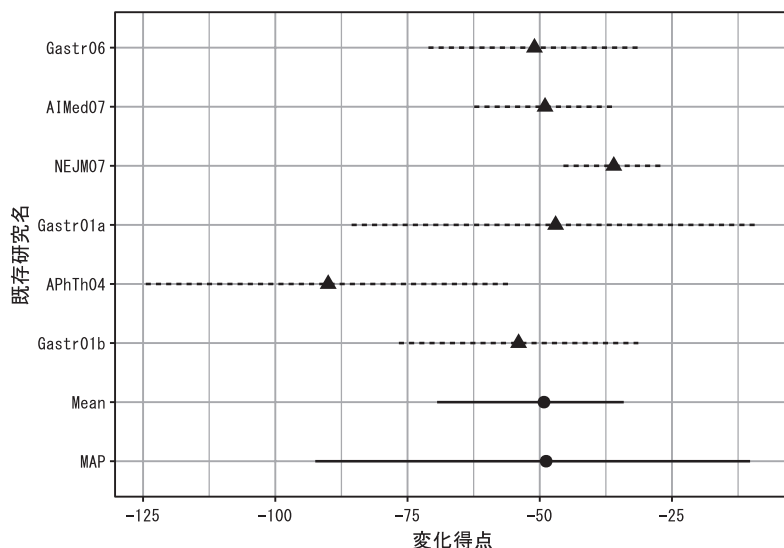


図4 MAP 事前分布の推定の結果 (MCMC 推定)：三角の点は個々の既存研究の事後平均、縦軸 Mean の点は既存研究の集団平均、縦軸 MAP の点は、事後予測事前分布の平均、実線は事後予測分布の95% 確信区間、破線と実線は95% 確信区間を示している。

頑健化されたMAP事前分布が得られたら、新規試験のデータの解析に併合する。今回のデータでは、新規試験において新規治療に39名が割付けられ、アウトカムは平均 $=-29.15$ 、標準誤差 $=16.69$ であった。対照群はMAP事前分布から計算される有効標本サイズに従い20名が割付けられ、平均 $=-76.01$ 、標準誤差は21.93であった。新規治療の事前分布には、分散を大きく設定した弱情報事前分布 $N(-50, 80)$ を指定して事後分布を求め、対照治療の事前分布にはMAP事前分布を指定して事後分布を求めた新規治療の事後分布は $N(-29.87, 16.40)$ 、MAP事前分布を使用した対照群の事後分布は $N(-74.49, 21.28)$ となる。本研究では、新規治療と対照治療の平均値の差が0以上である確率が.95以上である場合、新規治療が対照治療よりも有効であると判断するとする。新規治療と対照治療の事後分布から、各治療の平均値の差の事後分布が求められ、そこから新規治療と対照治療の平均値の差が0以上である確率は0.085であった。そのため、新規治療が対照治療よりも有効ではないと判断された。この結果は、過去の臨床試験の結果を事前分布として利用せずに、頻度主義統計学流に新規試験のデータのみで標準治療と新規治療の平均値の差の検定（両側検定）を行った場合でも同様に、 $p=0.097$ と5%水準で有意な差がないという結果であった。これは、過去の試験の試験内、試験間分散が大きいために、新規試験には情報があまり活用されなかったことが大きな要因である。

6. 最後に

本稿では、ベイズ統計学流の臨床試験デザインの一つかを概観した。ここで紹介した臨床試験デザインは、医薬品開発に限らず、心理社会的介入の有効性評価など、広く心理学の実験研究や調査研究に適用されうる。心理社会的介入の統制群には、治療待機群が用いられることが多く、冒頭で述べたような倫理的な問題を孕んでいる。欧米では、心理社会的介入の中でも、とりわけ認知行動療法は豊富な臨床試験データの蓄積がある。心理社会的介入の臨床試験においても、そのような過去の臨床試験の統制群のデータを活用することで、試験の効率化が期待される。一方、日本にお

