

第5回 DSRT
医薬品の臨床試験および製造販売後調査
における
ベイズ統計学の活用について

**議題 5 : PIII非劣性試験における
ベイズ統計学の利用可能性
(予習資料)**

ディスカッションの目的と背景

議論の背景と目的

- 背景：
 - 医薬品開発において非劣性試験を行う場合，以下の状況が想定できると考える
 - 実対照群の「有効性に対する情報」が十分に蓄積されている
 - 実対照群の「有効性に対する情報」を活用することで，より効率的な医薬品開発が可能性がある
- 目的：
 - 架空の薬剤Zの計画およびデータ解析を検討することを通して，非劣性試験におけるベイズ統計学の利用可能性を検討する
 - 一方で，注意点も多いと思われる．特に「実際に計画・実施してみなければわからない」ことも多いと思われるため，架空の事例を通して検討する

ベイズ統計学の 利用可能性のポイント

- 期待される利点
 - 必要症例数の抑えることで曝露される被験者が減少
 - 開発期間の短縮によって新薬が早く市場に出せる
 - 開発費を抑えることで、開発に対するインセンティブを高める可能性
- 想定される懸念
 - 誤った結論を下しかねない
 - 過去の情報が不適切（今回のデータと大きく異なる、現在の治療環境を反映していないデータである、など）な場合など
 - 頻度論の結果とどのように対応を取って解釈できるか

非劣性試験におけるベイズ統計学

- 医薬品開発における非劣性試験のFDAガイダンスではベイズ統計学の利用について、次のように述べられている。 [1]

Bayesian methods that incorporate historical information from past active control studies through the use of prior distributions of model parameters **provide an alternative approach to evaluating non-inferiority in the NI trial itself.** Although discussed in the literature and used in other research settings, CDER and CBER have not had much experience to date in evaluating NI trials of new drugs or therapeutic biologics that make use of a Bayesian approach for design and analysis. **If a sponsor is planning to conduct a Bayesian NI trial, early discussions with the Agency are advised.**

→ ベイズ統計学は、これまでの頻度論の方法の代替案になり得るかもしれない。実務上は多くのディスカッションポイントがあるはず

→ 架空の薬剤の例を用いることで、実際の利用可能性や使用する際に注意すべき点などについて議論したい

ベイズ流仮説検定： 医療機器ガイダンスの例

- 参考までに，医療機器開発のFDAガイダンス[2]では，ベイズ統計学の利用が検定や区間推定と併記されている。

5.2 Hypothesis testing

Statistical inference may include hypothesis testing, interval estimation, or both. **For Bayesian hypothesis testing, you may use the posterior distribution to calculate the probability that a particular hypothesis is true, given the observed data.**

ディスカッションの題材の背景

対象疾患と既存治療薬

- 適用
 - 非弁膜症性心房細動患者における虚血性脳卒中および全身性塞栓症の発症抑制
- 既存治療薬
 - ワーファリン
- 主な指標
 - 血性脳卒中および全身性塞栓症の発現
- イベント発現率
 - 1~3[%/年]

* 参考：ワーファリンを対照薬とした非劣性マージンの導出例はFDAガイダンス[1]では例として挙げられている。本ディスカッションでは非劣性マージンの妥当性は議論しない

ワーファリン対照とした非劣性試験の例数と試験期間

- ワーファリンを対照薬とした薬剤Zの非劣性試験を検討する
- 2017年までに実際に開発された薬剤のPIII試験の例数及び登録期間および平均追跡期間(または中央値)を表に示す
- 薬剤Zの開発計画時点ではキシメラガトランの**SPORTIF III,V試験のデータのみ**が使えるものとする

試験名(新薬名)	例数(warf/新薬)*	登録	追跡
SPORTIF III(キシメラガトラン)[3]	1703/1704	1.1年	1.5年
SPORTIF V(キシメラガトラン)[4]	1962/1960	1.3年	1.7年
RE-LY(ダビガトラン)[5]	6022/6015/6076	2.0年	2.0年
ARISTOTLE(アピキサバン)[6]	9081/9120	3.3年	1.8年
ROCKET-AF(リバロキサバン)[7]	7133/7131	2.5年	1.5年
ENGAGE AF-TIMI 48(エドキサバン)[8]	7036/7034/7035	2.0年	2.8年

