

生存時間データに対するベイズ流例数設計

○矢田 真城¹ 魚住 龍史² 浜田 知久馬³

¹エイツーヘルスケア株式会社 開発戦略本部 生物統計部

²京都大学大学院医学研究科 医学統計生物情報学

³東京理科大学 工学部 情報工学科

Bayesian sample size calculation for survival analysis

Shinjo Yada¹, Ryuji Uozumi², and Chikuma Hamada³

¹*Biostatistics Department, Development Strategy Division, A2 Healthcare Corporation*

²*Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine*

³*Department of Information and Computer Technology, Tokyo University of Science*

要旨

臨床試験では、試験のデザインや目的に従い試験結果の精度を保証するために、必要と考えられる例数を設定する。必要例数を解析的に導出することが難しい場合、シミュレーションにより必要例数を見積もる方法が用いられる。シミュレーションによるベイズ流例数設計では、“もし…としたら”という状況の下で仮想的に生成されたデータを解析し、その結果がモデル判定基準を満たすような例数を、当該臨床試験に必要な例数と見積もる。このとき、検定の有意水準、検出力、効果の差といった概念は、事後分布に関連付けて提示することができる。本稿では、シミュレーションによるベイズ流例数設計として、*indifference zone* を用いた方法について紹介する。これは、効果の差を表すモデルパラメータの信用区間と *indifference zone* との位置関係から必要例数を見積もる方法である。シミュレーションの実行手順に関してまとめ、適用例として生存時間データにおける例数設計をとりあげ、対応する SAS プログラムと簡単なシミュレーション結果を示す。

キーワード： *indifference zone*, *design prior*, *analysis prior*

1. はじめに

臨床試験における例数設計は、試験デザインの重要な一つの要素である (上坂, 2006). 統計学的な例数設計には、精度に基づく例数設計と検出力に基づく例数設計とがあるが、臨床試験では検出力に基づく例数設計がよく用いられている. 検定に用いる帰無仮説及び対立仮説を設定し、検定の有意水準 α , 検出力 $1-\beta$, 予想される群間での効果の差 Δ とそのばらつきを与え、適切な検定方法を選択することで、必要例数を算出することができる. ただし、必要例数を解析的に算出するためには、対立仮説のもとでの検定統計量の分布を知る必要がある. より複雑なモデルを想定した場合、対立仮説のもとでの検定統計量が従う分布の導出が難解になり、必要例数を明示できないことがありえる. このような場合、シミュレーションに基づき必要例数を見積もる方法がある (大橋等, 2016).

シミュレーションに基づくベイズ流の例数設計では、必要例数 N を仮に定めた後、1) 予め設定した事前分布に基づきモデルパラメータを生成させる、2) 生成させたモデルパラメータを用いてデータを発生させる、3) 発生させたデータを用いてベイズ流に解析する、4) 得られた事後分布から予め設定したモデル性能基準 (model performance criteria) と照らし合わせる、という手順をとり、モデル性能基準を満たしていなければ、 N を変えて 1) から 4) を再度実行する (Wang and Gelfand, 2002). このとき、検定の有意水準、検出力、群間での効果の差といった概念は事後分布に関連付けて定義される. そこで本稿では、事後分布に基づくベイズ流例数設計の 1 つとして、indifference zone による方法を取りあげ、シミュレーションによる例数設計の手順を示し、検定の有意水準、検出力との類似性をまとめる.

本稿の構成は以下のとおりである. 第 2 節で indifference zone を用いたベイズ流例数設計について説明する. 第 3 節にて、indifference zone を用いたベイズ流例数設計として、生存時間解析における適用例を具体的に紹介する. 最後に第 4 節で本稿のまとめを行う.