いては、心理社会的介入の無作為化比較試験の蓄積が乏しいのが現状である。したがって、日本の現状においては、将来に過去の試験情報が有効活用されるように、質の高い無作為化比較試験を蓄積することが必要な段階であろう。日本で心理社会的介入の無作為化比較試験のコストを削減するために、ベイズ流の適応的ランダム化を活用した臨床試験の効率化は有効な方法の一つであろう。また、心理社会的介入の有効性の検討に限らず、既存研究の情報の活用方法は、調査研究も含め、心理学研究に広く適用可能である。例えば、今回はパワー事前分布について臨床試験の文脈で解説を行ったが、パワー事前分布の活用は、回帰モデル全般に適用されうる（Ibrahim & Chen, 2000）。また、項目反応理論の能力パラメータの推定にパワー事前分布を活用する方法も心理統計学の枠組みで提案されている（Matteucci & Veldkamp, 2015）。ただし、既存試験を新規試験の推定に活用する際には、過去の試験データと新規試験の間の比較可能性を担保することが、バイアスを減少させるために重要となる。比較可能性が担保される要件として Pocock（1976）は1）同じ治療を受けていること、2）同じ適格基準が用いられていること、3）アウトカムの評価方法が同一であること、4）性別年齢など疾患と関連する主要な背景要因の分布が同様であること、5）同様の組織で試験が実施されていること、6）新規試験との結果の違いを生むと予想される他の要因が存在しないことをあげている。同一の施設内で実施されている過去の臨床試験であれば、新規試験と比較可能性が高い状況がありえるが、過去の臨床試験の実施施設が異なれば現実的には6条件を満たすデータが得られることは稀である。修正パワー事前分布やMAP事前分布は過去の試験データから得られる情報の大きさを、試験結果の類似度に応じて制御する方法であるため、先の6条件を満たさない条件でもバイアスを低減するのに有用である（武田, 2015）。

ベイズ統計学流の枠組みを用いた臨床試験デザインによって、臨床試験にかかるコストを削減し、効率のよい試験運用が期待される。一方で、ベイズ統計学流臨床試験デザインのデメリットとしては、試験デザインをモニタリングするため、ベイズ統計学流のモデル構築やシミュレー

ションに精通するための技術的なハードルが高いことがあげられる。しかし、本稿で紹介した前後比較試験での、予測確率に基づくモニタリングやMAP事前分布を用いた先行試験の活用は、Rのph2bayesパッケージおよびRBestパッケージ(Sebastian, 2018)を用いて、比較的容易に実行可能な環境が整ってきている。