*新薬が複数用量のある場合はwarfarinの例数/低用量の例数/高用量の例数と表す。9

この疾患でベイズ統計学を用いる 動機

- この疾患のエンドポイントではイベントの発現率が非常に低く、6000 – 9000例/群の程度の例数が必要になる
- ワーファリン群の情報は事前の情報として複数の過去試験の情報が得られている
- より早期に新薬を開発するために、ベイズ統計学を用いて、効率的なサンプルサイズで科学的な結論を得たい

架空の薬剤Zの開発計画概要

架空の薬剤Zの概要

- 適用
 - 非弁膜症性心房細動患者における虚血性脳卒中および全身性塞栓症の発症抑制
- 既存治療薬
 - ワーファリン
- 主な指標
 - 血性脳卒中および全身性塞栓症の発現
- イベント発現率
 - 1~3[%/年](ワーファリン)
 - P2試験の結果より、薬剤Zの有効性はワーファリンと同等程度であると想定されており、安全性はワーファリンより優れることが期待されている

計画中の国際共同試験

- 目的
 - 薬剤Zのワーファリンに対する非劣性
- 対象
 - 非弁膜症性心房細動患者
- 治療群
 - 薬剤Zおよびワーファリン
- 主要評価項目
 - 血性脳卒中または全身性塞栓症の発現割合(複合エンドポイント)
- 解析方法
 - 1.5年後のイベント発現割合をベイズ統計学を用いて評価する
- 非劣性マージン【ディスカッション対象外】
 - FDAガイドラインで例示されているものから計算

本試験に関わる ワーファリンの試験は2種類

- ワーファリンvsプラセボ
 - 6試験(本資料のバックアップ資料参照)
 - メタアナリシスに基づき、非劣性マージンの算出に用いられている(FDAガイダンス)【ディスカッションの対象外】
- ワーファリンvs キシメラガトラン
 - 2試験(スライド9の表中の上2試験)
 - 直近のP3試験であり、デザインや患者集団が本試験の計画している集団を代表していると考えられる
 - ワーファリンの事前分布に用いる



プラセボ対象6試験



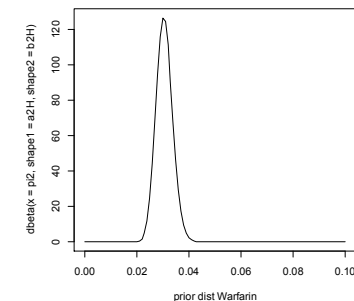
非劣性
マージンの
構成に利
用



SPORTIF III + V試験



事前分布
の構成に
利用



ワーファリンの事前情報の構成

- キシメラガトラン vs ワーファリン(1998 – 2003)[3,4]

試験名	例数(N)	event	平均 観察期間	event/N
SPORTIF III	1703	56	1.5	3.29%
SPORTIF V	1962	37	1.7	1.89%

- ワーファリン群の1.5年および1.7年のイベント発現割合は1.89%および3.29%と観察された。

ワーファリン群の交換可能性*

- ワーファリンはリスクベネフィットの観点から、マーカー (INR)による投与量のコントロールが必要な薬剤である
- ワーファリンのイベント発現割合はいかにワーファリンがコントロールされていたかに強く依存する
- 過去試験のうち、よくワーファリンがコントロールされたキシメラガトランのP3の2試験のワーファリン群の患者が最も本試験のワーファリン群の患者を代表していると考えられる
- また、プラセボ対象のワーファリンの試験の情報は、非劣性マージン導出に用いられるため、同じ試験の情報を非劣性マージンと事前分布に用いることを避けたい
- 従ってここでは、SPORTIF III と V のワーファリン群データのみを利用するものとした

*交換可能性については、参考文献[2]を参照し、ここではSPORTIF III と V は交換可能であると仮定する。

ワーファリンの過去の情報

- キシメラガトランvsワーファリン(1998 - 2003)[3,4]

試験名	例数(N)	event	平均 観察期間	event/N
SPORTIF III	1703	56	1.5	3.29%
SPORTIF V	1962	37	1.7	1.89%
合計	3665	93	1.6	2.54%

- ワーファリンの事前情報として, 上記2つの臨床試験を併合した症例のワーファリン群の発現割合(イベント数/合計例数)を用いる.
- 2試験で観察された3665例(そのうち93イベント発現)を本試験のワーファリン群の1.5年後のイベント発現割合の事前情報として使う