2. ベイズ流例数設計

2.1. Indifference zone

本稿では、対照群に対する試験群の有効性を示すことを目的とした、並行 2 群比較試験における例数設計を対象とする. "indifference zone" を用いたベイズ流例数設計において、基本となるアイデアは、検定の際に設定する帰無仮説「 $H_0: \Delta=0$ 」をある範囲をもった区間 $\Delta \in [\delta_L, \delta_U]$ に置き換えることにある. ここに、 Δ は対照群 (control) に対する試験群 (treatment) の 2 群間での効果の差、 δ_U は試験群がこの値を超えると対照群に比べて臨床的に優れているといえる限界値であり、 δ_L はこの値を下回ると試験群が対照群よりも臨床的に劣ると考えられる閾値である. この indifference zone $[\delta_L, \delta_U]$ と 2 群間での効果の差 Δ の 95% 信用区間との位置関係から、試験群と対照群との優劣を結論付けることが可能となる. 効果の差 Δ が大きいほど対照群に対する試験群の優越性が示されている状態であるとき、 Δ の 95% 信用区間に基づく判定基準 (decision rule) は図 2.1 のようになる (Spiegelhalter et al., 1994). "accept treatment" (試験群を採択), "reject control" (対照群を棄却), "equivalence" (同等), "reject treatment" (試験群を棄却), "accept control" (対照群を採択), "no decision" (判定不能) のいずれかの判定が下される.

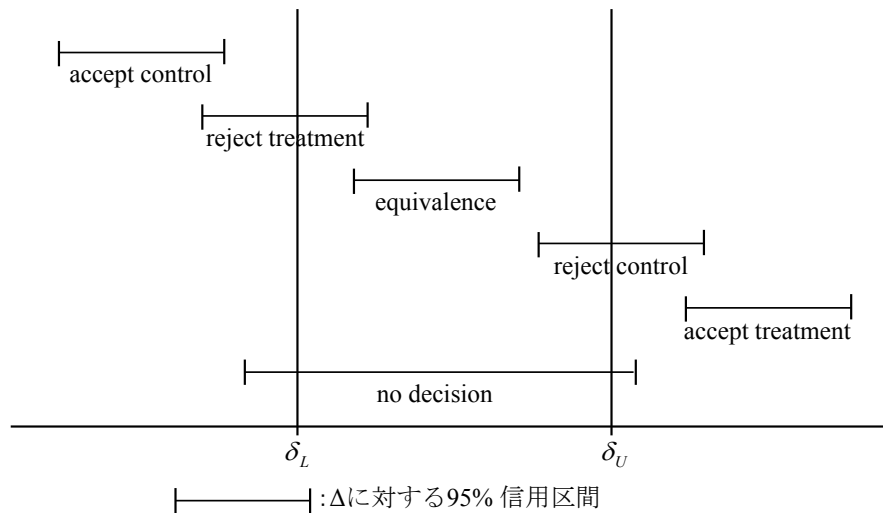


図 2.1 : 効果の差 Δ の 95%信用区間に基づく indifference zone と判定基準との関係

2.2. シミュレーションによる例数設計

例数設計の大きな流れとしては次のようになる。

- 手順 1 データに対してあてはめるモデルを想定し, 2 群間での効果の差 Δ , indifference zone $[\delta_L, \delta_U]$ 及びモデルパラメータの事前分布を設定する.
- 手順 2 事前分布に基づき N 人分の被験者の観測値を仮想的に生成する
- 手順 3 手順 2 で生成された仮想的な観測値を用いてベイズ流にパラメータを更新し, Δ の 95%信用区間を推定する.
- 手順 4 Δ の 95%信用区間と indifference zone $[\delta_L, \delta_U]$ との位置関係から, 図 2.1 に従い 6 つの判定基準のいずれに該当するかを判定する. これが 1 回のシミュレーションによって得られる判定結果となる.
- 手順 5 手順 2 から手順 4 までを数千から数万回繰り返すことで, 判定基準ごとの経験的確率を計算することができ, 例数 N に対する検出力などを推定することができる.

ベイズ流アプローチでは, 解析事前分布 (analysis prior あるいは fitting prior) とデザイン事前分布 (design prior あるいは sampling prior) という 2 種類の事前分布を設定し, 例数設計を行う. 解析事前分布は, 事前情報をモデル化し (ときにヒストリカルデータに代表される情報をモデル化し), 事後分布を得るために用いられるべきであり, デザイン事前分布はパラメータの不確実性をモデル化するために用いられるべきである (Brutti et al., 2008).

