

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

# 本日のタイムライン

- ディスカッションの背景と目的[2分]
- ディスカッションの題材の背景[5分]
- 架空の薬剤Zの開発計画概要[10分]
- 解析計画案の作成[10分]
- 試験計画段階での検討(事後分布事後確率の計算)① [15分]
- 試験計画段階での検討(事後分布事後確率の計算)②[☆]
- ベイズ統計学による意思決定の動作特性[10分]
- 試験計画段階における例数設計[15分]
- 結果とその解釈[10分]
- 非劣性試験におけるベイズ統計学のメリット・デメリット[15分]

補足：各テーマごとに5分程度のディスカッションのお時間が含まれます。 ☆は割愛する可能性があります。

# 注意

- 時間が限られているため、本日のディスカッションで、すべてのスライドを説明することは出来ません。
- ディスカッションのための時間をより多く確保するために、スライドの一部を割愛することがありますが、ご了承ください。
- ★のついているページは割愛する可能性があるページです。

# ディスカッションの背景と目的

# ディスカッションの背景と目的

- 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ことがある
    - 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行える可能性
- 目的：
  - 架空の薬剤Zの開発において「実際に計画・実施してみる」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・使用上の注意点を検討する

# ベイズ統計学の利点および懸念

- 期待される利点の例
  - 必要症例数を抑えることで曝露される被験者が減少
    - 試験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などの点で不明確な点がある

# 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.**

- ベイズ統計学は代替案になり得るかもしれないが、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.**

例) 帰無仮説が $H_0: \pi_1 + \Delta \leq \pi_2$ の時，ベイズ統計学を用いて $P(\pi_1 + \Delta > \pi_2 | X_1, X_2)$ を評価する



# 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

- **ワーファリンのイベント発現率(%/年)**

- 1~3[%/年]

# ワーファリン対照とした 非劣性試験の例数と試験期間

- 2018年までに実際に開発された薬剤のワーファリンを対照とした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
- ワーファリンの事前情報には、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(ワーファリン/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

\*新薬が複数用量ある場合はワーファリンの例数/低用量の例数/高用量の例数を表す。<sup>11</sup>

# これまでのポイント

- イベントの発現を抑えるのが薬の目的
- 対照薬はワーファリンという薬
- イベントの発現が少ないので、サンプルサイズは数千例になる
- ワーファリンは2試験が事前情報に使える設定

# 架空の薬剤Zの開発計画概要

# 架空の薬剤Zの概要

- 薬剤Zへの期待
  - 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
  - イベント発現率は1~3[%/年](ワーファリン)程度と想定されている
  - 投与量コントロールが必要なワーファリンと異なり、より簡便に投与できる可能性がある
- 事前情報として利用可能な試験
  - 先のスライドで示した6試験は、実際には2018年現在までにすべて終了しているが、薬剤Zの開発時点では、SPORTIF III と V のみが終了し、事前分布として利用可能な結果が得られていると仮定する。

# 計画中の臨床試験

- **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確率\*(複合エンドポイント)
- **解析方法**
  - 1.5年後のイベント発現確率をベイズ統計学を用いて評価する
  - 本来はハザードを用いるが議論を簡単にするため、割合を用いる
- **非劣性マージン【ディスカッション対象外】**
  - FDAガイダンスの記載を参考に計算

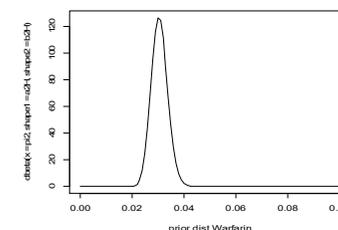
\*ハザード・発現率でなく発現確率を使っている点については、のちのスライドで説明

# 解析方法について

- 通常は生存時間解析の統計手法を用いた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、1.5年後の発現確率を用いる
- FDAの非劣性のガイダンス記載のマージンもハザード比に対応するものであるが、解析方法に発現確率を用いるため、1.5年後の発現確率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)
- 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

# 本試験に関わる ワーファリンの試験は2種類

- キシメラガトラン vs ワーファリン [再掲]
  - 2試験(スライド10の表中の上2試験)
  - 直近のPIII試験であり, デザインや患者集団が本試験で計画している集団を代表していると考える
  - ワーファリンのデータを用いて事前分布を構成する
- 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[年]	event/N
SPORTIF III [3]	1703	56	1.5	<b>3.29%</b>
SPORTIF V [4]	1962	37	1.7	<b>1.89%</b>

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

# ワーファリン群の事前情報

- 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確率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
- 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PIII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(交換可能\*)と考えられる
- 本試験でも、同様のプロトコルになるように計画する
- 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を事前情報として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SPORTIG III と V は交換可能であると仮定し、事前情報として利用できるとする



# 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) [再掲]

試験名	例数(N)	event	平均 観察期間[年]	event/N
SPORTIF III [3]	1703	56	1.5	3.29%
SPORTIF V [4]	1962	37	1.7	1.89%
合計 (事前情報)	<b>3665</b>	<b>93</b>	<b>1.6</b>	<b>2.54%</b>

- 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発現確率(2.54%; 3665例中93例発現)を用いる
- 背景表はAppendix参照および[3,4]

# これまでのポイント

- エンドポイントは1.5年後のイベント発現確率
- 非劣性マージンは固定(0.94%, FDAガイダンスに基づく)
- ワーファリンの事前情報は過去の2試験を統合したデータ  
(3665例中93例発現2.54%)に基づく
- 事前分布と非劣性マージンの導出は異なる試験が用いられる

# ディスカッションポイント

- 一度ここで整理したいと思います。
- 現時点で疾患・薬剤の概要，過去の試験について不明確な点や疑問点がありますでしょうか？
- 事前情報の前提条件として，より詳細に検討・確認すべき点がありますでしょうか？
  - 過去の2試験が同じとみなせるか
  - 過去の2試験が本試験と同じとみなせるか

# 解析計画案の作成

# 記載の定義一覧

記載	定義
$\pi_w, \pi_z$	各群の1.5年後のイベント発現確率(w:ワーファリン, z: 薬剤Z)
$n_{wH}, X_{wH}$	過去試験のワーファリン群における例数,および1.5年後のイベントの例数
$a_{wH}$	ワーファリン群の事前分布の重み [0,1]
$Beta(\alpha_w, \beta_w)$	$\pi_w$ の事前分布
$\alpha_w = \lfloor a_{wH} X_{wH} \rfloor$	事前分布の形状パラメータ $\lfloor x \rfloor$ は $x$ 以下の整数
$\beta_w = \lfloor a_{wH} (n_{wH} - X_{wH}) \rfloor + 1$	事前分布の形状パラメータ $\lfloor x \rfloor$ は $x$ 以下の整数
$ESS_w = \alpha_w + \beta_w$	prior effective sample size
$(1 - \pi_z)^{-1}$	$\pi_z$ の事前分布
$n_w, n_z$	本試験の各群の割り付け例数
$X_w, X_z$	本試験の各群の1.5年後のイベントの例数
$X_w/n_w, X_z/n_z$	本試験の各群の1.5年後のイベント発現確率
$Beta(\alpha_w + X_w, \beta_w + (n_w - X_w))$	$\pi_w$ の事後分布
$Beta(1 + X_z, (n_z - X_z))$	$\pi_z$ の事後分布
$P(\pi_w + \Delta > \pi_z   X_w, X_z)$	薬剤Zのワーファリンに対する非劣性の事後確率

# 解析計画(ベイズ統計学/頻度論共通)

- ワーファリン群および薬剤Z群にそれぞれ $n_w, n_z$ 人ずつに割り付ける
- それぞれの1.5年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  
とする  $X_i \sim \text{Bin}(n_i, \pi_i)$

# 解析計画(ベイズ統計学/頻度論共通)

- 薬剤Zの発現確率のワーファリンに対する非劣性を結論づけたい
  - ワーファリンの発現確率 $\pi_w$ に許容される非劣性マージン $\Delta$ を加算した割合よりも $\pi_z$ は小さい
- 頻度論では帰無仮説を次のようにおく
$$H_0: \pi_w + \Delta \leq \pi_z$$
- 示したい結論(頻度論での対立仮説)を
$$H_1: \pi_w + \Delta > \pi_z$$
とする

# 解析計画(ベイズ統計学：結論)

- 各群の発現確率 $\pi_w, \pi_z$ を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
- 得られた情報の下で、示したい仮説 ( $\pi_w + \Delta > \pi_z$ ) が正しい確率

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

を算出する

- この確率が事前に定めた閾値 $p^*$ を超えた場合に非劣性が示されたと結論する

補足：解析方法については[9]に基づくが、確率計算はMC推定を用いた。

# 解析計画(ベイズ統計学：事前分布)

- $\pi_w$ の事前分布(ワーファリン群)

- 事前情報： $n_{wH}$ (= 3665)人中 $X_{wH}$ (= 93)例でイベント発現したことより,  
 $\pi_w$ の事前分布を次のように構成する[9]

$$\pi_w \sim \text{Beta}(\alpha_w, \beta_w)$$
$$\alpha_w = [a_{wH}X_{wH}], \beta_w = [a_{wH}(n_{wH} - X_{wH})] + 1$$

- ここで $a_{wH}$ ( $0 \leq a_{wH} \leq 1$ )は**事前情報の(尤度の)重み**を表す
  - prior effective sample size(は $\text{ESS} = \alpha_w + \beta_w$ である(詳細は[9,10]))
  - $\alpha_w = 0$ のときはbeta分布とはならないことに注意

- $\pi_z$ の事前分布(薬剤Z)

- $\pi_z$ の事前分布 $f_z$ を次のように構成する[9]

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

# ワーファリン事前分布

$$\pi_w \sim \text{Beta}(\alpha_w, \beta_w)$$
$$\alpha_w = [a_{wH} X_{wH}]$$
$$\beta_w = [a_{wH} (n_{wH} - X_{wH})] + 1$$