解析方法について

- 通常は生存時間解析を用いて、ハザードの比較が主要な解析となるが、ここでは計算を簡単にするため、18か月後(1.5年後)の発現割合を用いる
- FDAのガイドライン記載のマージンもハザード比に対応するものであるが、解析方法に割合を用いるため、18か月後の割合に変換したマージン $\Delta = 0.94\%$ をここでは用いる
 - 注意：マージンの設定の妥当性については、ここではディスカッションの対象としない

解析計画案の作成

解析計画(ベイズ統計/頻度論共通)

- ワーファリン群および薬剤Z群にそれぞれ n_w, n_z 人ずつに割り付け, 解析対象集団とする
- それぞれの18か月後のイベント発現例数を X_w, X_z とおく
- $i = w, z$ に対して

$$X_i \sim \text{Bin}(n_i, \pi_i)$$

とおく

解析計画(頻度論共通)

- 与えられた非劣性マージン $\Delta(> 0)$ に対して, 示したい結論(頻度論での対立仮説)を

$$\pi_w + \Delta > \pi_z$$

とする.

- 結論として薬剤Zの発現割合のワーファリンに対する非劣性を示したい
 - ワーファリンの発現割合 π_w に許容される非劣性マージン Δ を加算した割合よりも π_z は小さい

解析計画(ベイズ統計学：結論)

- 各群の発現割合 π_w, π_z を事前の情報および本試験のデータに基づく確率変数として扱う

– 注意：頻度論では定数として扱う点異なる

- 示したい結論 $\pi_w + \Delta > \pi_z$ が得られた情報に基づいて仮説が正しい確率

$$P(\pi_w + \Delta > \pi_z | X_w, X_z)$$

を表す

- この確率が事前に定めた閾値 p^* を超えた場合に非劣性と結論づけることとする

解析計画(ベイズ統計学：事前分布)

- ワーファリン群の過去の情報では, $n_{wH} (= 3665)$ 人の症例のうち $X_{wH} (= 98)$ 例で18か月後のイベント発現したとする
- この情報に基づいて事前分布を次のように構成する

$$\pi_w \sim \text{Beta}(\alpha_w, \beta_w)$$

$$\alpha_w = \lfloor a_{wH} X_{wH} \rfloor, \beta_w = \lfloor a_{wH} (n_{wH} - X_{wH}) \rfloor + 1$$

- ここで $\alpha_w (0 \leq \alpha_w \leq 1)$ は事前情報の尤度の重みを表す
 - prior effective sample size(は $ESS = \alpha_w + \beta_w$ である(詳細はAppendix)
 - $\alpha_w = 0$ のときはbeta分布とはならないことに注意
- 薬剤Zの π_z の事前分布 f_z を次のように構成する[8]

$$f_z(\pi_z) \propto (1 - \pi_z)^{-1}$$

解析計画(ベイズ統計学：事後分布)

- 本試験で得られた両群の情報 (n_w, X_w) および (n_z, X_z) を用いて π_w と π_z の事後分布を次のように構成する

$$\pi_w \sim \text{Beta}(\alpha_w + X_w, \beta_w + (n_w - X_w))$$

$$\pi_z \sim \text{Beta}(1 + X_z, (n_z - X_z))$$

- これらの事後分布を用いて

$$P(\pi_w + \Delta > \pi_z | X_w, X_z)$$

を評価し[8], 結論を得る

- 薬剤Zのイベントは最低でも1例起きる($X_z > 0$)と想定する

解析計画(ベイズ統計学：閾値)

- 事後確率 $P(\pi_w + \Delta > \pi_z | X_w, X_z)$ とFisherの正確性検定(片側での優越性)のp値(p と表す)には以下の関係がある [8]

$$\lim_{a_{wH}, \Delta \rightarrow +0} P(\pi_w + \Delta > \pi_z | X_w, X_z) = 1 - p$$

- この関係から結論の事後確率は, Fisherの正確性検定の拡張版と解釈することもできる
- 頻度論で用いられる有意水準片側2.5%に対応する閾値として p^* を97.5%と設定し

$$P(\pi_w + \Delta > \pi_z | X_w, X_z) > 0.975$$

のときに非劣性が示されたと結論する

解析計画(ベイズ統計学：閾値)

- 論文抜粋[8]

Theorem 1.