解析事前分布は, 文字どおり集積されたデータに対しモデルをあてはめる, 即ち解析するために用いることを想定し設定される事前分布である (Wang and Gelfand, 2002). たとえば対照群の治療効果をパラメータ θ_1 で, 試験薬の治療効果をパラメータ θ_2 で表すモデルを考える. 対照群については試験開始前にかんりの情報を得ているが, 試験薬については対照群ほどの豊富な情報はもちあわせていないかもしれない. このように, 各モデルパラメータに対する事前分布の検討においては, それまでの経験則やヒストリカルデータに対する解析結果のレビューから始めるべきであり, それが解析事前分布を設定するときに役立つ (Berry et al, 2011). 一方, デザイン事前分布は "もし…だったとしたら" という見解に基づき設定される事前分布である (Wang and Gelfand, 2002). 即ち, 試験をデザインするときに用いられる事前分布であり, 過去に行われた試験の結果や文献調査により収集された情報をもとに検討される (Berry et al, 2011).

上記手順でいえば、デザイン事前分布は手順2で、解析事前分布は手順3で、それぞれ使用する。手順2において、デザイン事前分布に基づきモデルパラメータのもとで*i*番目の被験者 ($i = 1, 2, \dots, N$) に対する観測値を仮想的に生成させる。Type I エラーに完全に対応する "Bayesian analogs" を求めるためには、デザイン事前分布に従いパラメータを固定させて観測値を生成させるが、よりベイズ流に行うのであれば、デザイン事前分布からサンプリングされたモデルパラメータを用いて観測値を生成させることも考えられる。続く手順3にてパラメータを更新し Δ の 95%信用区間を推定する際に、解析事前分布を使用する。

頻度論におけるシミュレーションによる例数設計では、帰無仮説及び想定する対立仮説を決め、対立仮説のもとで仮想的にデータを生成させる。生成させたデータに対し治療効果に差があるかの仮説検定を行い、有意水準 α で検定が有意となるか判定する。この過程をシミュレーション回数 N_{rep} 繰り返し、有意になる割合を調べればこれが検出力の推定値となる。これに対し、ベイズ流のシミュレーションによる例数設計では、デザイン事前分布に基づき仮想的にデータを生成させる。生成されたデータに対し解析事前分布を用いてベイズ流に解析し、モデル性能基準と照らし合わせる。この過程をシミュレーション回数 N_{rep} 繰り返し、検出力を推定する。indifference zone を用いたベイズ流例数設計では、"モデル性能基準と照らし合わせる" ところが、 Δ と indifference zone との位置関係から6つの判定基準のいずれに該当するかを判定するステップに該当する。

3. 生存時間解析における適用例の紹介

3.1. 試験デザインと解析方法

適用例として、生存時間解析における例数設計を紹介する (Berry et al, 2011)。対照群の5年生存率が0.65、試験群の5年生存率が0.80、対照群の例数：試験群の例数 = 1:1 の条件の下で、例数設計を行う場合を考える。生存時間分布に形状パラメータ γ 、尺度パラメータ η のワイブル分布

$$f(t) = \frac{\gamma}{\eta^\gamma} t^{\gamma-1} \exp\left(-\left(\frac{t}{\eta}\right)^\gamma\right) \quad (t > 0 \text{ のとき ; それ以外は全て } 0) \quad (3.1)$$

を仮定する。ここで

$$\lambda = -\gamma \log \eta$$

とおくとき、生存関数 $S(t)$ 及びハザード関数 $h(t)$ は λ を用いて $S(t) = \exp(-\exp(\lambda)t^\gamma)$ 、 $h(t) = \gamma \exp(\lambda)t^{\gamma-1}$ となる。ここで治療群を表す共変量を x ($x=0$ のとき対照群、 $x=1$ のとき試験群) として、共変量をモデルに組み込むために $\lambda = -(\beta_0 + \beta_1 x)$ とおくと、対照群に対する試験群のハザード比は

$$HR = \frac{\gamma \exp(-\beta_0 - \beta_1)x t^{\gamma-1}}{\gamma \exp(-\beta_0) t^{\gamma-1}} = e^{-\beta_1 x} \quad (3.2)$$

となり、 $\beta_1 > 0$ のとき $HR < 1$ となり試験群のほうが対照群よりも死亡のリスクが低いことを表す。共変量をモデルに組み込む際、 $\lambda = \beta_0 + \beta_1 x$ とせず $\lambda = -(\beta_0 + \beta_1 x)$ とおいたのは、そうすることで β_1 の値が大きいほど試験群のほうが対照群よりも死亡のリスクが低い、つまり試験群のほうが対照群に比べて臨床的に望ましい状態であることを表すため、2群の効果の差を表す指標として β_1 をとれば、 β_1 の 95%信用区間と6つの判定基準との関係に図2.1がそのまま適用できるためである ($\lambda = \beta_0 + \beta_1 x$ とおくと、 β_1 が大きいほど試験群のほうが対照群に比べて劣る状態を表すため、図2.1において treatment と control を反転させねばならなくなる)。対照群における時点 t での生存率を $S_1(t)$ 、試験群における時点 t での生存率を $S_2(t)$ と表すと、上記モデルのも