本試験と過去試験交換可能であることを想定して

$$n_{wH} = 3665,$$

$$X_{wH} = 93,$$

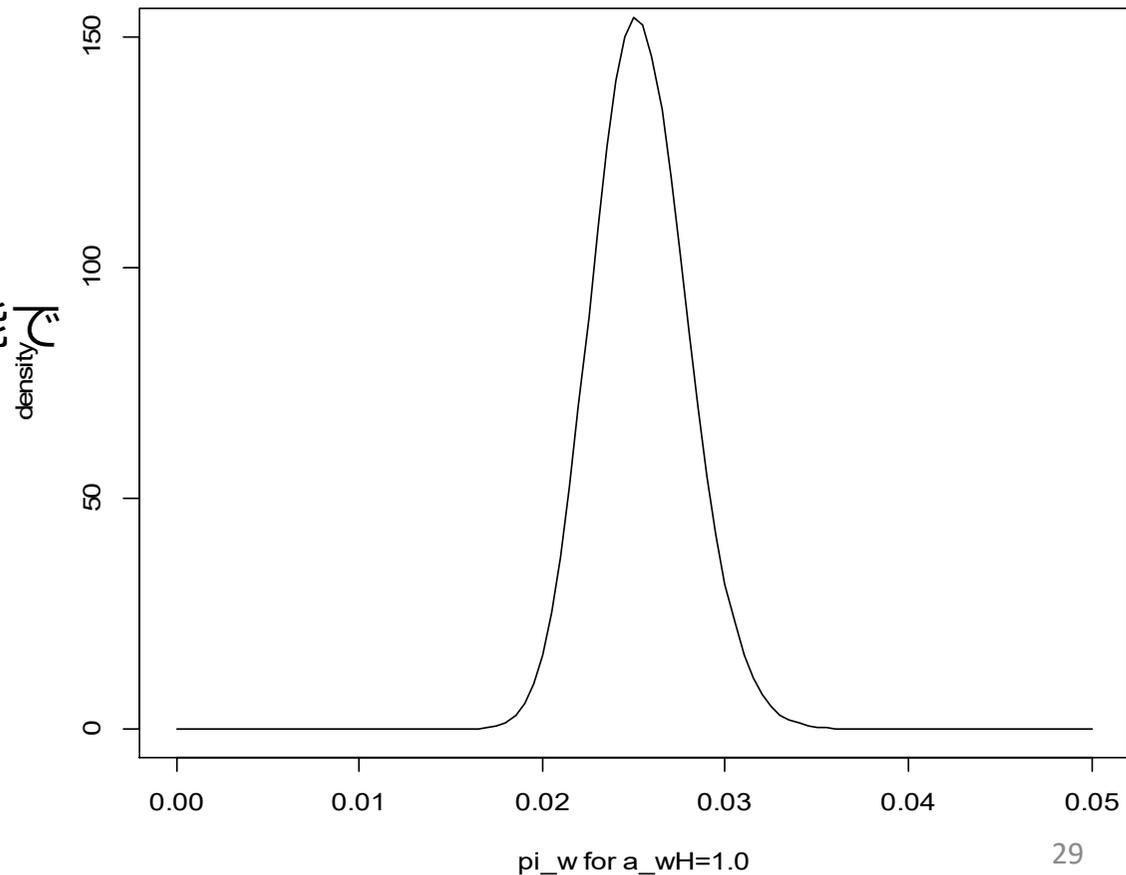
$$a_{wH} = 1,$$

より

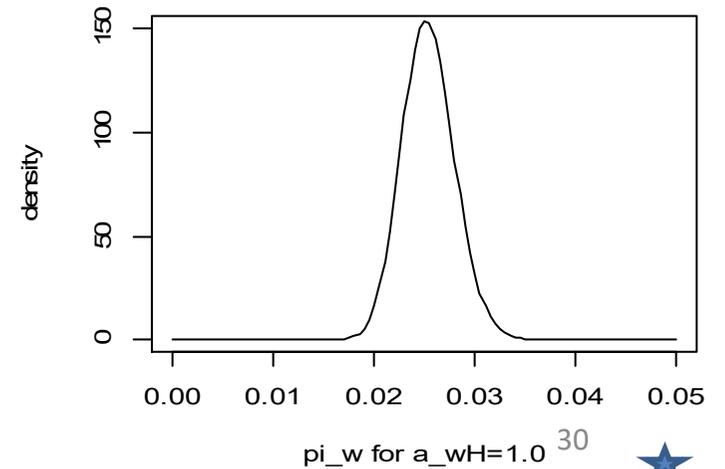
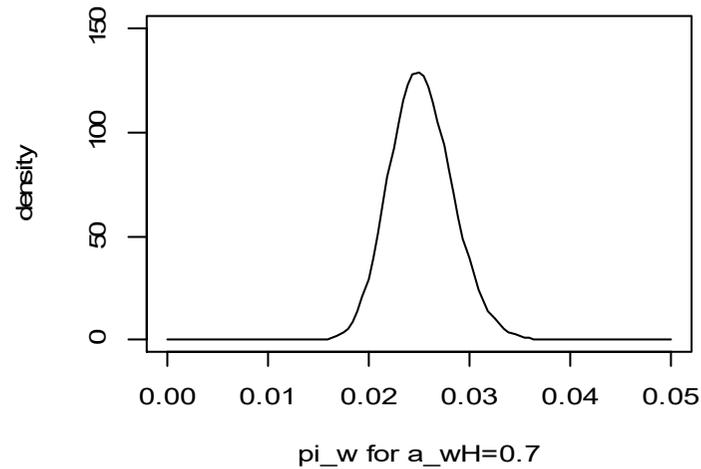
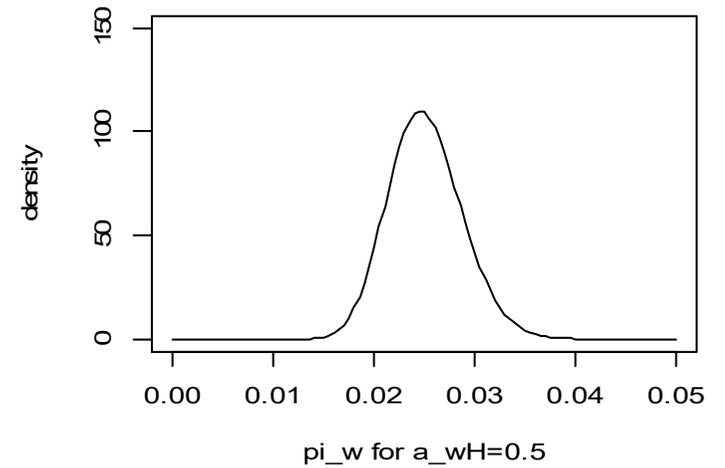
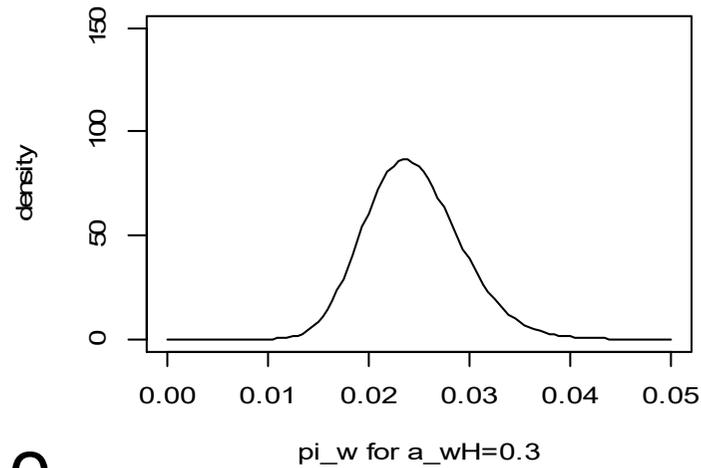
$$\text{Beta}(93, 3573)$$

を図示した。

\* $\pi_w \leq 0.05$ のみ表示している



参考：  
事前分布の  
重みを  
 $a_{wH}=0.3,$   
 $0.5, 0.7, 1.0$   
としたとき



# 解析計画(ベイズ統計学：事後分布)

- 本試験で観測される各群の情報  $(n_w, X_w)$  および  $(n_z, X_z)$  を用いて  $\pi_w$  と  $\pi_z$  の 事後分布\* を次のように 更新 する

$$\pi_w | X_w \sim \text{Beta}(\alpha_w + X_w, \beta_w + (n_w - X_w))$$
$$\pi_z | X_z \sim \text{Beta}(1 + X_z, (n_z - X_z))$$

- 事後分布を用いて 事後確率\*

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

を算出し[9], それに基づき結論する

\*今後, 事後分布と事後確率という言葉が頻出します.

# 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の直接確率検定(片側での優越性)のp値( $p$ と表す)には以下の関係がある [9]

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

- この関係から結論の事後確率は、**Fisherの直接確率検定の拡張版と解釈**することもできる
- この関係から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

$$P(\pi_w + \Delta > \pi_z | X_w, X_z) > p^* = 0.975$$

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[9]

## 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

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

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

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



# 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年後の発現確率( $\pi_z$ :薬剤Z,  $\pi_w$ :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

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

のとき**薬剤Zの発現確率はワーファリンに劣らない**と結論する

- 比較のため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(95%信頼区間利用)

# ディスカッションポイント

- 現時点で不明確な点や欠けている情報はありますか？
- ワーファリンの事前分布の設定は受け入れ可能でしょうか？どのようにすればより良い事前分布が作れるでしょうか？
- $a_{wH}$  の設定方法を決めるには、どうしたらよさそうか（どういう点を考えるとよさそうか）
- 比較のために、以下の解析を行うとします。「何を比較」していることになるでしょうか？
  - 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(95%信頼区間利用)

# 試験計画段階での検討(事後分布事後確率の計算)①

## 記載の定義一覧(再掲)

記載	定義
$\pi_w, \pi_z$	各群の1.5年後のイベント発現確率(w:ワーファリン, z: 薬剤Z)
$n_{wH}, X_{wH}$	過去試験のワーファリン群における例数,および1.5年後のイベントの例数
$\alpha_{wH}$	ワーファリン群の事前分布の重み [0,1]
$Beta(\alpha_w, \beta_w)$	$\pi_w$ の事前分布
$\alpha_w = \lfloor \alpha_{wH} X_{wH} \rfloor$	事前分布の形状パラメータ $\lfloor x \rfloor$ は $x$ 以下の整数
$\beta_w = \lfloor \alpha_{wH} (n_{wH} - X_{wH}) \rfloor + 1$	事前分布の形状パラメータ $\lfloor x \rfloor$ は $x$ 以下の整数
$ESS_w = \alpha_w + \beta_w$	prior effective sample size
$(1 - \pi_z)^{-1}$	$\pi_z$ の事前分布
$n_w, n_z$	本試験の各群の割り付け例数
$X_w, X_z$	本試験の各群の1.5年後のイベントの例数
$X_w/n_w, X_z/n_z$	本試験の各群の1.5年後のイベント発現確率
$Beta(\alpha_w + X_w, \beta_w + (n_w - X_w))$	$\pi_w$ の事後分布
$Beta(1 + X_z, (n_z - X_z))$	$\pi_z$ の事後分布
$P(\pi_w + \Delta > \pi_z   X_w, X_z)$	薬剤Zのワーファリンに対する非劣性の事後確率

# 例)想定結果の下での事後確率

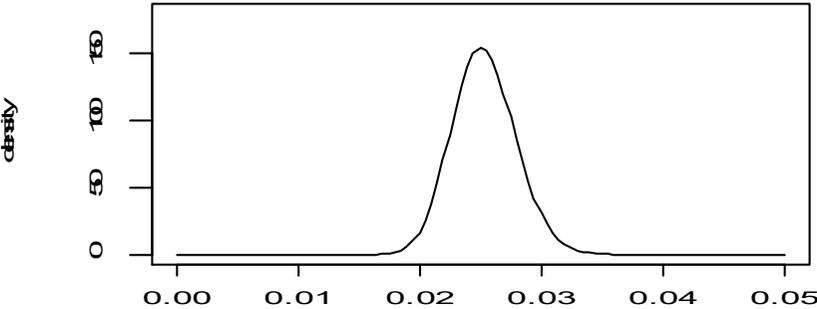
- ここではまだ試験計画の段階であるが，例を用いて理解を深める
- 以下の想定結果の下で事後確率を計算する
  - 例数1000例/群で薬剤Zとワーファリンが事前分布に用いたデータと同じ発現割合(1.5年後2.54%)が観測された，即ち
$$n_z = n_w = 1000, X_z = X_w = 25$$
  - $a_{wH} = 1$ とする
  - このとき，事後確率は0.937 ( $< 0.975$ ) となり，非劣性を結論づけることができない

# 事前分布(上段)と事後分布(下段)

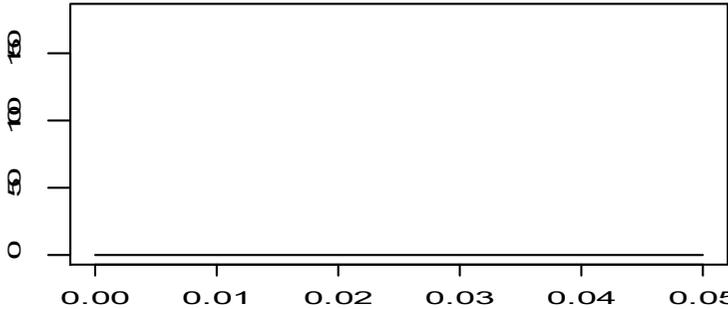
ワーファリン  $ESS_w = 3666$   
 $Beta(93, 3573)$

薬剤Z  
 $(1 - \pi_z)^{-1}$

prior



density



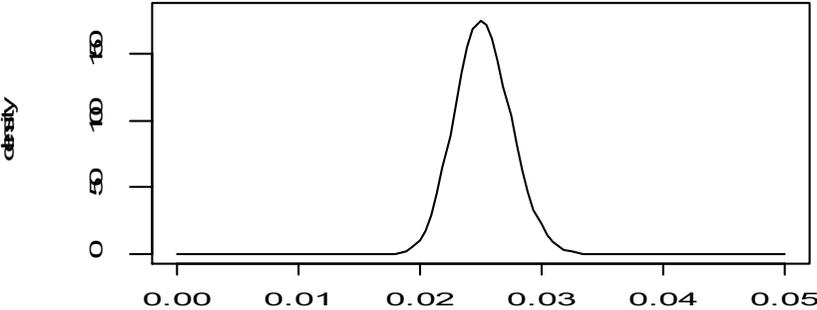
$X_w=25$   $n_w=1,000$

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

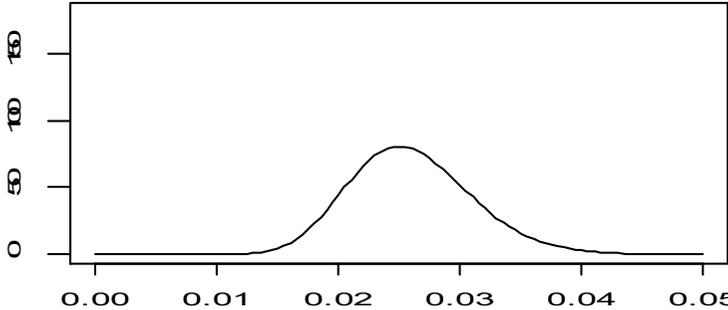
$Beta(118, 4548)$

$Beta(26, 975)$

post



density



post dist Warfarin

post dist drug Z

# ワーファリンの事前・事後分布

- ワーファリンの事前分布 :

- 事前情報 :  $n_{wH} = 3665$  例中  $X_{wH} = 93$  例でイベント発現
- $a_{wH} = 1$  のとき, 事前パラメータは

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

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

- よって, 事前分布は

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

- ワーファリンの事後分布 :

$$\begin{aligned} \pi_w | X_w &\sim \text{Beta}(93 + 25, 3573 + (1000 - 25)) \\ &= \text{Beta}(118, 4548) \end{aligned}$$

# 薬剤Zの事前・事後分布

- 薬剤Zの事前分布：
  - 以下を用いる[9]

$$f_Z(\pi_Z) \propto (1 - \pi_Z)^{-1}$$

- 薬剤Zの事後分布：

$$\alpha_Z = 1 + 25 = 26$$

$$\beta_Z = 0 + (1000 - 25) = 975$$

$$\pi_Z | n_Z, X_Z \sim \text{Beta}(\alpha_Z, \beta_Z) = \text{Beta}(26, 975)$$

## 例で使用した計算の設定

記号	パラメータの値
$\pi_w, \pi_z$	
$n_{wH}$	3665
$X_{wH}$	93
$a_{wH}$	<b>1</b>
$Beta(\alpha_w, \beta_w)$	$Beta(93, 3573)$
$\alpha_w = \lfloor a_{wH} X_{wH} \rfloor$	93
$\beta_w = \lfloor a_{wH} (n_{wH} - X_{wH}) \rfloor + 1$	3573
$ESS_w = \alpha_w + \beta_w$	3666
$(1 - \pi_z)^{-1}$	
$n_w, n_z$	<b>1000, 1000</b>
$X_w, X_z$	25, 25
$X_w/n_w, X_z/n_z$	<b>2.54%</b>
$Beta(\alpha_w + X_w, \beta_w + (n_w - X_w))$	$Beta(118, 4548)$
$Beta(1 + X_z, (n_z - X_z))$	$Beta(15, 975)$
$P(\pi_w + \Delta > \pi_z   X_w, X_z)$	0.936955