1 If $X_1 > 0, X_2 < n_2$, then between $P(\pi_1 > \pi_2 - \Delta | X_1, X_2)$ and the 1-sided P value of the Fisher exact test with $H_0 : \pi_1 \leq \pi_2$ versus $H_1 : \pi_1 > \pi_2$, the following relation holds

$$\lim_{a_{01}, a_{02}, \Delta \rightarrow +0} P(\pi_1 > \pi_2 - \Delta | X_1, X_2) = 1 - P.$$

2 Suppose $\alpha_1, \beta_1, \alpha_2, \beta_2, m_1, m_2 \in \mathbb{N}, 0 < k_1 \leq m_1$, and $0 \leq k_2 < m_2 - \beta_2$. Then, between $P(\pi_1 > \pi_2 - \Delta | X_1, X_2)$ given $X_1 = k_1 - \alpha_1, n_1 = m_1 - (\alpha_1 + \beta_1) + 1, X_2 = k_2 - \alpha_2 + 1, n_2 = m_2 - (\alpha_2 + \beta_2) + 1$ and the 1-sided P value of the Fisher exact test with $X_1 = k_1, X_2 = k_2, n_1 = m_1, n_2 = m_2$, the following relation holds

$$\lim_{\Delta \rightarrow +0} P(\pi_1 > \pi_2 - \Delta | X_1, X_2) = 1 - P.$$

再掲：解析方法について

- 通常は生存時間解析を用いて、ハザードの比較が主要な解析となるが、ここでは計算を簡単にするため、18か月後(1.5年後)の発現割合を用いる
- FDAのガイドライン記載のマージンもハザード比に対応するものであるが、解析方法に割合を用いるため、18か月後の割合に変換したマージン $\Delta = 0.94\%$ をここでは用いる(詳細な導出はAppendixと[1]を参照)