とでは

$$S_1(t) = \exp(-\exp(-\beta_0)t^\gamma)$$

$$S_2(t) = \exp(-\exp(-\beta_0 - \beta_1)t^\gamma)$$

により

$$\beta_0 = -\log\left(\frac{-\log S_1(t)}{t^\gamma}\right) \quad (3.3)$$

$$\beta_1 = \log\left(\frac{-\log S_1(t)}{t^\gamma}\right) - \log\left(\frac{-\log S_2(t)}{t^\gamma}\right) \quad (3.4)$$

である.

いま, i 番目の被験者における, イベント発現時間あるいは打ち切り時間を t_i , 群を表す変数を x_{1i} (対照群であれば $x_{1i} = 0$, 試験群であれば $x_{1i} = 1$), 打ち切りを表す変数を v_i (イベント発現のとき $v_i = 1$, 打ち切りのとき $v_i = 0$) と表すとき ($i = 1, 2, \dots, n$), 被験者の生存時間が互いに独立であるとして, n 人分のデータ D が与えられたもとの尤度は, $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, $\mathbf{x}_i = (1, x_{1i})^T$ として

$$L(\boldsymbol{\beta}, \gamma | D) = \gamma^d \exp\left[\sum_{i=1}^n \{v_i(\mathbf{x}_i^T \boldsymbol{\beta}) + v_i(\gamma - 1) \log t_i - t_i^{-\gamma} \exp(\mathbf{x}_i^T \boldsymbol{\beta})\}\right] \quad (3.5)$$

となる. ここに

$$d = \sum_{i=1}^n v_i$$

とおいた. モデルパラメータ $\boldsymbol{\beta}$ の事前分布として平均ベクトル $\boldsymbol{\mu}_0$, 分散共分散行列 $\boldsymbol{\Sigma}_0$ の 2 変量正規分布を, ワイブル分布の形状パラメータ γ の事前分布として平均 α_0 / κ_0 , 分散 α_0 / κ_0^2 のガンマ分布をそれぞれとりあげると, $\boldsymbol{\beta}$ と γ の同時事後分布は

$$p(\boldsymbol{\beta}, \gamma | D) \propto \gamma^{\alpha_0 + d - 1} \exp\left[\sum_{i=1}^n \{v_i(\mathbf{x}_i^T \boldsymbol{\beta}) + v_i(\gamma - 1) \log t_i - t_i^{-\gamma} \exp(\mathbf{x}_i^T \boldsymbol{\beta})\} - \kappa_0 \gamma - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0) / 2\right]$$

で与えられる (Ibrahim et al., 2001). $\boldsymbol{\beta}$ の事後分布は閉じた形式をもたず解析的に求められないため, MCMC 法 (Markov Chain Monte Carlo method) を用いて事後分布を導出する.

以降では簡単のため, ワイブル分布(3.1)の形状パラメータが固定値 $\gamma = 2$ をとる, 即ち生存時間分布がレイリー分布に従うものとする (Leemis and McQueston, 2008). あてはめるモデルを決めたら, そのモデルに基づいて, 対照群と試験群の効果の差を表すパラメータと, 対応する indifference zone $[\delta_L, \delta_U]$ を設定することになる. 対照群と試験群の効果の差を表すパラメータとして β_1 をとりあげると, $\beta_1 < 0$ のとき $HR > 1$ となり試験群のほうが対照群よりも死亡のリスクが高くなることから, indifference zone の下限値 δ_L を 0 とおく. 対照群の 5 年生存率が 0.65, 試験群の 5 年生存率が 0.80 のとき, (3.4)より

$$\beta_1 = \log(-\log S_1(5)/5^2) - \log(-\log S_2(5)/5^2) = \log(-\log(0.65)/5^2) - \log(-\log(0.80)/5^2) = 0.658$$

であることから, indifference zone の上限値 δ_U を 0.658 とおく.