# 例) 想定と結果のまとめ

- ワーファリン群のみ3665例の情報を与えて、事前分布を構成した
- 本試験でワーファリン群および薬剤Z群の各群に1000例をリクルートを想定した
- 本試験の発現確率はワーファリンの事前情報と同じであると仮定して2.54%発現したと想定した
- 結果として、非劣性の事後確率は0.937であり、非劣性を結論できなかった

# ディスカッションポイント

- 現時点で不明確な点や欠けている情報はありますか？
- 計画段階で事後確率の特徴を把握するために動かしてみたいパラメータはありますか？

応用編(当日の時間次第では割愛します。)

## 試験計画段階での検討(事後分布事後確率の計算)②



# 今後の計算で**変化させるもの**と依存して**変化するもの**

記号	今後変化させるもの
$\pi_w, \pi_z$	
$n_{wH}$	固定
$X_{wH}$	固定
<b><math>a_{wH}</math></b>	<b>変化させる</b>
$Beta(\alpha_w, \beta_w)$	$a_{wH}$ に依存
$\alpha_w = [a_{wH}X_{wH}]$	$a_{wH}$ に依存
$\beta_w = [a_{wH}(n_{wH} - X_{wH})] + 1$	$a_{wH}$ に依存
$ESS_w = \alpha_w + \beta_w$	$a_{wH}$ に依存
$(1 - \pi_z)^{-1}$	固定
<b><math>n_w, n_z</math></b>	<b>変化させる</b>
$X_w, X_z$	$X_w/n_w, X_z/n_z, n_w, n_z$ に依存
<b><math>X_w/n_w, X_z/n_z</math></b>	<b>変化させる</b>
$Beta(\alpha_w + X_w, \beta_w + (n_w - X_w))$	$X_w/n_w, X_z/n_z, n_w, n_z, a_{wH}$ に依存
$Beta(1 + X_z, (n_z - X_z))$	$X_w/n_w, X_z/n_z, n_w, n_z, a_{wH}$ に依存
$P(\pi_w + \Delta > \pi_z   X_w, X_z)$	$X_w/n_w, X_z/n_z, n_w, n_z, a_{wH}$ に依存



# 重み( $a_{wH}$ )と事前/事後分布と事後確率

- まずは事前情報の尤度の重みのみを変えたら  
( $a_{wH} = 0.0, 0.5, 0.7, 1.0$ )どのように
  - 事前分布
  - 事後分布
  - 事後確率

$a_{wH}$	Prior for $\pi_w$	Prior for $\pi_z$	Post for $\pi_w$	Post for $\pi_z$	$P(\pi_w + \Delta > \pi_z   X_z, X_w)$
0	$\pi_w^{-1}$	$(1 - \pi_z)^{-1}$	<i>Beta</i> (25, 976)	<i>Beta</i> (26, 975)	0.885497
0.5	<i>Beta</i> (46, 1787)	$(1 - \pi_z)^{-1}$	<i>Beta</i> (71, 2762)	<i>Beta</i> (26, 975)	0.923707
0.7	<i>Beta</i> (65, 2501)	$(1 - \pi_z)^{-1}$	<i>Beta</i> (90, 3476)	<i>Beta</i> (26, 975)	0.931591
1	<i>Beta</i> (93, 3573)	$(1 - \pi_z)^{-1}$	<i>Beta</i> (118, 4548)	<i>Beta</i> (26, 975)	0.936955

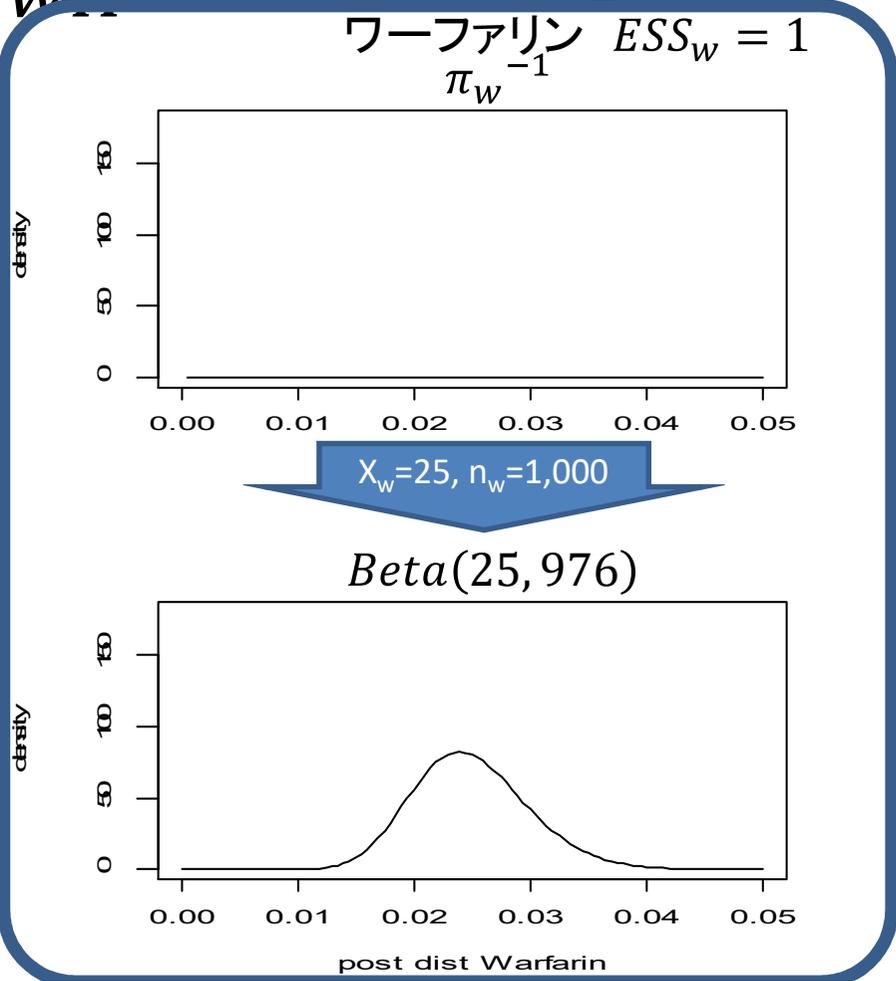
\*他のパラメータ( $n_{wH} = 3665, X_{wH} = 93, n_w = n_z = 1000, X_w = X_z = 25$ )は固定