ワーファリン事前分布

$$\pi_w \sim \text{Beta}(\alpha_w, \beta_w)$$

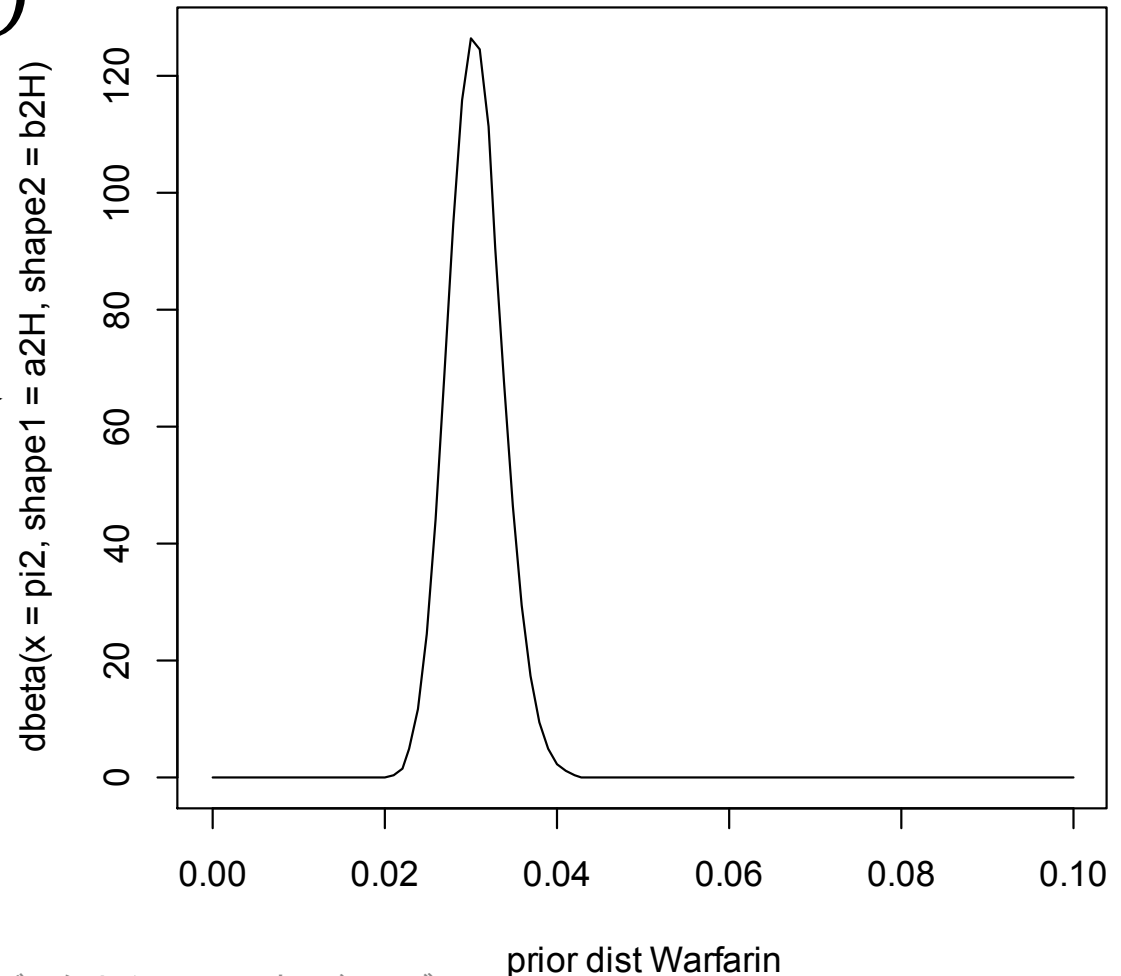
ここで

$$\alpha_w = [a_{wH} X_{wH}] + 1$$

$$\beta_w = [a_{wH} (n_{wH} - X_{wH})]$$

a_{wH} は事前情報の利用割合を表す

右図では本試験と過去試験交換可能であることを想定して $a_{wH} = 1$ を用いた



解析計画案の要約

- エンドポイント：血性脳卒中または全身性塞栓症
- 解析手法：投与開始から1.5年後の発現割合(π_z : 薬剤Z, π_w : ワーファリン)をベイズ統計を用いて次のように判定する

$P(\pi_x < \pi_w + \Delta | X_z, X_w) > 0.975$ のとき, 薬剤Zはワーファリンに劣らないと結論する

- 感度分析に以下の解析を行う
 - 事前分布の重みを0とした解析
 - 頻度論に基づく解析を行う

試験計画段階での検討 (事後分布の性質)

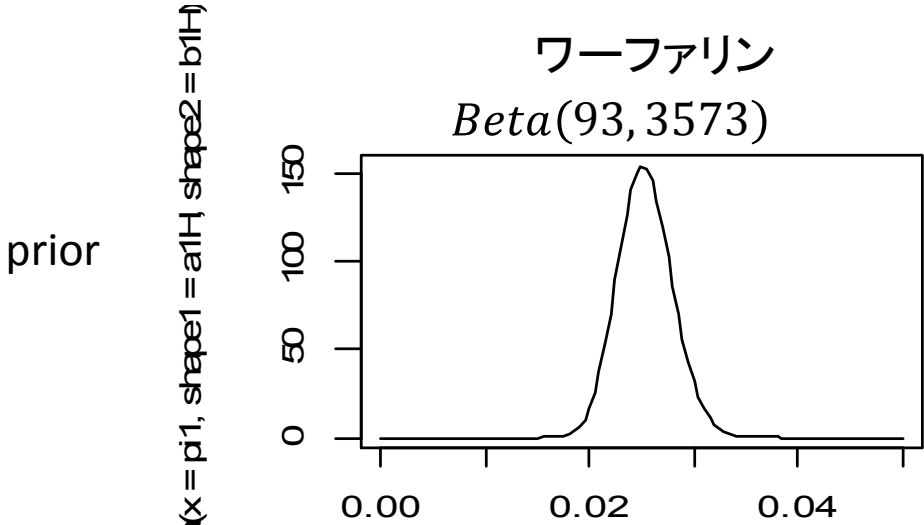
例)いったん結果を想定して 事後確率を計算してみる

- ここではまだ試験計画の段階であるが、例としていったん以下の試験を仮定して事後確率を計算してみる
 - 例数1000例/群で薬剤Zとワーファリンが事前分布に用いたデータと同じ発現率(1.5年後2.54%)が観測されると仮定する
 - つまり $n_z = n_w = 1000, X_z = X_w = 25$ および $a_{wH} = 1$ とおく
 - 事後確率はMonte Carlo(MC)推定 10^6 回により $P(\pi_x < \pi_w + \Delta | X_z = 25, X_w = 25) = 0.936955$ となる.