ベイズ流アプローチによる例数設計では, 前述のとおり, モデルパラメータの事前分布を設定しなければならない. Spiegelhalter et al. (1994) は, 事前分布の設定に関して, "reference prior", "clinical prior" (臨床的事前分布), "skeptical prior" (懐疑的事前分布), "enthusiastic prior" (楽観的事前分布) の 4 種類から構成される "community of priors" の利用を推奨している. reference prior は「無情報」を客観的に表現した事前分布であり, いわゆる無情報事前分布 (non-informative prior) である. skeptical prior は試験薬に対して懐疑的な見解な立場での事前分布であり, より保守的な事前分布といえる. これに対して, enthusiastic prior は試験薬が対照

薬に比べて優れているとの信念に基づいた事前分布であり、臨床試験担当者にとって楽観的な事前分布といえる。本稿では、モデルパラメータ β_1 の懐疑的事前分布として平均 0 の正規分布を用いる。事前分布として用いる正規分布の分散は、 $P(\beta_1 > \delta_U) = 0.05$ となるよう 0.40^2 と設定する。楽観的事前分布として平均 0.658、分散 0.40^2 の正規分布を、無情報事前分布として平均 0、分散 10^4 の正規分布をそれぞれ用いる。モデルパラメータ β_0 は対照群の治療効果を規定する局外パラメータである。そこで、 β_1 が 3 つの事前分布いずれの場合においても平均 $-\log(-\log S_1(5)/5^\gamma) = 4.061$ 、分散 0.20^2 の正規分布を用いる。以上の設定のもとで、3.2 の手順 2 から手順 5 を実行することで、Type I エラー、検出力に対する "Bayesian analogs" を求めることも可能となる。 β_1 のデザイン事前分布として懐疑的事前分布 $N(0, 0.40^2)$ を用いたとき、対照群を却下する経験的確率がベイズ流 Type I エラー (Bayesian Type I error) の推定値となる。また、 β_1 のデザイン事前分布として楽観的事前分布 $N(0.658, 0.40^2)$ を用いたとき、対照群を却下する経験的確率がベイズ流 Type II エラー (Bayesian Type II error) の推定値であり、試験群を却下する経験的確率がベイズ流検出力 (Bayesian power) の推定値となる。

3.2. SAS プログラムコード

プログラム 3.1 は、上記事前分布に基づき 1 群あたり 140 人の被験者の仮想的な観測値を生成させるための SAS プログラムの一例である。なお、プログラムの作成にあたっては、Minnesota 大学公衆衛生大学院の web サイト (<http://www.biostat.umn.edu/~brad/software/BRugs/>) にて公開されている BUGS コードを参考にした。少し補足すると、 β_1 に対してはデザイン事前分布として懐疑的事前分布、楽観的事前分布を記載し、楽観的事前分布が採用されるよう、懐疑的事前分布の設定に関するコードをコメントアウトしている。プログラム 3.1 を実行することにより生成されるデータセット SIM では、イベント発現または打ち切りの時間 (年) を表す変数が TIME、群を表す変数が X (対照群であれば 0、試験群であれば 1)、打ち切りを表す変数が CENSOR (打ち切りであれば 0、イベント発現であれば 1) である。レイリー分布に従う乱数 T を生成させ、 $T > 5$ であれば打ち切り例とみなして CENSOR に 0 を代入し、TIME に 5 と代入する。逆に $T \leq 5$ であればイベント発現例とみなして CENSOR に 1 を代入し、TIME に T を代入している。

プログラム 3.1 : 擬似乱数による生存時間データの生成

```
data SIM ;
  gamma = 2 ;
  EST1 = 0.65 ; EST2 = 0.80 ;
  call streaminit(1234) ;
  do i = 1 to 140 ;
    do treatment = 1 to 2 ;
      x = treatment-1 ;
      beta0_mean = -log(-log(EST1))/(5**gamma) ;
      beta0 = RAND('NORMAL',beta0_mean,0.20) ;
      /*optimistic design prior*/
      beta1_mean = log(-log(EST1))/(5**gamma)-log(-log(EST2))/(5**gamma) ;
      beta1 = RAND('NORMAL',beta1_mean,0.40) ;
      /*skeptical design prior*/
```

```

/* beta1 = RAND('NORMAL',0,0.40);*/
lambda = -(beta0+beta1*x);
t = RAND('WEIBULL', gamma,(1/exp(lambda))**(1/gamma));
output ;
end ;
end ;
run ;
data SIM ;set SIM ;
keep i x time censor ;
if round(t,1e-8)>5 then do ; censor = 0; time = 5 ; end ;
else if round(t,1e-8)<=5 then do ; censor = 1 ; time = t ; end ;
run ;