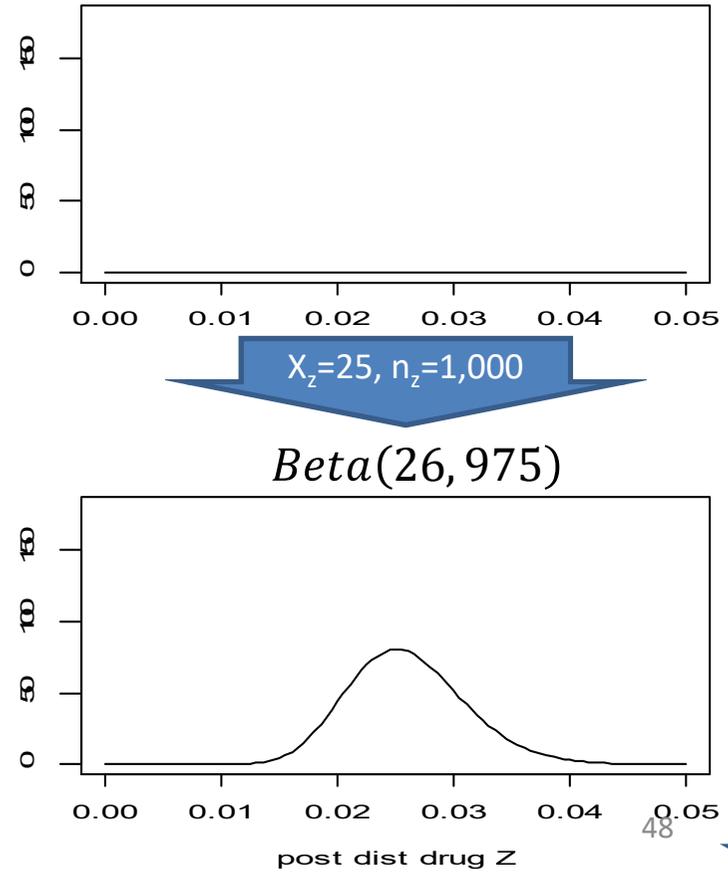
$a_{WH} = 0$  のとき

ワーファリン  $ESS_w = 1$   
 $\pi_w^{-1}$

prior



薬剤Z  
 $(1 - \pi_z)^{-1}$



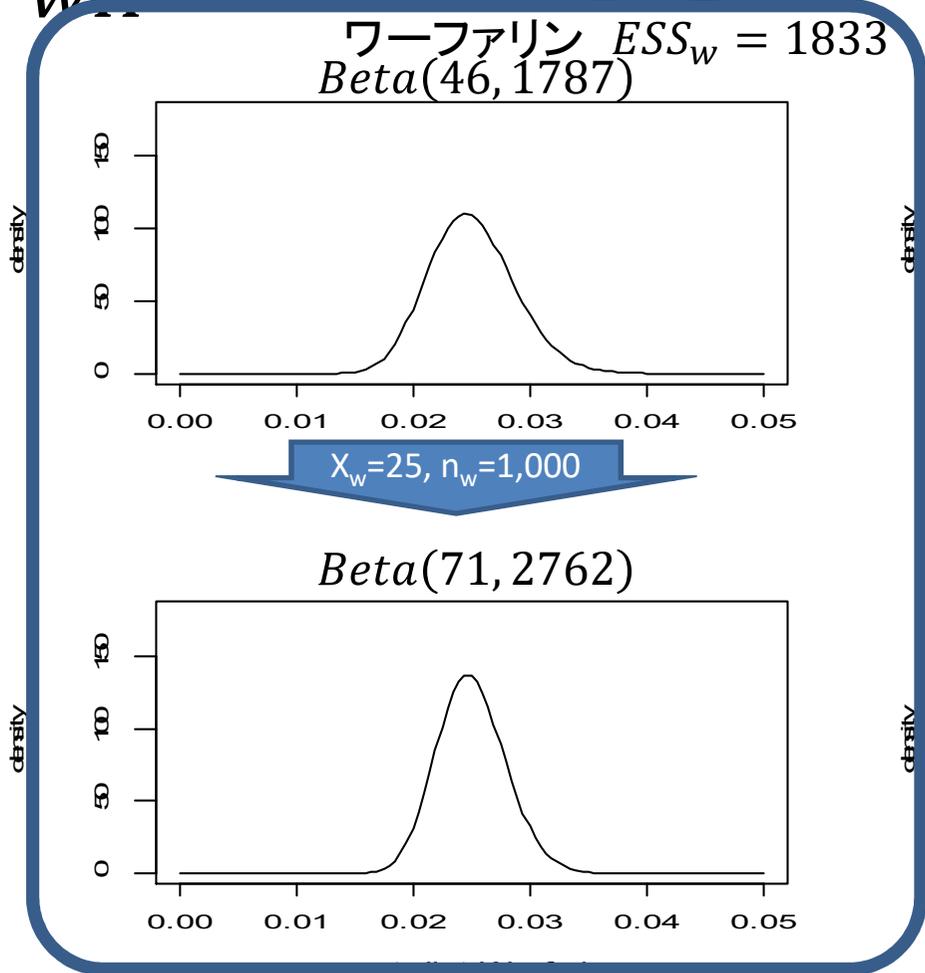
post



# $a_{WH} = 0.5$ のとき

ワーファリン  $ESS_w = 1833$   
 $Beta(46, 1787)$

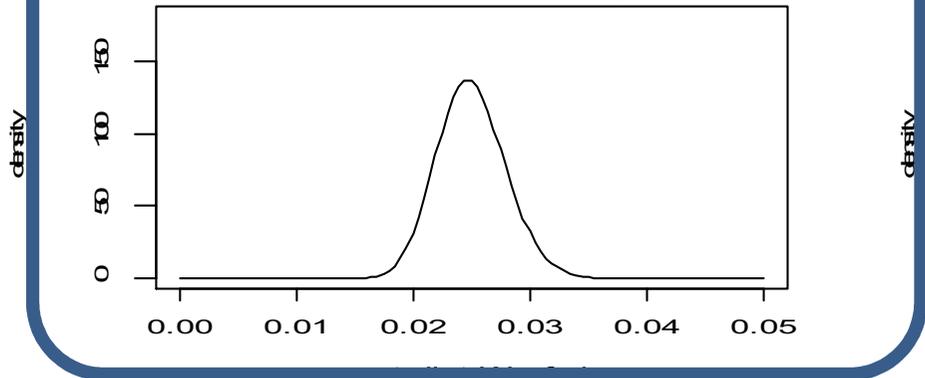
prior



$X_w = 25, n_w = 1,000$

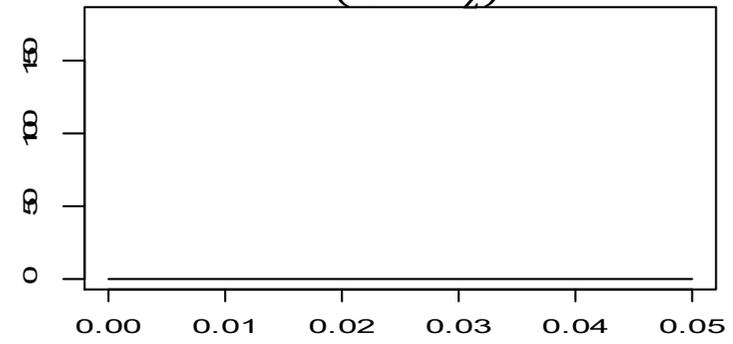
$Beta(71, 2762)$

post



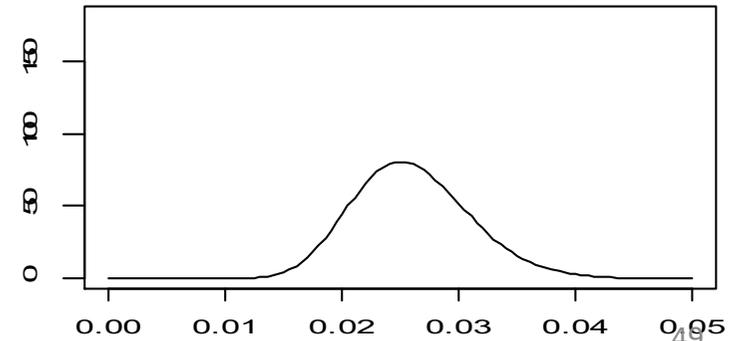
薬剤Z

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



$X_z = 25, n_z = 1,000$

$Beta(26, 975)$



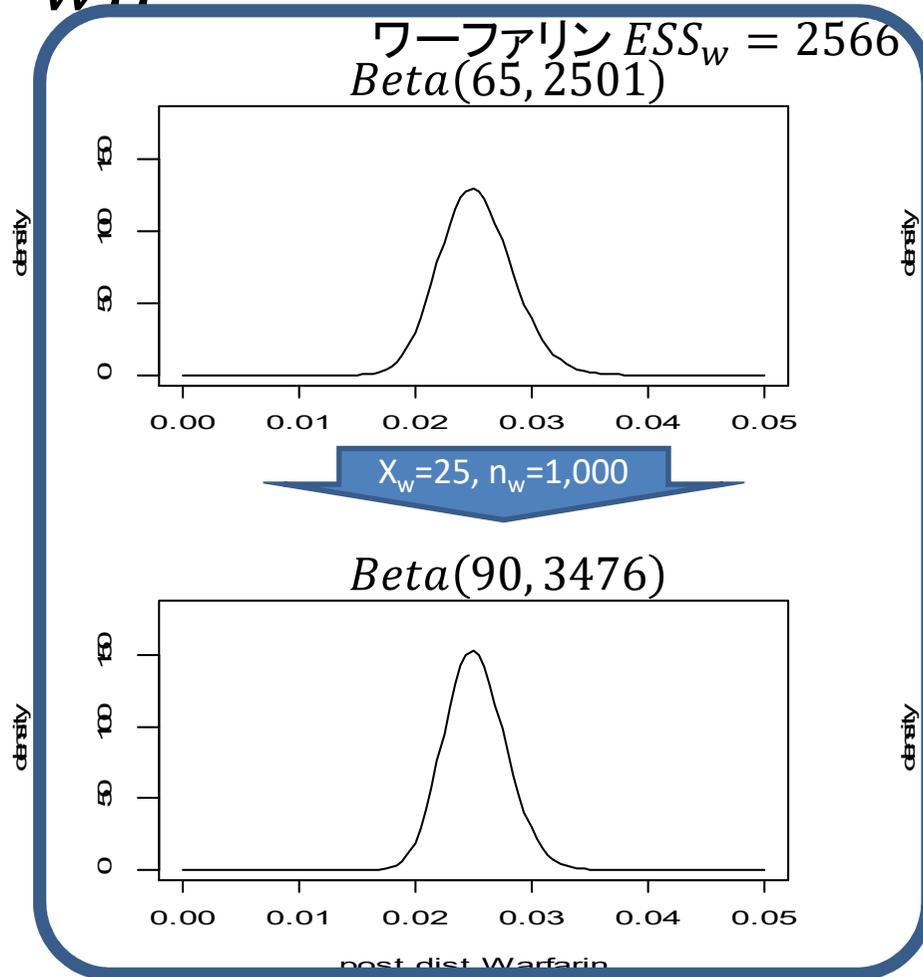
post dist drug Z



$a_{WH} = 0.7$  のとき

ワーファリン  $ESS_w = 2566$   
 $Beta(65, 2501)$

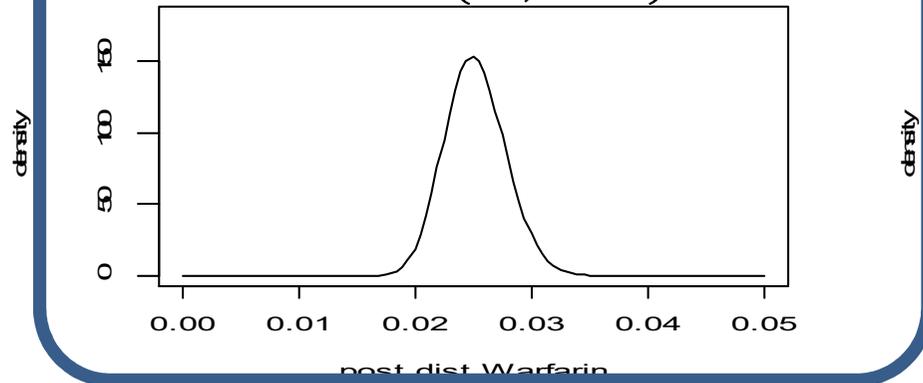
prior



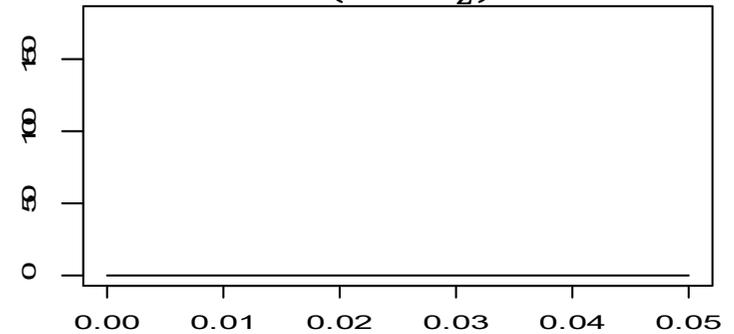
$X_w = 25, n_w = 1,000$

$Beta(90, 3476)$

post

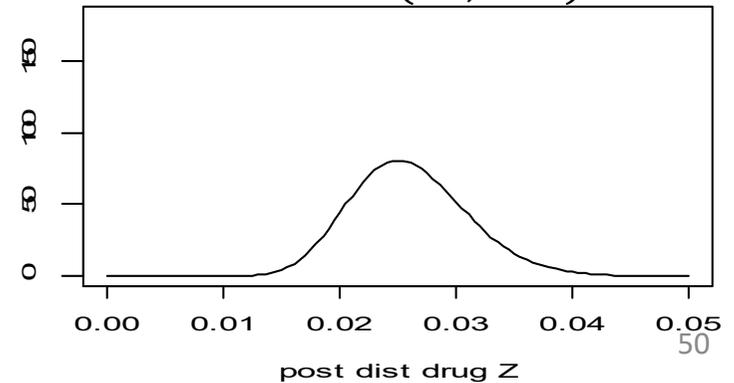


薬剤Z  
 $(1 - \pi_Z)^{-1}$



$X_z = 25, n_z = 1,000$

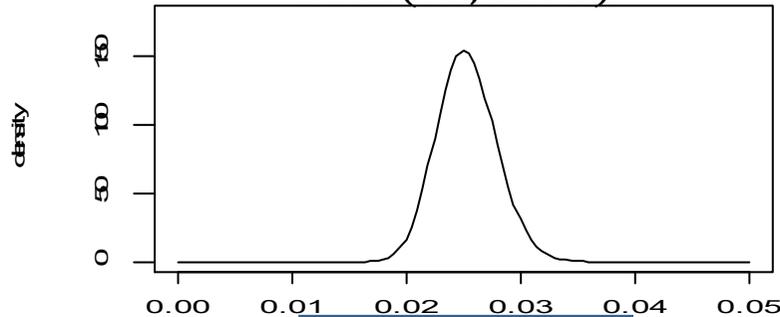
$Beta(26, 975)$



# $a_{WH} = 1$ のとき (再掲)

prior

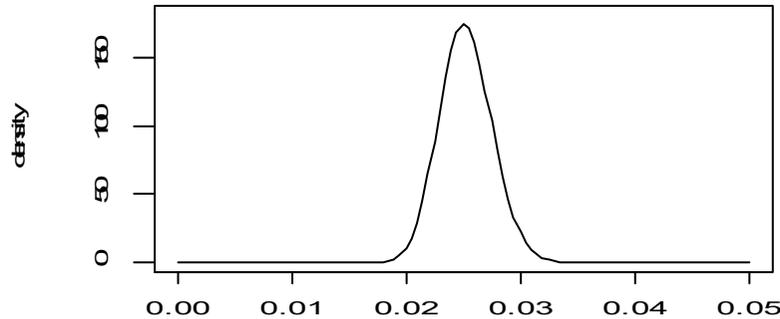
ワーファリン  $ESS_w = 3666$   
 $Beta(93, 3573)$



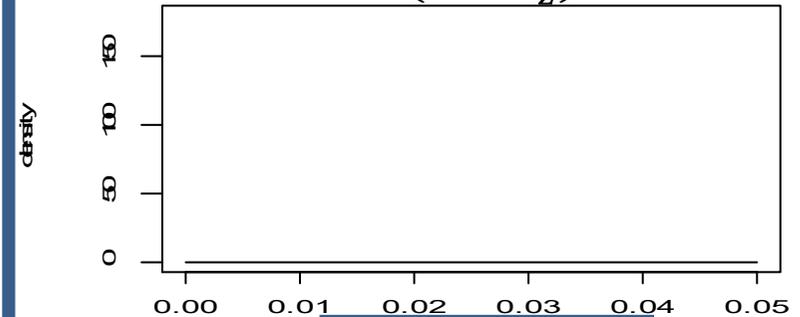
$X_w=25, n_w=1,000$

$Beta(118, 4548)$

post

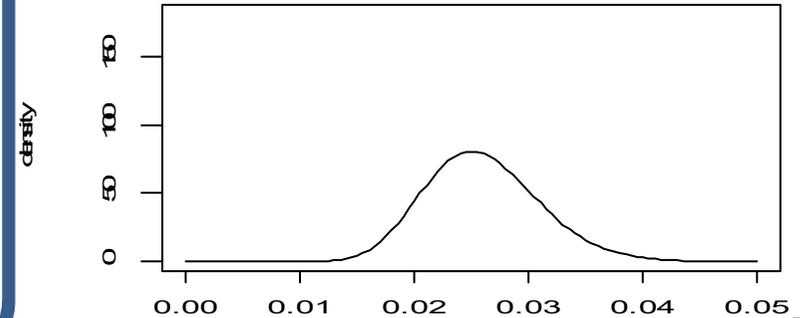


薬剤Z  
 $(1 - \pi_z)^{-1}$



$X_z=25, n_z=1,000$

$Beta(26, 975)$

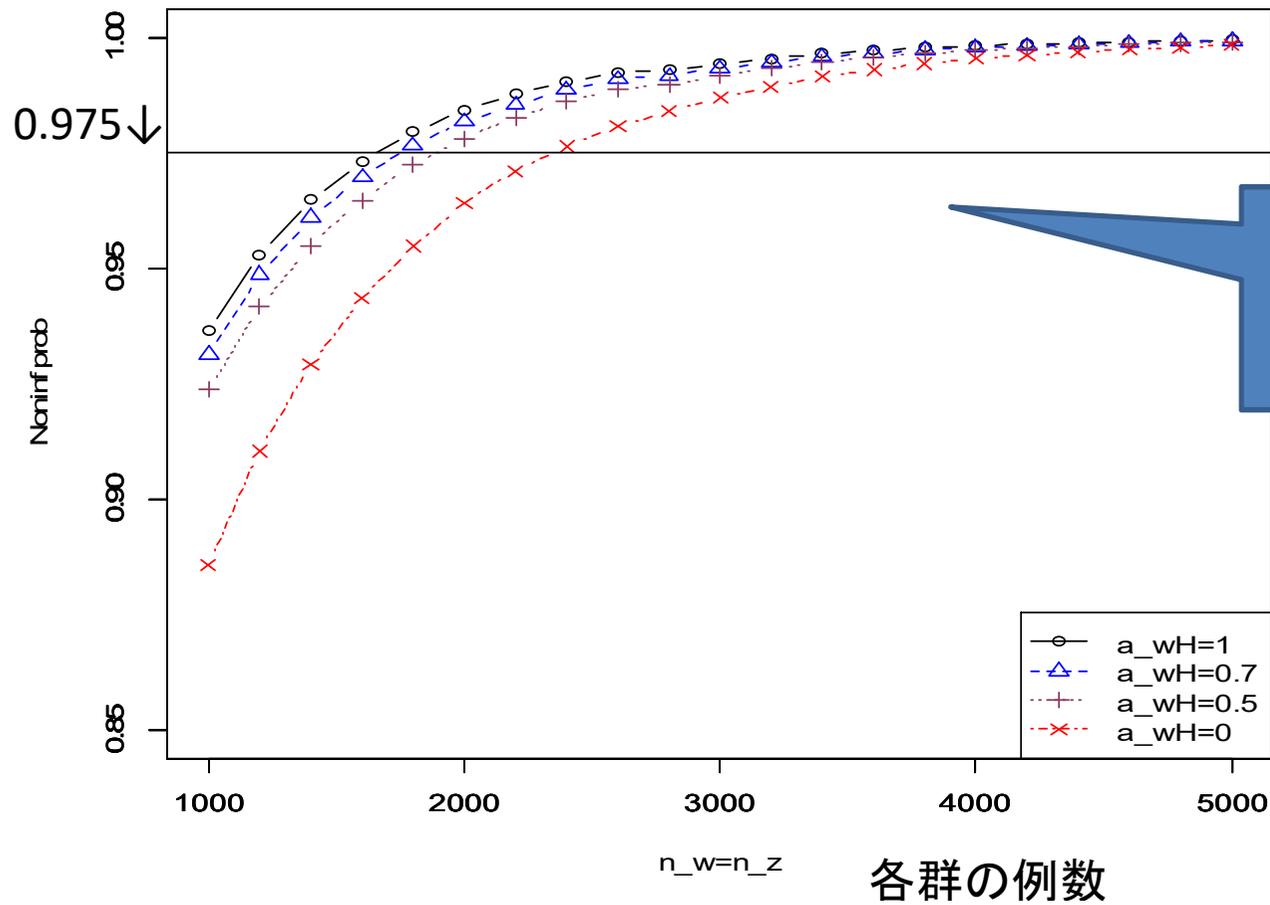


# 重み( $a_{WH}$ )と例数(N/群)と事後確率

- 事前情報の尤度の重みと本試験の例数を変えたら  
( $a_{WH} = 0, 0.5, 0.7, 1.0, n_w = n_z = 1000, \dots, 5000$ )どのよ  
うに
  - 事後確率が変化するかを見てみます
- 以下の点に注意してください
  - 各群の発現確率 $X_w/n_w = X_z/n_z = 2.54\%$ を固定したうえで  
 $n_w, n_z$ を変化させます。結果として $X_w, X_z$ も変化します。
  - ここでは、1:1に割り付けていることを想定しています。