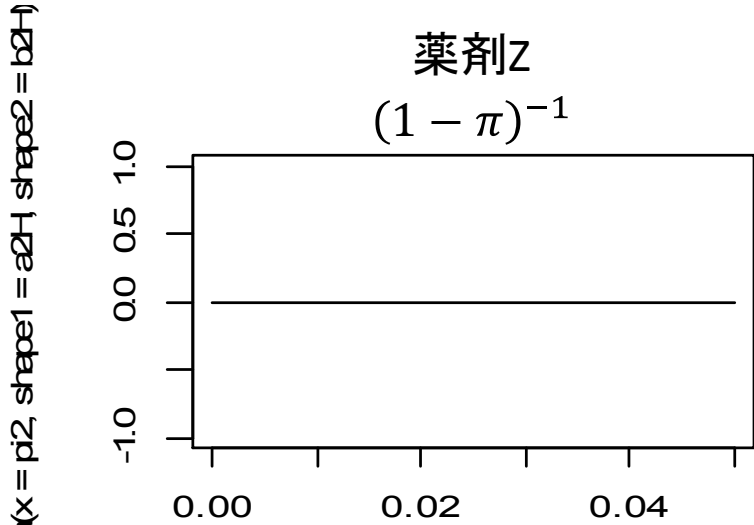
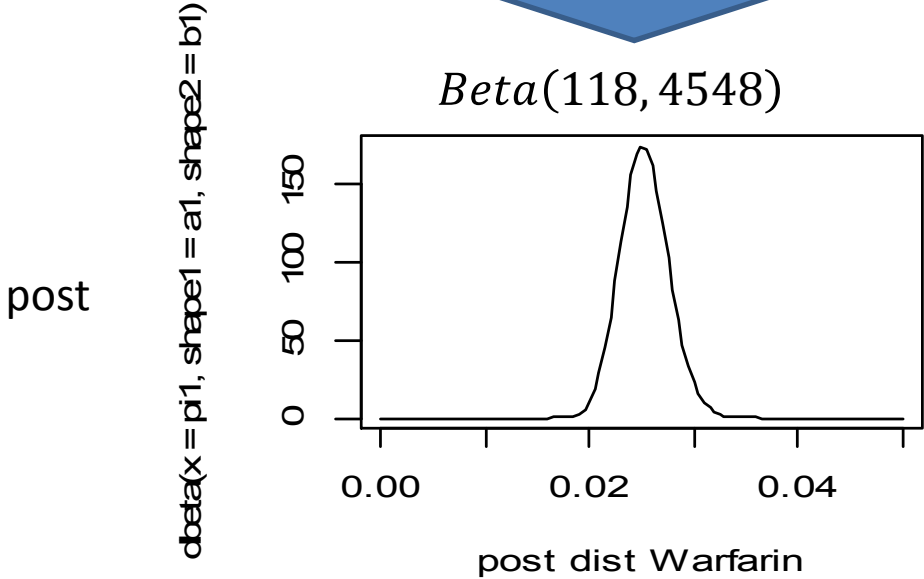
例)いったん結果を想定して 事後確率を計算してみる

- 従って、この例数でこのデータが得られたという仮定のもとでは非劣性を結論づけることができない。
 - 念のため、seedを変えて計算したところ、0.936393, 0.936393となった。
 - 参考文献[8]では正確な確率計算方法も提案されているが、本試験の例数の規模では計算時間が問題となりMC推定した。
- 頻度論の計画段階と同様に色々な仮定を置いて、事前に特性を評価することが、しばしば行われる