```

デザイン事前分布に基づき N 人分の仮想的な観測値が生成できれば、それらを用いてベイズ流に解析し、2 群の効果の差 Δ (今回は β_1 が該当する) の 95%信用区間を推定する. プログラム 3.2 は、プログラム 3.1 で生成された 1 組の仮想的な観測値を用いて β_1 の 95%信用区間を推定するための SAS プログラムの一例である. β_1 に対する事前分布として、無情報事前分布、懐疑的事前分布、楽観的事前分布を記載し、無情報事前分布が解析事前分布として採用されるよう、他の 2 つの事前分布の設定に関するコードをコメントアウトした. (3.5)より対数尤度関数が

$$\log L = \sum_{i=1}^n \left\{ v_i \log \gamma - (\beta_0 + \beta_1 x_i) v_i + v_i (\gamma - 1) \log t_i - t_i^\gamma \exp(-(\beta_0 + \beta_1 x_i)) \right\}$$

で表されることを用い、MCMC プロシジャにて対数尤度関数を直接記述している. 得られた 10000 組のモンテカルロ標本を用いて β_1 の 95%信用区間を推定した. 今回のように、生存時間分布に標準的な分布をあてはめるのであれば、対数尤度関数を直接記述する以外に、SAS 関数 LOGPDF 及び LOGSDF を用いて対数尤度を記述することも可能である. その他 MCMC プロシジャの詳細については、SAS Institute Inc. (2015) を参照されたい.

プログラム 3.2 : MCMC プロシジャによる生存時間データの解析

```

proc mcmc data = SIM outpost = OUT nbi = 1000 nmc = 10000 seed = 1234 ;
gamma = 2 ; EST1 = 0.65 ; EST2 = 0.80 ;
beta0_mean = -log(-log(EST1)/(5**gamma)) ;
parms beta0 beta1 0 ;
prior beta0~normal(beta0_mean,sd = 0.2) ;
/*optimistic fitting prior*/
/* beta1_mean = log(-log(EST1)/(5**gamma))-log(-log(EST2)/(5**gamma)) ; */
/* prior beta1~normal(beta1_mean,sd = 0.40) ; */
/*skeptical fitting prior*/
/* prior beta1~normal(0,sd = 0.40) ; */

```

```

/*reference fitting prior*/
  prior beta1~normal(0,sd = 1e+2);
  lambda = -(beta0+beta1*x);
  llike = censor*(log(gamma) + (gamma-1)*log(time) + lambda)
          - exp(lambda)*(time**gamma);
  model general(llike);
run;
proc univariate data = OUT noprint;
  var beta1; output out = OUT pctlpre = P_ pctlpts=2.5 97.5;
run;

```

2 群の効果の差 β_1 の 95%信用区間が推定されれば, indifference zone $[\delta_L, \delta_U]$ との位置関係から, 図 2.1 に従い 6 つの判定基準のいずれに該当するかを判定することになる. プログラム 3.3 はその一例である. 変数 D1, D2, ..., D6 は, 対照群を採択, 試験群を棄却, 同等, 対照群を棄却, 試験群を採択, 判定不能に対する判定フラグであり 1 のとき該当, それ以外全て 0 をとる. 以上のステップをシミュレーション回数 N_{rep} 繰り返して得られる変数 D1, D2, ..., D6 の平均値が, 各判定基準の経験的確率となる.