# 重み( $a_{wH}$ )と例数(N/群)と事後確率



この想定の場合、  
重み( $a_{wH}$ )を50%以上、  
増やしてもあまり、  
変わらなさそう

$a_{wH}$	事前分布の $ESS_w$
1	3666
0.7	2566
0.5	1833
0	1



# 事前情報と本試験の 発現確率の不一致の影響

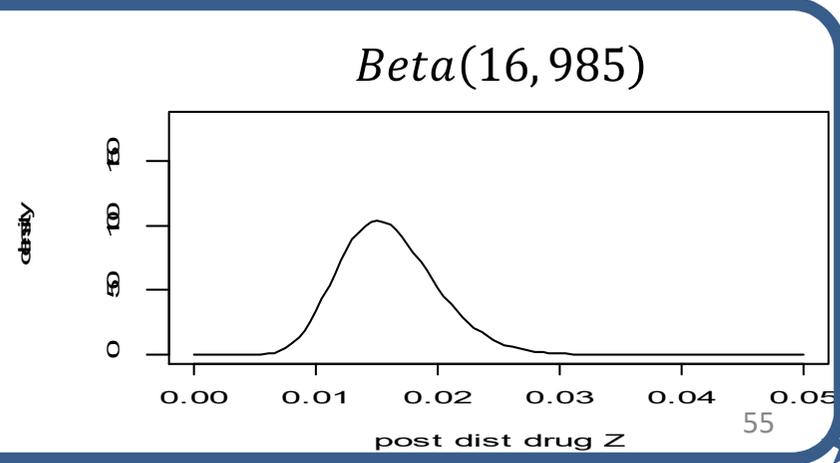
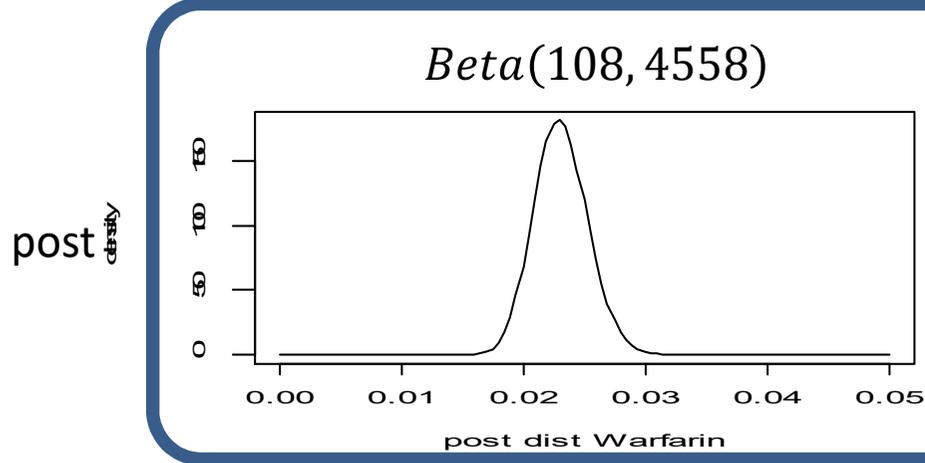
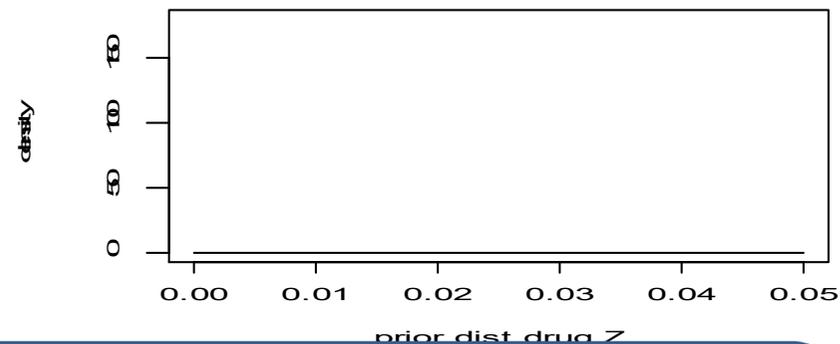
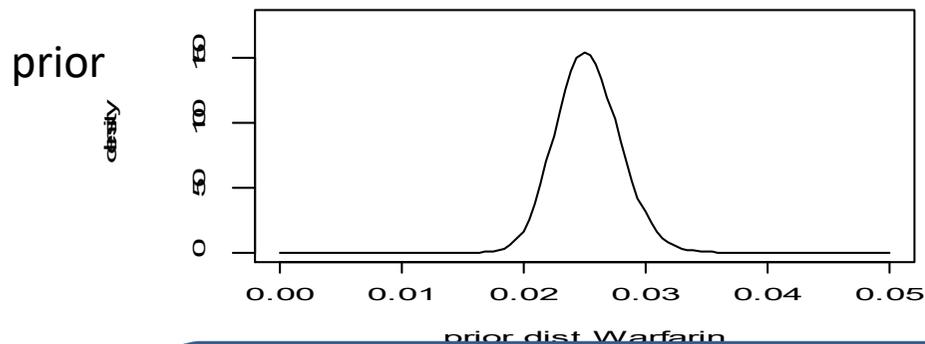
- これまで事前情報と本試験の両群の発現確率は一致すること (2.54%) を仮定してきた
- ワーファリンの事前分布の尤度の重みは  $a_{wH} = 1$  のままで、**本試験で観測される発現頻度が事前情報と異なった場合** ( $X_w/n_w = 1.5\%, 3.5\%$ ) の事後確率を計算する
- 前述のとおり薬剤Zの効果はワーファリンと同等であると仮定しているため、本試験で観測される**薬剤Zも同じ発現割合** ( $X_z/n_z = 1.5\%, 3.5\%$ ) が観測されると仮定する
- 本試験の例数は、1000例/群で均等割り付けとする  
例)  $X_w/n_w = X_z/n_z = 1.5\%$  のとき  $X_w = X_z = 15$  例



$X_w/n_w = X_z/n_z = 1.5\%$ のとき  $P(\pi_z < \pi_w + \Delta | X_z, X_w) = 0.999181$

ワーファリン  $ESS_w = 3666$   
 $Beta(93, 3573)$

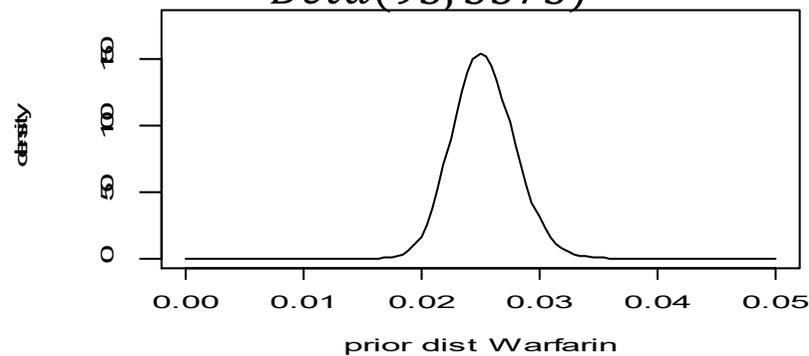
薬剤Z  
 $(1 - \pi_z)^{-1}$



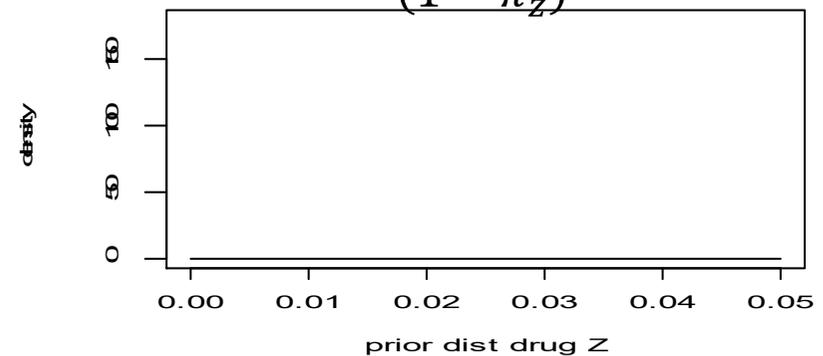
$X_w/n_w = X_z/n_z = 3.5\%$  のとき  $P(\pi_z < \pi_w + \Delta | X_z, X_w) = 0.571491$

ワーファリン  $ESS_w = 3666$   
 $Beta(93, 3573)$

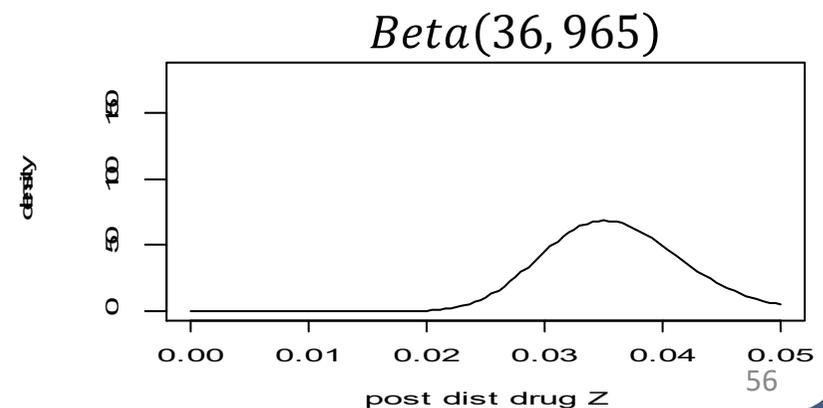
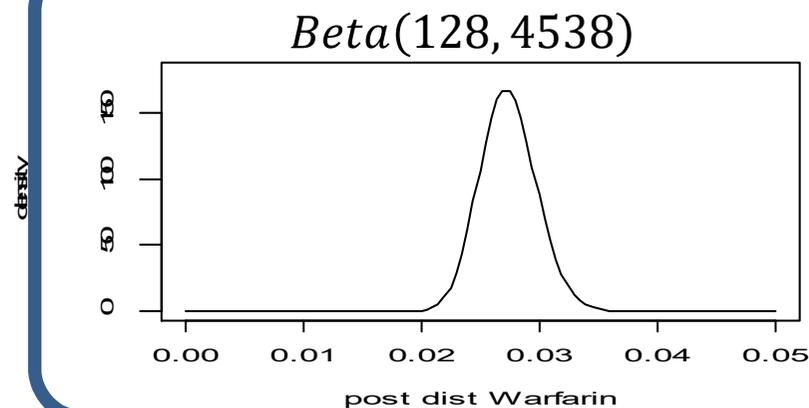
prior