引用文献

- Barker, A., Sigman, C., Kelloff, G., Hylton, N., Berry, D., & Esserman, L. (2009). I-SPY 2: An adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clinical Pharmacology & Therapeutics*, 86, 97–100. <https://doi.org/10.1038/clpt.2009.68>
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1, 515–534. <https://doi.org/10.1214/06-BA117A>
- Gsteiger, S., Neuenschwander, B., Mercier, F., & Schmidli, H. (2013). Using historical control information for the design and analysis of clinical trials with overdispersed count data. *Statistics in Medicine*, 32, 3609–3622. <https://doi.org/10.1002/sim.5851>
- Ibrahim, J. G., & Chen, M.-H. (2000). Power prior distributions for regression models. *Statistical Science*, 15, 46–60. <https://doi.org/10.1214/ss/1009212673>
- Ibrahim, J. G., Chen, M.-H., & Chu, H. (2012). Bayesian methods in clinical trials: A Bayesian analysis of ECOG trials E1684 and E1690. *BMC Medical Research Methodology*, 12, 183. <https://doi.org/10.1186/1471-2288-12-183>
- Ito, M., Horikoshi, M., Kato, N., Oe, Y., Fujisato, H., Nakajima, S., ... Ono, Y. (2016). Transdiagnostic and transcultural: Pilot study of unified protocol for depressive and anxiety disorders in Japan. *Behavior Therapy*, 47, 416–430. <https://doi.org/10.1016/j.beth.2016.02.005>
- Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., & Jones, D. R. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine*, 24, 2401–2428. <https://doi.org/10.1002/sim.2112>
- Lee, J. J., & Liu, D. D. (2008). A predictive probability design for phase II cancer clinical trials. *Clinical Trials*, 5, 93–106. <https://doi.org/10.1177/1740774508089279>
- Matteucci, M., & Veldkamp, B. P. (2015). The approach of power priors for ability estimation in IRT models. *Quality & Quantity*, 49, 917–926. <https://doi.org/10.1007/s11135-014-0059-y>
- Morita, S., Thall, P. F., & Müller, P. (2012). Prior effective sample size in conditionally independent hierarchical models. *Bayesian Analysis (Online)*, 7. <https://doi.org/10.1214/12-BA720>
- Nagashima, K. (2016). Ph2bayes: Bayesian Single-Arm Phase II Designs. R package version 0.0.1. <https://CRAN.R-project.org/package=ph2bayes>.
- Neelon, B., & O'malley, J. (2009). Bayesian analysis using power priors with application to pediatric quality of care. *Journal of Biometrics & Biostatistics*, 1, 103. <https://doi.org/10.4172/2155-6180.1000103>
- Neuenschwander, B., Branson, M., & Spiegelhalter, D. J. (2009). A note on the power prior. *Statistics in Medicine*, 28, 3562–3566. <https://doi.org/10.1002/sim.3722>
- Neuenschwander, B., Capkun-Niggli, G., Branson, M., & Spiegelhalter, D. J. (2010). Summarizing historical information on controls in clinical trials. *Clinical Trials (London, England)*, 7, 5–18. <https://doi.org/10.1177/1740774509356002>
- Pocock, S. J. (1976). The combination of randomized and historical controls in clinical trials. *Journal of chronic diseases*, 29, 175–188.
- Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D., & Neuenschwander, B. (2014a). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70, 1023–1032. <https://doi.org/10.1111/biom.12242>
- Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D., & Neuenschwander, B. (2014b). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70, 1023–1032. <https://doi.org/10.1111/biom.12242>
- Sebastian, W. (2018). RBest: R Bayesian Evidence Synthesis Tools. R package version 1.3-2. <https://CRAN.R-project.org/package=RBest>.
- Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation*. John Wiley & Sons.
- 武田 健 (2015) 臨床試験におけるヒストリカルコントロールデータの利用 計量生物学, 36, 25–50.
- 手良向聡 (2017) なぜベイズを使わないのか!? 臨床試験デザインのために 京都: 金芳堂.
- Thall, P. F., & Simon, R. (1994). Practical Bayesian guidelines for phase IIB clinical trials. *Biometrics*, 50, 337–349. <https://doi.org/10.2307/2533377>
- Thall, P. F., & Wathen, J. K. (2007). Practical Bayesian adaptive randomisation in clinical trials. *European Journal of Cancer (Oxford, England: 1990)*, 43, 859–866. <https://doi.org/10.1016/j.ejca.2007.01.006>
- U.S. Department of Health and Human Services, & Food and Drug Administration. (2010). Food and drug administration. Guidance for industry and FDA staff: Guidance for the use of Bayesian statistics in medical device clinical trials.

Yin, G. (2012). *Clinical trial design: Bayesian and frequentist adaptive methods*. John Wiley & Sons.

Zhou, X., Liu, S., Kim, E. S., Herbst, R. S., & Lee, J. J. (2008). Bayesian adaptive design for targeted therapy development in lung cancer a step toward personalized

medicine. *Clinical Trials*, 5, 181–193. <https://doi.org/10.1177/1740774508091815>

— 2018. 2. 4 受稿, 2018. 3. 13 受理 —