プログラム 3.3 : 信用区間に基づく判定基準フラグの導出

```

data OUT; set OUT;
  gamma = 2; EST1 = 0.65; EST2 = 0.80;
  deltaL = 0;
  deltaU = log(-log(EST1)/(5**gamma))-log(-log(EST2)/(5**gamma));
  array a1[6] d1-d6;
  do i = 1 to 6;
    a1[i] = 0;
    if P_97_5 < deltaL then a1[1] = 1;
    else if p_97_5 < deltaU and p_97_5 > deltaL and p_2_5 < deltaL then a1[2] = 1;
    else if p_2_5 > deltaL and p_97_5 < deltaU then a1[3] = 1;
    else if p_2_5 > deltaL and p_2_5 < deltaU and p_97_5 > deltaU then a1[4] = 1;
    else if p_2_5 > deltaU then a1[5] = 1;
    else if p_2_5 < deltaL and p_97_5 > deltaU then a1[6] = 1;
  end;
run;

```


3.3. シミュレーション結果

表 3.1 は、デザイン事前分布に楽観的事前分布を、解析事前分布に無情報事前分布を、それぞれ用いたときのシミュレーション結果をまとめたものである。シミュレーション回数は 1000 回とした。デザイン事前分布に楽観的事前分布を用いる、即ち、楽観的事前分布に基づいて仮想データを生成するため、対照群に対する試験群の真のハザード比は(3.2)より 0.518 であり、「対照群を棄却」あるいは「試験群を採択」と判定された経験的確率を合算した値がベイズ流検出力の推定値となる。表 3.1 より、1 群あたりの例数 $N=80$ のときには、「対照群を棄却」と判定される割合は 0.63 であり、「対照群を棄却」よりもさらに強い「試験群を採択」と判定される割合も 0.01 となる。「対照群を棄却」あるいは「試験群を採択」と判定された割合 $0.63 + 0.01 = 0.64$ が、このシミュレーション結果から推定されるベイズ流検出力となる。デザイン事前分布に楽観的事前分布を設定し、対照群に比べて試験群が有効な状況であるにも関わらず、「試験群を棄却」と判定された割合は 0.02 であり、よってベイズ流 Type II エラーは 0.02 と推定される。例数の増加に伴ってベイズ流検出力も上がり、 $N=80$ のとき 0.64、 $N=100$ のとき 0.77、 $N=120$ のとき 0.82、 $N=140$ のとき 0.87、 $N=160$ のとき 0.90 となった。また、 $N=80$ のときは「判定不能」と判定された割合は 0.34 だったが、例数の増加に伴いその割合は減少していき、 $N=160$ のときには 0.07 であった。 N が大きくなるにつれて Δ の信用区間の幅が狭くなり、その結果、判定不能以外の判定結果となる割合が高くなっている。

表 3.1：無情報事前分布を解析事前分布として用いたときの 6 つの判定基準に対する経験的確率

N	Accept control	Reject treatment	Equivalence	Reject control	Accept treatment	No decision
80	0.00	0.02	0.00	0.63	0.01	0.34
100	0.00	0.02	0.00	0.76	0.01	0.22
120	0.00	0.02	0.00	0.81	0.01	0.16
140	0.00	0.03	0.00	0.86	0.01	0.11
160	0.00	0.03	0.00	0.89	0.01	0.07

表 3.2 は解析事前分布に懐疑的事前分布を用いたときの、表 3.3 は解析事前分布に楽観的事前分布を用いたときの、シミュレーション結果の要約である。シミュレーション条件は、解析事前分布の設定以外、全て表 3.1 のときと同じである。ベイズ流検出力は、 $N=80$ のとき懐疑的事前分布では 0.49、楽観的事前分布では 0.89、 $N=120$ のとき懐疑的事前分布では 0.64、楽観的事前分布では 0.95、 $N=140$ のとき懐疑的事前分布では 0.80、楽観的事前分布では 0.96、 $N=160$ のとき懐疑的事前分布では 0.86、楽観的事前分布では 0.97 となった。 $N=80, 100, 120, 140, 160$ いずれの場合も、3 つの事前分布の中では、楽観的事前分布を用いたときのベイズ流検出力が高く、懐疑的事前分布を用いたときのベイズ流検出力が低かった。事前分布の設定の仕方が必要な例数が異なることが確認できる。

表 3.2：懐疑的事前分布を解析事前分布として用いたときの 6 つの判定基準に対する経験的確率

N	Accept control	Reject treatment	Equivalence	Reject control	Accept treatment	No decision
80	0.00	0.11	0.00	0.49	0.00	0.40
100	0.00	0.12	0.00	0.64	0.00	0.25
120	0.00	0.11	0.00	0.73	0.00	0.17
140	0.00	0.10	0.00	0.80	0.00	0.10
160	0.00	0.10	0.00	0.86	0.00	0.04