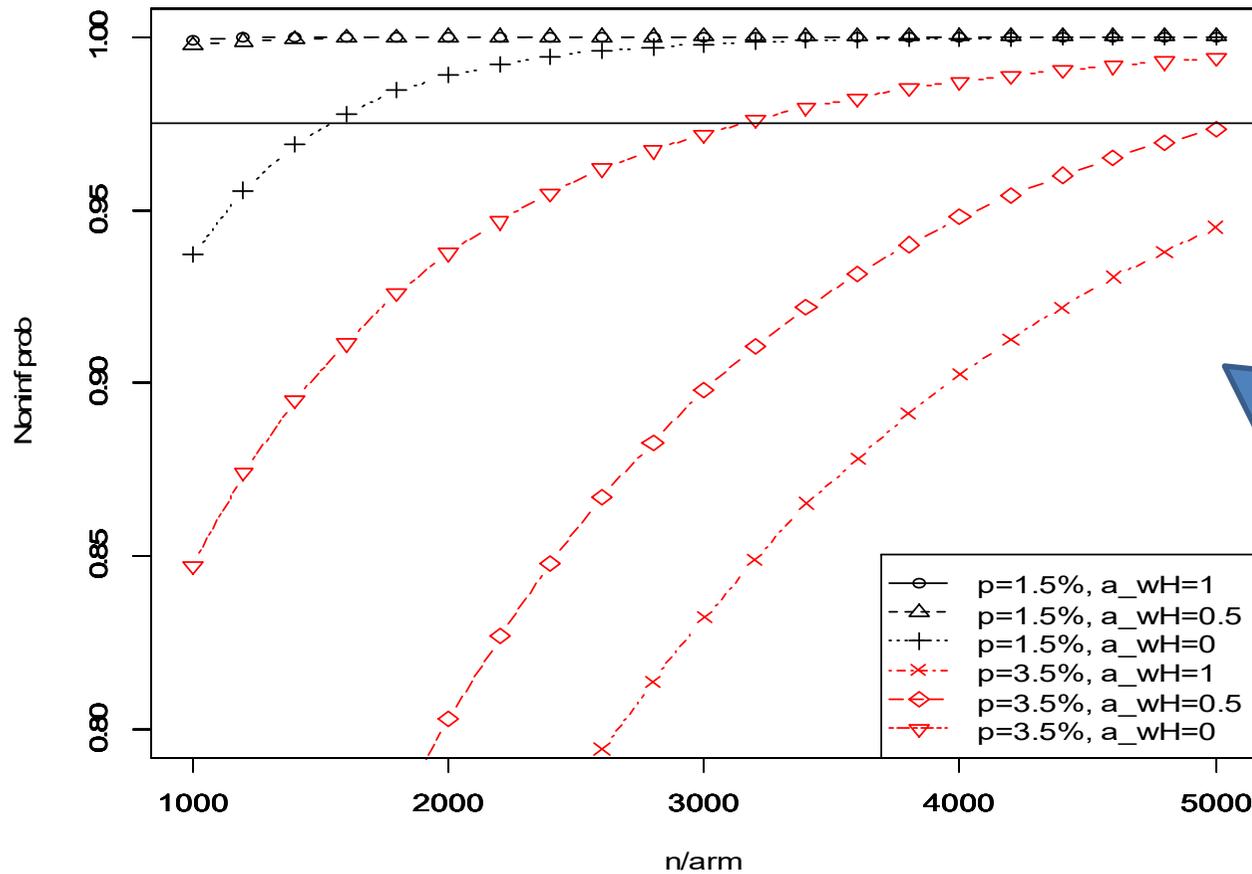
薬剤Z  
 $(1 - \pi_z)^{-1}$



post



# 例数(N/群)と観測発現確率と事後確率



確率変数	事前分布
$\pi_w$	$Beta(93, 3573)$
$\pi_z$	$(1 - \pi_z)^{-1}$

事前情報と本試験の発現確率が異なると, 事後確率は大きく変化する

特に, 本試験の観測される発現確率が過去試験より大きいと, 事後確率は低くなる

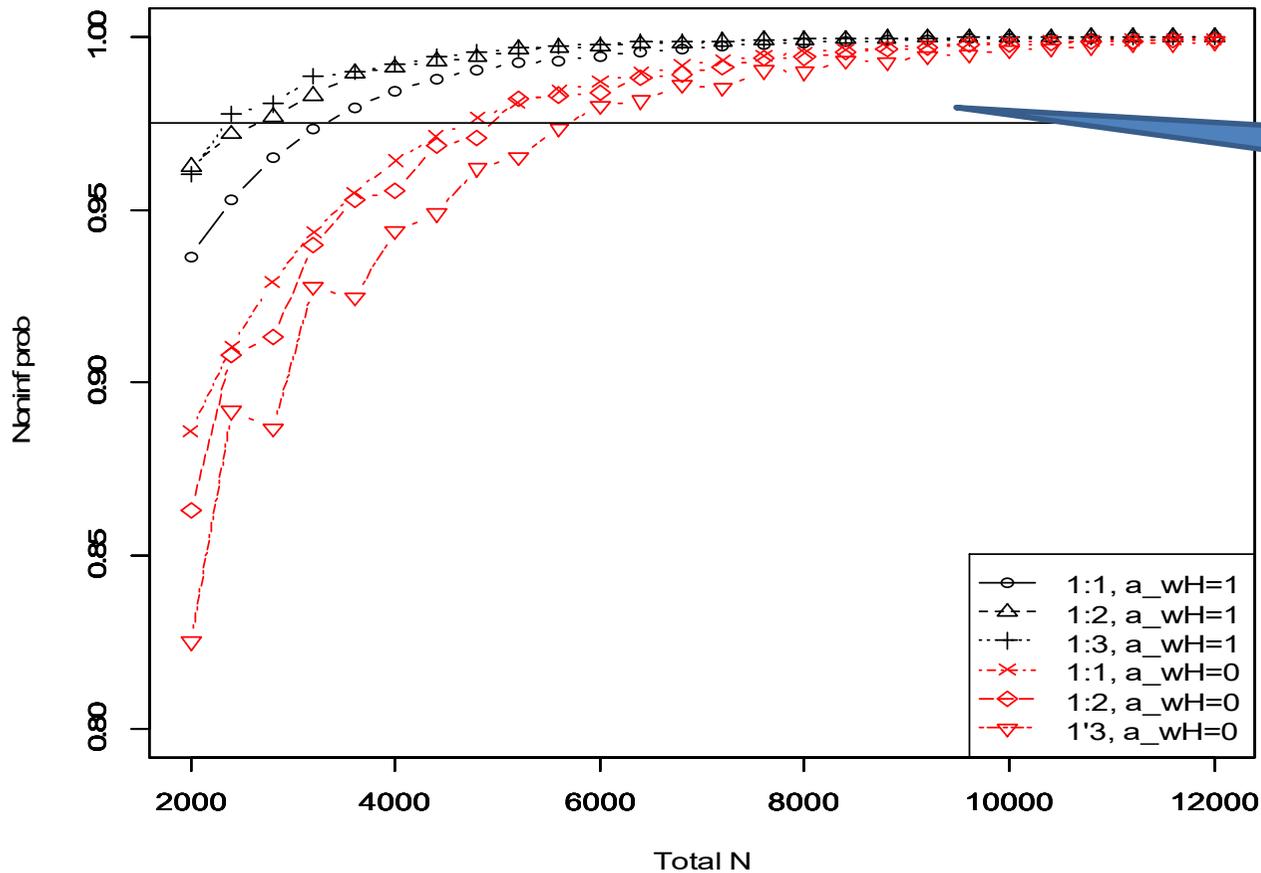


# 非対称割り付け(1:2,1:3)

- 今回の状況では、ワーファリン群は事前情報を多く持っているが、薬剤Z群は事前情報を持たないため、今回の試験のデータは「薬剤Z群の方が多し」方が適切かもしれない
- 以下の場合の事後分布を複数の合計例数毎に求める
  - 割り付け割合を、1:1, 1:2, 1:3に割り付けた場合
  - 重みを0および1とした場合



# 合計例数(N)と割り付け割合と事後確率



ほとんど同じになる

確率変数	事前分布
$\pi_w$	$Beta(93, 3573)$
$\pi_z$	$(1 - \pi_z)^{-1}$

$a_{wH}$	$ESS_w$
1	3666
0	1



# 今回の想定における事後確率の性質

- 異なる想定で事後確率を計算した
  - 事前情報の重み
    - 事前情報の発現確率と観測される発現確率が一致しているとき, 1:1割り付けでは重み( $a_{wH}$ )は0.5以下のとき**事後確率を下げた**
  - 事前情報の発現確率と観測される発現確率の不一致
    - 発現確率が事前情報より高い場合は, **事後確率を大きく下げた**
  - 不均等割り付けの影響
    - 重み( $a_{wH}$ )を1として( $ESS_w = 3666$ ), 本試験の割り付けを1:1から1:2や1:3にした場合は, **事後確率を上げた**が, 1:2と1:3で顕著な差は見られなかった
    - 特に本試験の合計例数が多い (例えば $ESS_w$ の2倍以上)の場合は, **事後確率に差が見られなかった**

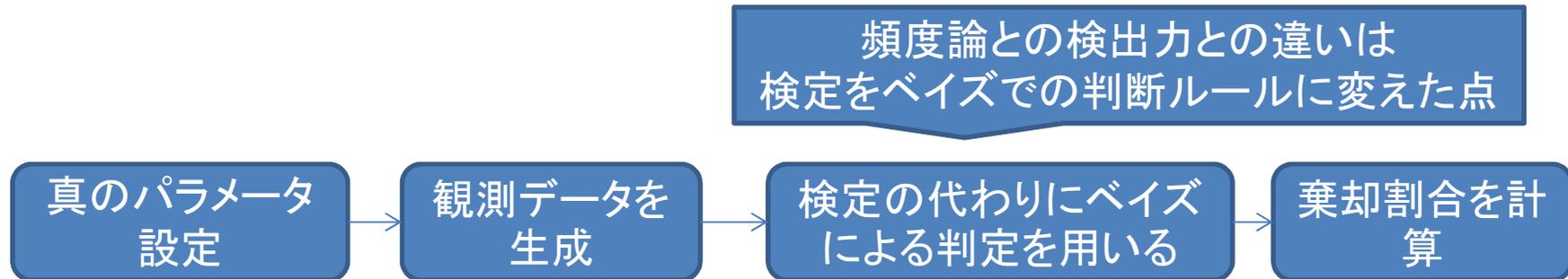
※今回の設定に強く依存した結果であり, 一般的な性質とは限らないことに注意



# ベイズ統計学による意思決定方法の 動作特性

# 動作特性(Operating Characteristics)

- 頻度論では、パラメータを1つに**固定**し、有意な結果が得られる確率をもって検出力としている
- 一方、ベイズ統計学ではパラメータは確率変数であるため**頻度論と全く同じ定義での検出力を定義することはできない**
- ここでは、事後確率での判定を頻度論の検定のように扱うことで（下図参照）、頻度論のように便宜的に評価する(動作特性)



# 何を確率変数として扱うべきか？

- 薬剤Zの開発事例では、2つの不確実な因子が意思決定に用いられている
  - ワーファリンの事前情報として用いられたイベント数 $X_{wH}$
  - 本試験で得られる予定の薬剤Zおよびワーファリンのイベント数 $X_z, X_w$
- ここではイベント数 $X_z, X_w$ の不確実性のみを考慮したシミュレーションを実施することとした
  - 事前情報として用いられた観測されたイベント数( $X_{wH}$ )は十分な調査例数( $n_{wH}$ )に基づいているため
  - 事前情報は重み( $a_{wH}$ )で間接的にコントロール可能であるため

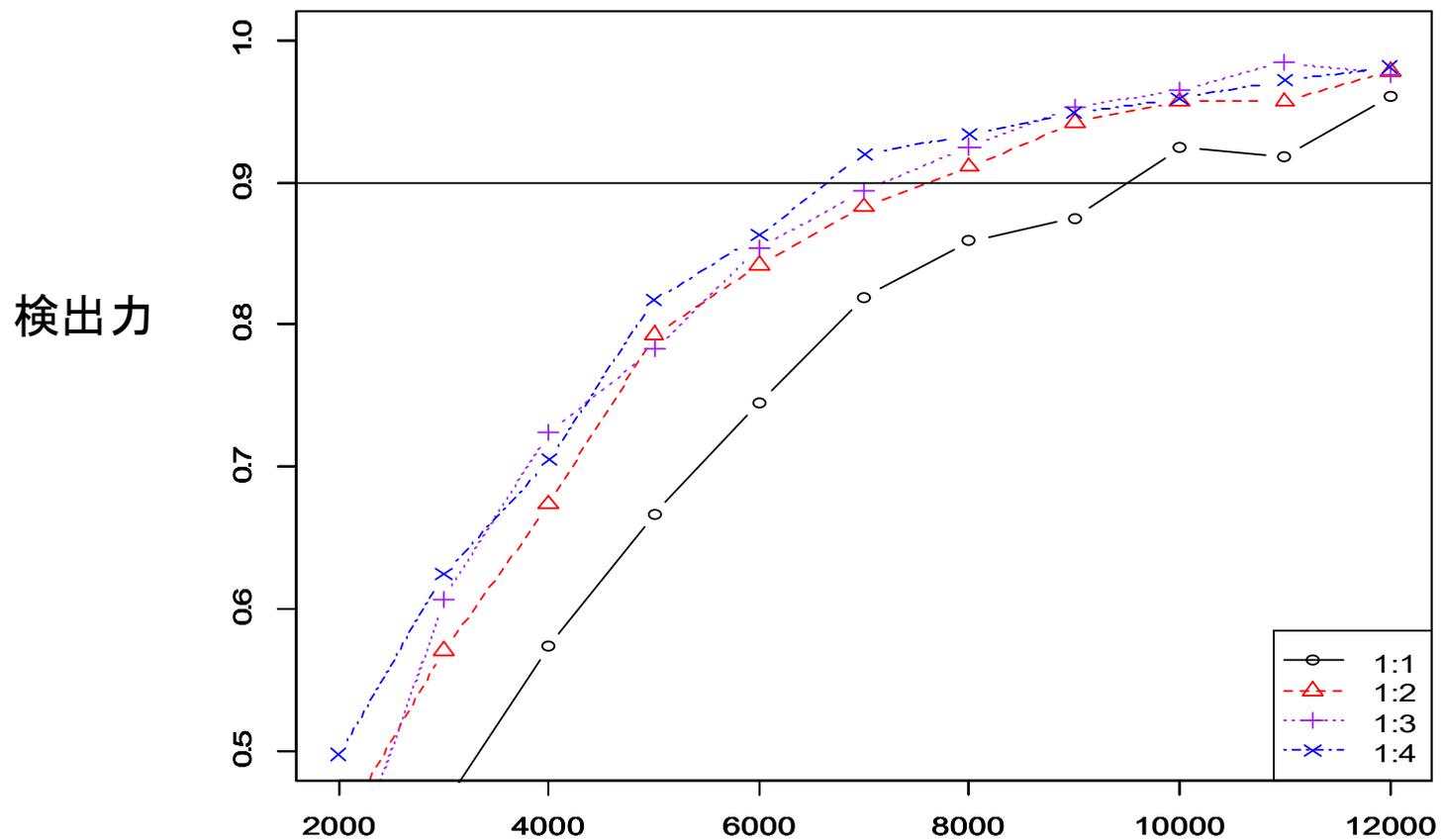
# 動作特性を把握するための シミュレーション

- 以下のシナリオでシミュレーションを行った
  - ワーファリン群と薬剤Zの割り付け比を1:1から1:4とおく
  - 重み( $a_{wH}$ )を1.0, 0.5, 0.0とおく
  - 各群で観測される発現割合は同じであると仮定し,  $X_w/n_w$ および $X_z/n_z$ は次の3つのシナリオを想定する
    - 過去2試験を統合したイベント割合と同じ: 2.5%
    - 過去試験より $\pm 1\%$ 変動したイベント割合: 1.5%および3.5%
  - 検出力を計算するとき
    - 本試験の真の発現確率を $p_w = X_w/n_w, p_z = X_z/n_z$ と設定する
    - $X_w \sim B(n_w, p_w)$ および $X_z \sim B(n_z, p_z)$ からサンプリングする
  - $\alpha$ エラーを計算するとき
    - 本試験の真の発現確率を $p_w = X_w/n_w, p_z = X_z/n_z + \Delta$ と設定する
    - $X_w \sim B(n_w, p_w)$ および $X_z \sim B(n_z, p_z)$ からサンプリングする