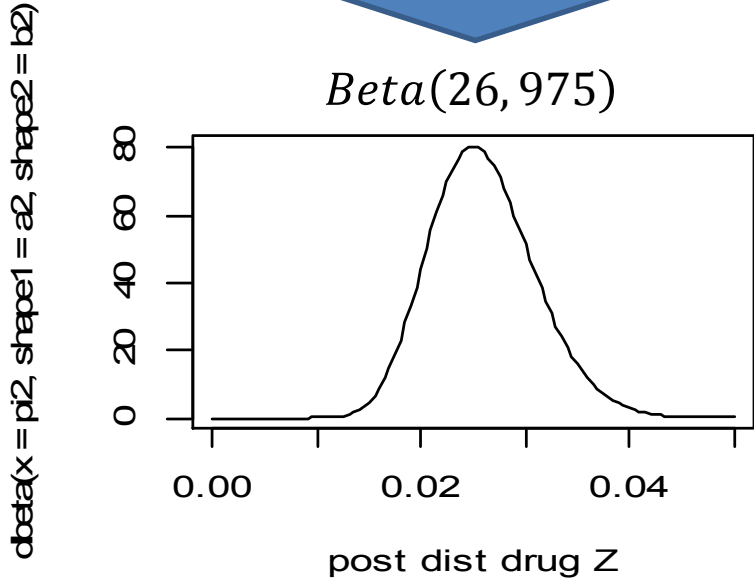
事前分布(上段)と事後分布(下段)



$X_w=25$
 $n_w=1,000$



$X_z=25$
 $n_z=1,000$



事前・事後分布の計算

- ワーファリンの事前分布は過去のデータより
 - 例数 $n_{wH} = 3665$ のうち $X_{wH} = 93$ に発現があった
 - いったんすべての情報を用いる($a_{wH} = 1$)
 - ここからパラメータを算出し, 次の分布とする

$$\alpha_w = [a_{wH}X_{wH}] = 93$$

$$\beta_w = [a_{wH}(n_{wH} - X_{wH})] + 1 = 3665 - 93 + 1 = 3573$$

- 次の分布を事前分布とする

$$\pi_w \sim \text{Beta}(93, 3573)$$

- ワーファリンの事後分布は,
 - パラメータにイベント数($X_{wH} = 25$)とイベントのなかった例数($n_w - X_w = 1000 - 25 = 975$)をそれぞれ足して

$$\pi_w \sim \text{Beta}(118, 4548)$$

事前・事後分布の計算

- 薬剤Zの事前分布は

- 以下を用いる[8]

$$f_z(\pi_z) \propto (1 - \pi_z)^{-1}$$

- 分布はimproperとなるがパラメータは次のように定義する

$$\alpha_{zH} = 1, \beta_w = 0$$

- 薬剤Zの事後分布は

- 事前分布にイベント数($X_z = 25$)とイベントのなかった例数($n_z - X_z = 1000 - 25 = 975$)をそれぞれ足して

$$\pi_w \sim \text{Beta}(26, 975)$$

- 次に事前情報の尤度の重みを変えたら($a_{wH} = 0, 0.5, 0.7$)どのように変化するかを試してみる

予習資料としてはここまでとなります。
分からない点・不明確な点はDSRT当日
にご質問の時間を設けております。

参考文献

1. FDAガイダンス, *"Non-Inferiority Clinical Trials to Establish Effectiveness Guidance for Industry"*
2. FDAガイダンス, *"Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials"*
3. Executive Steering Committee on behalf of the SPORTIF III Investigators, *"Stroke prevention with the oral direct thrombin inhibitor ximelagatran compared with warfarin in patients with non-valvular atrial fibrillation (SPORTIF III): randomised controlled trial."* *Lancet* 2003; 362: 1691-8.
4. Albers G. W., et al., *"Ximelagatran vs warfarin for stroke prevention in patients with nonvalvular atrial fibrillation: a randomized trial."* *JAMA* 2005; 293: 690-8.
5. Connolly, S. J., et al., *"Dabigatran versus Warfarin in Patients with Atrial Fibrillation"*, September 17, 2009, *N Engl J Med* 2009; 361:1139-1151

参考文献

5. Granger, C. B., et al., "*Apixaban versus Warfarin in Patients with Atrial Fibrillation*", September 15, 2011, *N Engl J Med* 2011; 365:981-992
6. Patel, M. R., et al., "*Rivaroxaban versus Warfarin in Nonvalvular Atrial Fibrillation*", September 8, 2011, *N Engl J Med* 2011; 365:883-891
7. Giugliano R. P, et al., "*Edoxaban versus warfarin in patients with atrial fibrillation.*" *N Engl J Med*. 2013; 369.22: 2093-104.
8. Doi M., et al., "*Bayesian noninferiority test for 2 binomial probabilities as the extension of Fisher exact test*", *Statistics in Medicine*, Volume 36, Issue 30, 30 December 2017, Pages 4789–4803