表 3.3 : 楽観的事前分布を解析事前分布として用いたときの 6 つの判定基準に対する経験的確率

N	Accept control	Reject treatment	Equivalence	Reject control	Accept treatment	No decision
80	0.00	0.01	0.00	0.89	0.00	0.11
100	0.00	0.01	0.00	0.92	0.00	0.07
120	0.00	0.02	0.00	0.95	0.00	0.03
140	0.00	0.01	0.00	0.96	0.00	0.03
160	0.00	0.02	0.00	0.97	0.00	0.01

4. まとめ

本稿では、シミュレーションによるベイズ流例数設計として *indifference zone* を用いた方法についてまとめ、生存時間解析における適用例を紹介した。適用例では、一度に全患者が登録され、登録された患者のフォローアップ期間は全て等しい状況を想定している。しかし実際には、患者の登録は逐次的であり、フォローアップ期間も全ての患者で等しいとは限らない。より現実の臨床試験で起こりうる状況に応じた例数設計が求められる（魚住等，2016）。

全ての試験に共通する困難な問題は、例数の計算に用いる効果と変動の大きさの設定であり、このための考え方には、a) 臨床的に意味のある最小の差を定めること、b) 他の試験データによるエフェクトサイズの推定、のように 2 通りある（上坂，2006）。*indifference zone* を用いた方法では、臨床的に意味のある最小の差を定めることが、*indifference zone* の設定に対応しているといえる。b) の考え方は、先行する試験の結果を利用することが可能な場合に有用である。その際、今回紹介した方法であれば、先行する試験のデータをベイズ流に解析し、得られたモデルパラメータの事後分布を用いてデザイン事前分布を設定し、当該試験における必要例数を検討することができる。効果と変動の大きさの設定にあたり、参考可能な試験は存在するが試験薬に関する情報はない、文献上での情報であるため、詳細なデータは入手できず、試験の結果しかわからない状況も起こりうる。Hobbs and Carlin (2008) は、このような場合への適用例もあわせて紹介している。ヒストリカルデータを、そのデータがもつ情報量に応じて新規試験におけるデザインと解析内容に柔軟に組み入れることが可能なことが、ベイズ流例数設計の大きな魅力ではないだろうか。

参考文献

- [1] Berry, S.M., Carlin, B.P., Lee, J.J. and Müller, P. (2011). *Bayesian Adaptive Methods for Clinical Trials*. New York: Chapman & Hall/CRC.
- [2] Brutti, P., De Santis, F. and Gubbiotti, S. (2008). Robust Bayesian sample size determination in clinical trials. *Statistics in Medicine*. **27**(13):2290-2306.
- [3] Hobbs, B.P. and Carlin, B.P. (2008). Practical Bayesian design and analysis for drug and device clinical trials. *Journal Biopharmaceutical Statistics*. **18**:54-80.
- [4] Ibrahim, J.G., Chen, M.H., Sinha, D. (2001). *Bayesian Survival Analysis*. Springer-Verlag, New York.
- [5] Leemis, L.M. and McQueston, J.T. (2008). Univariate Distribution Relationships. *The American Statistician*. **62**: 45-53.
- [6] SAS Institute Inc. (2015). *SAS/STAT(R) 14.1 User's Guide*. Cary, NC, USA: SAS Institute Inc.

- [7] Spiegelhalter, D. J., Freedman, L. S., Parmar, M. K. B. (1994). Bayesian approaches to randomized trials (with discussion). *Journal of Royal Statistical Society: Series A*. **157**:357-416.
- [8] Wang, F. and Gelfand, A. E. (2002). A Simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science*. **17**(2):195-208.
- [9] 上坂浩之(2006). 医薬品開発のための臨床試験の計画と解析. 朝倉書店.
- [10]魚住龍史・矢田真城・浜田知久馬 (2016). SAS プロシジャを用いた生存時間データに対する例数設計の变革. SAS ユーザー総会論文集.
- [11]大橋靖雄・浜田知久馬 (1995). 生存時間解析 –SAS による生物統計. 東京大学出版会.
- [12]大橋靖雄・浜田知久馬・魚住龍史 (2016). 生存時間解析 応用編 –SAS による生物統計. 東京大学出版会.
- [13]中村剛 (2001). Cox 比例ハザードモデル. 朝倉書店.

連絡先

E-mail: yada-s@a2healthcare.com