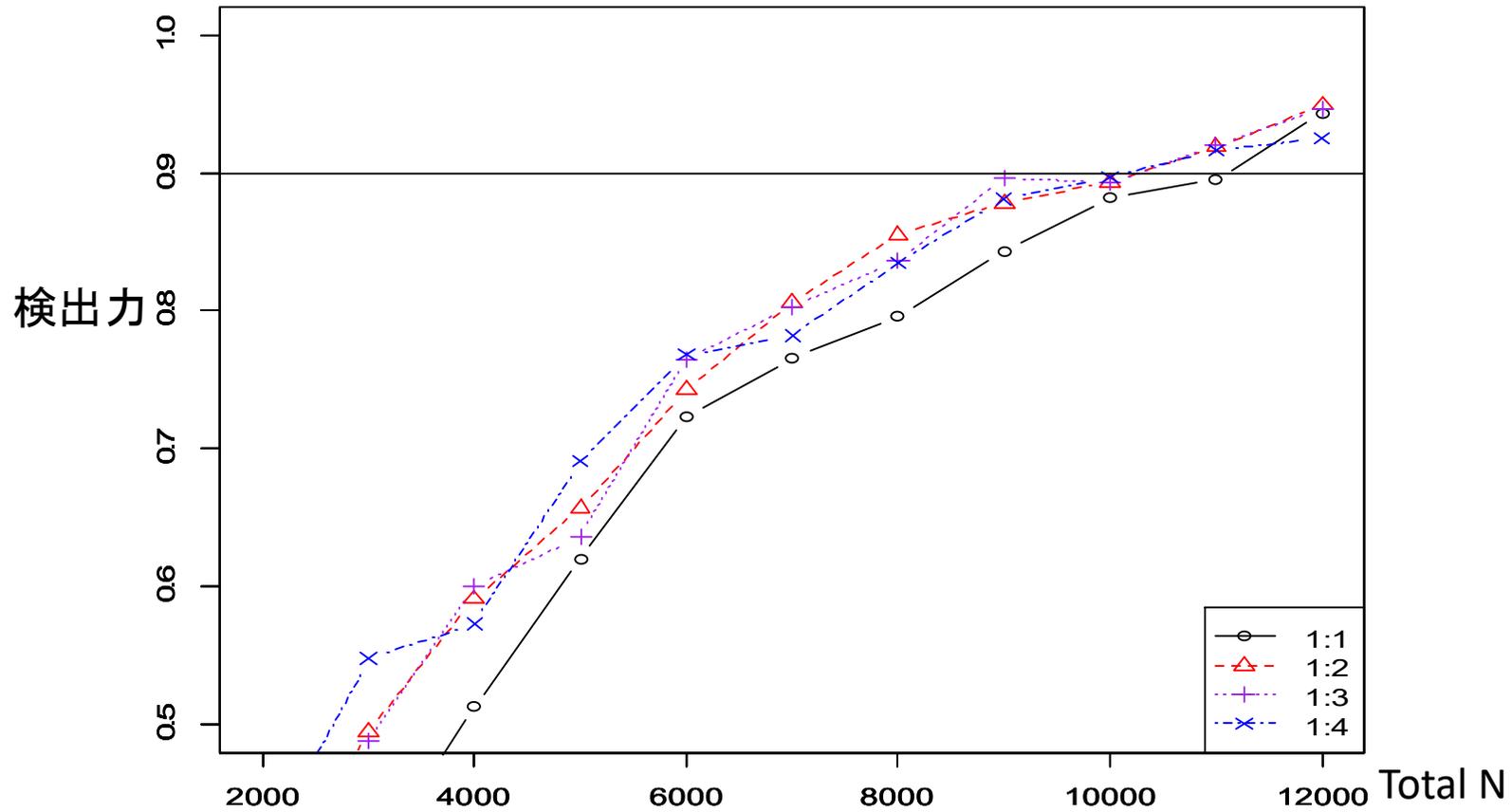
## 本試験の観測発現割合が過去試験と一致する 場合の検出力と $\alpha$ エラー

- 本試験の両群の発現確率が過去試験のワルファリン群と同じ2.5%
- 割り付け比( $n_w:n_z$ )は1:1, 1:2, 1:3, 1:4の4種類
- 重みは1.0, 0.5, 0.0の3種類

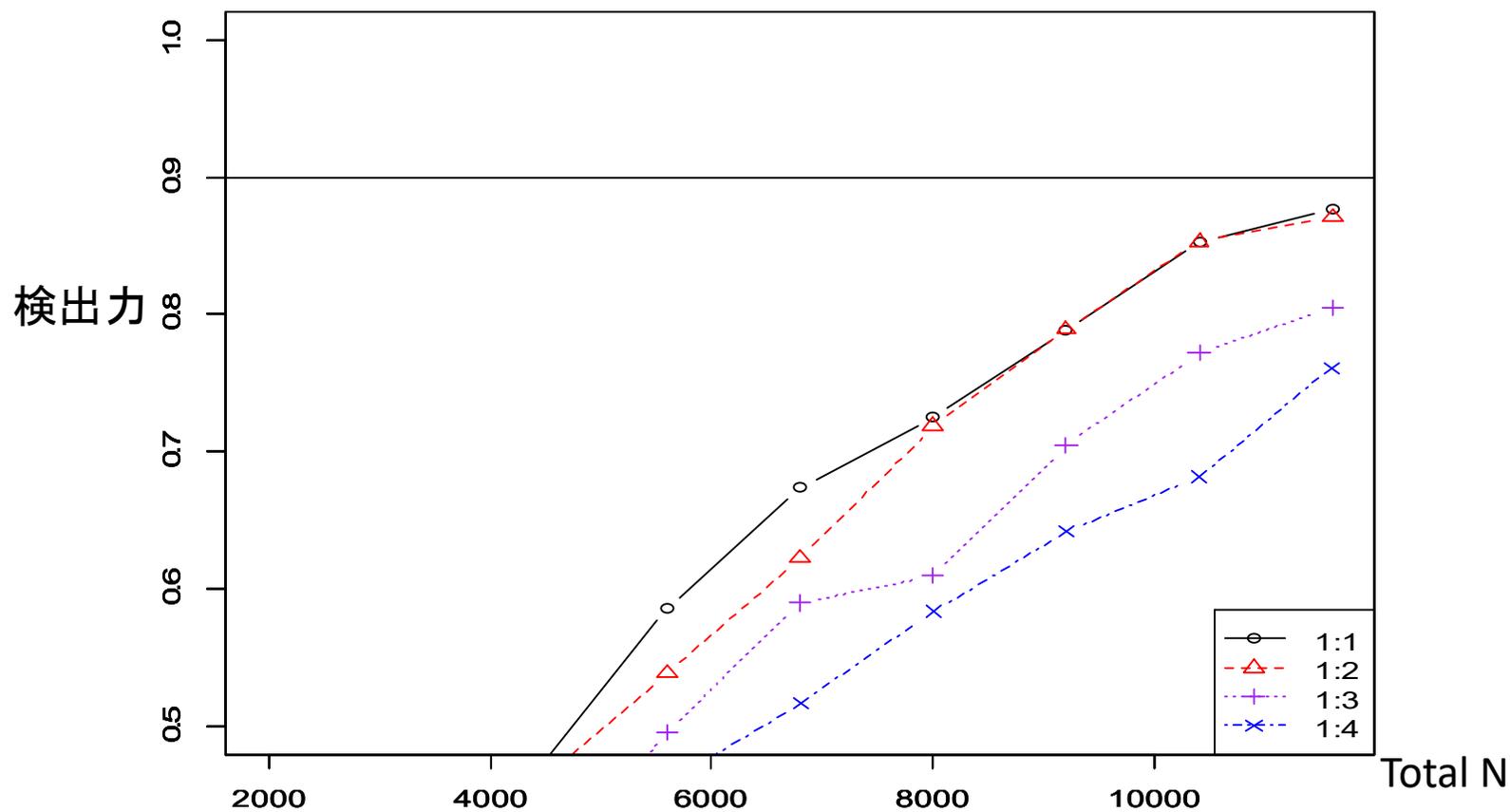
# 本試験の観測発現割合が過去試験と一致する 場合の検出力( $a_{WH} = 1.0$ )



# 本試験の観測発現割合が過去試験と一致する場合の検出力( $a_{WH} = 0.5$ )

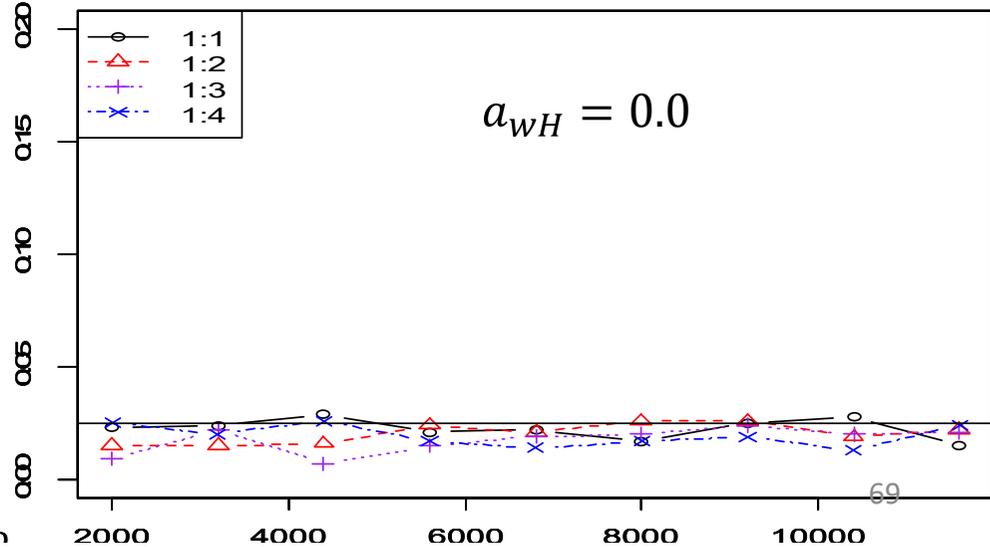
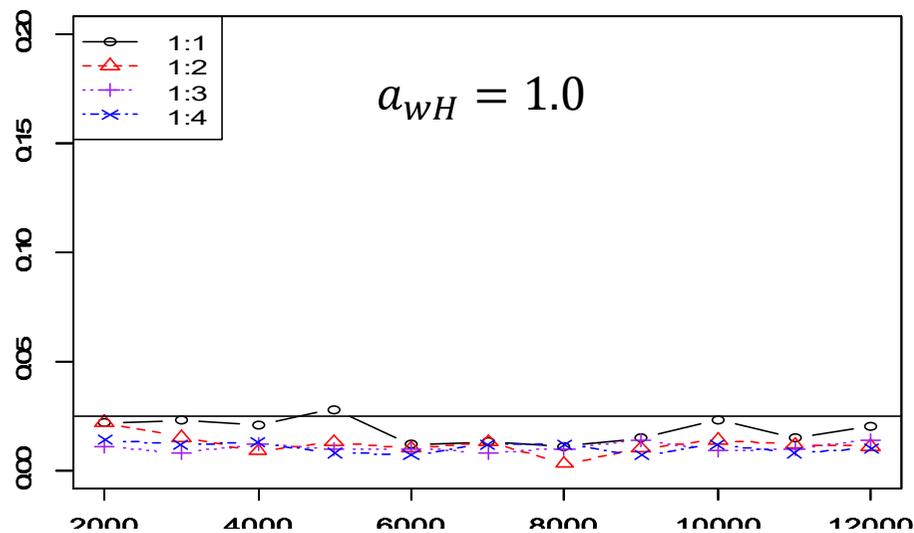
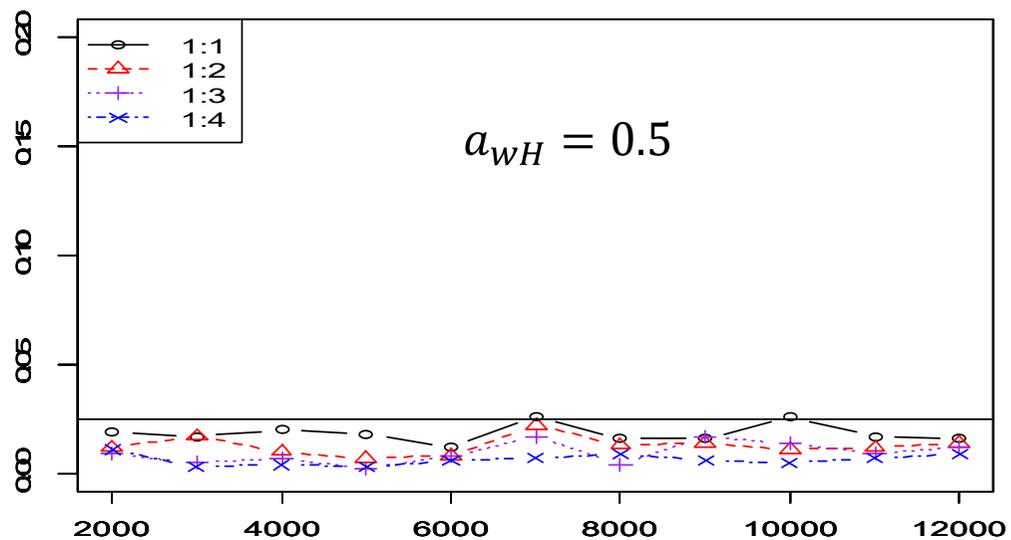


# 本試験の観測発現割合が過去試験と一致する 場合の検出力( $a_{WH} = 0.0$ )



# $\alpha$ エラー

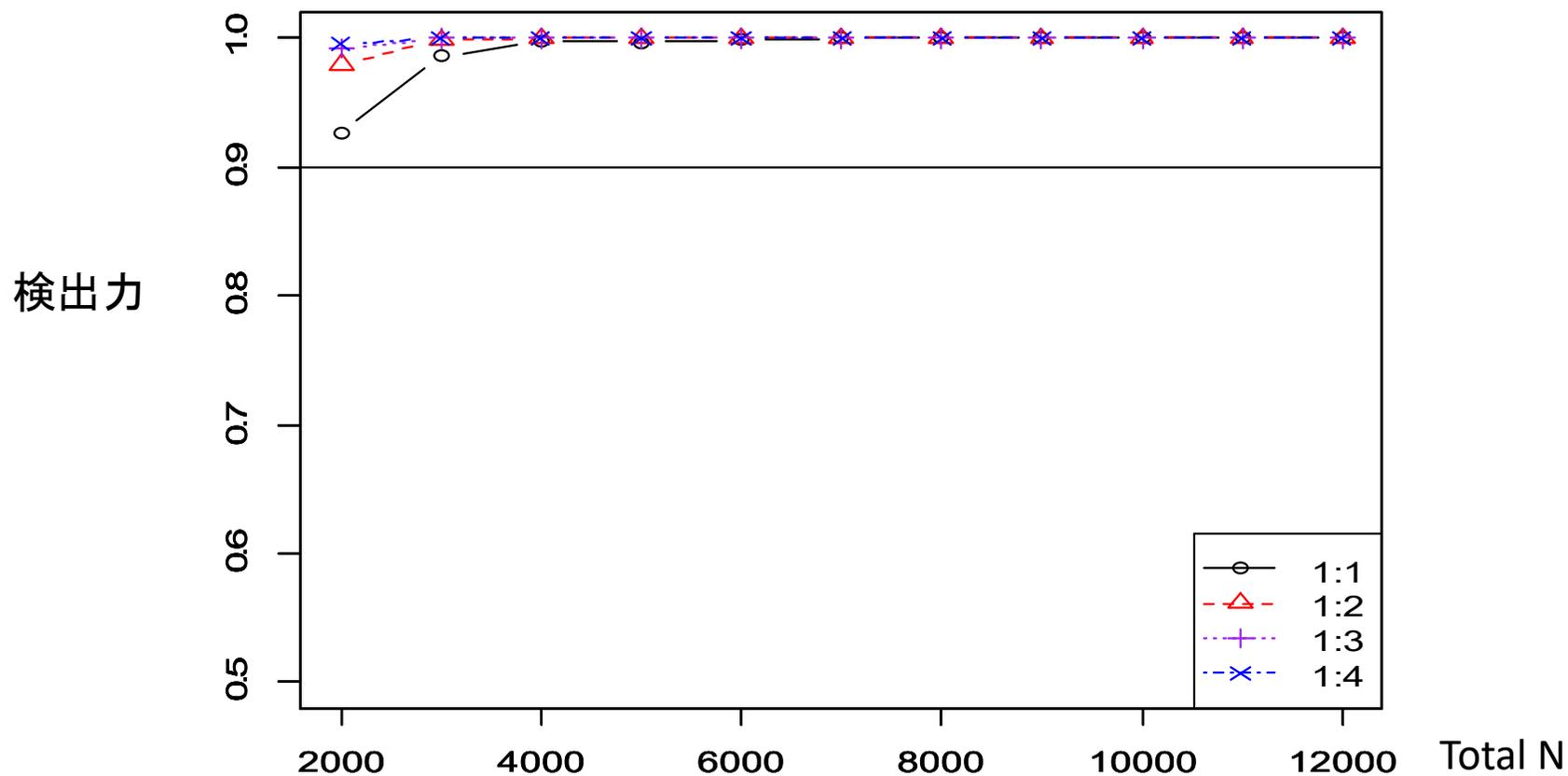
一様に0.025付近におさまる



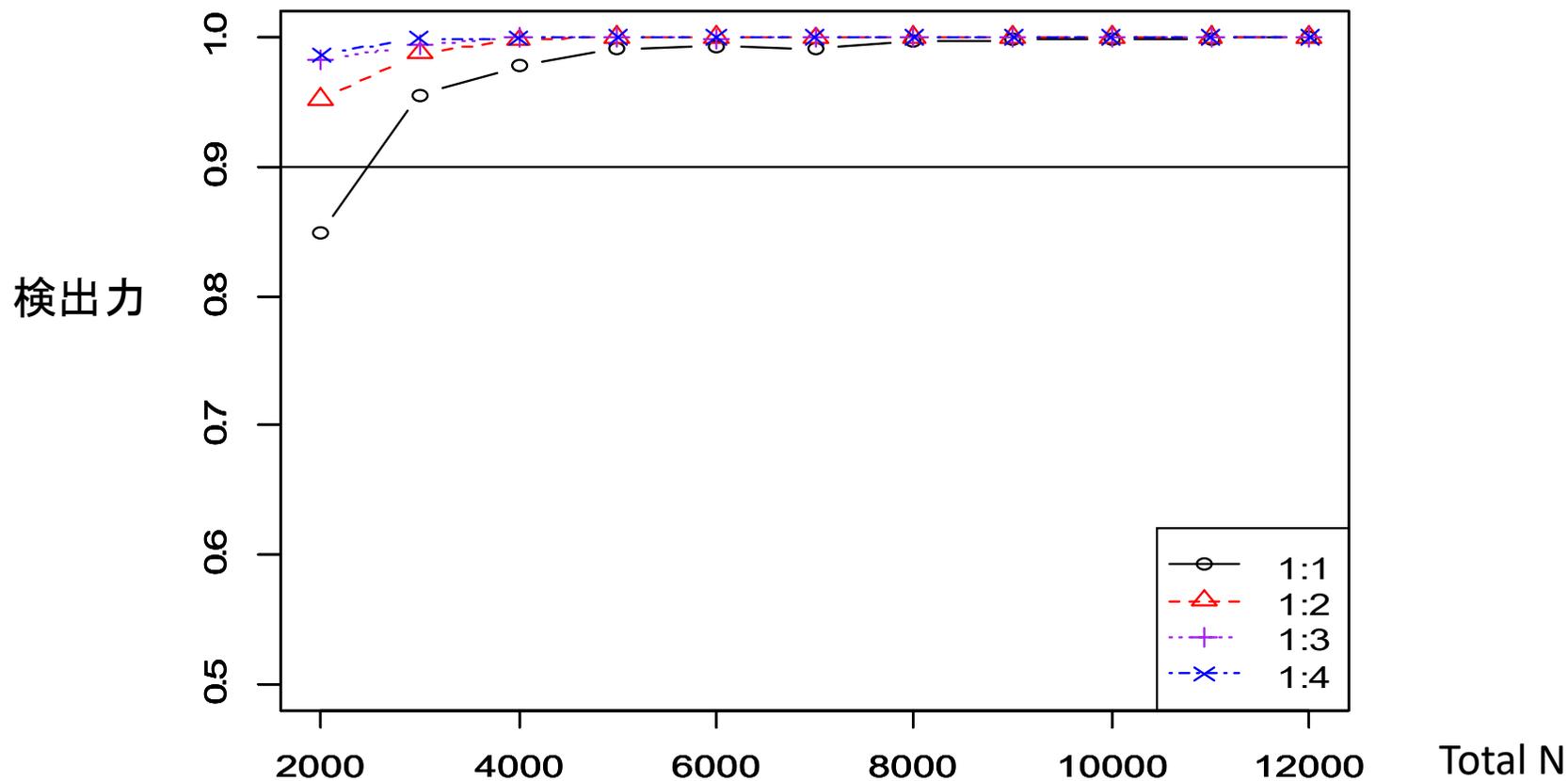
## 本試験の観測発現割合が過去試験より低い 場合の検出力と $\alpha$ エラー

- 本試験の両群の発現確率が過去試験のワルファリン群より低い1.5%
- 割り付け比( $n_w:n_z$ )は1:1, 1:2, 1:3, 1:4の4種類
- 重みは1.0, 0.5, 0.0の3種類

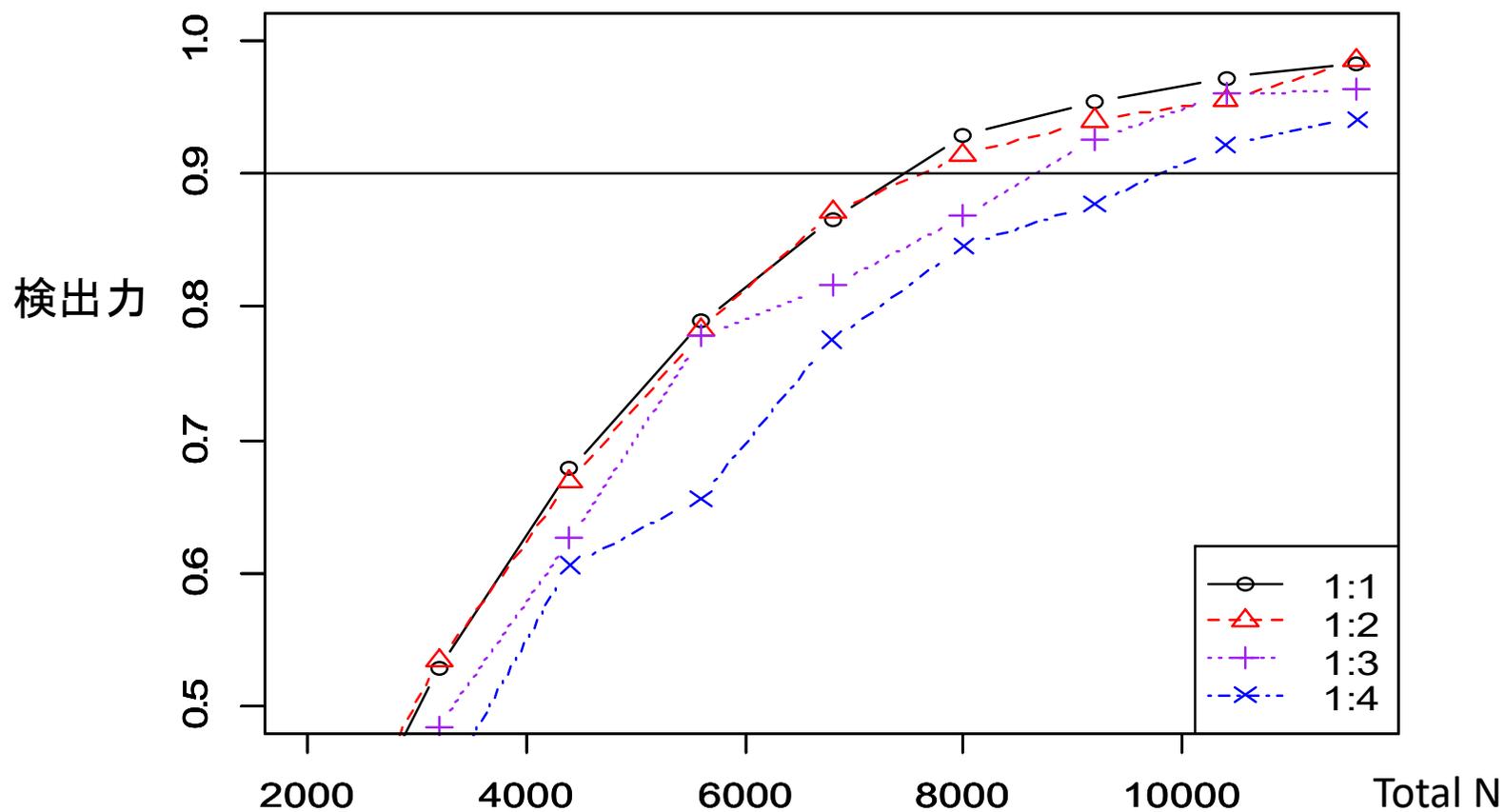
# 本試験の観測発現割合が過去試験より低い場合の検出力( $a_{WH} = 1.0$ )



# 本試験の観測発現割合が過去試験より低い場合の検出力( $a_{wH} = 0.5$ )

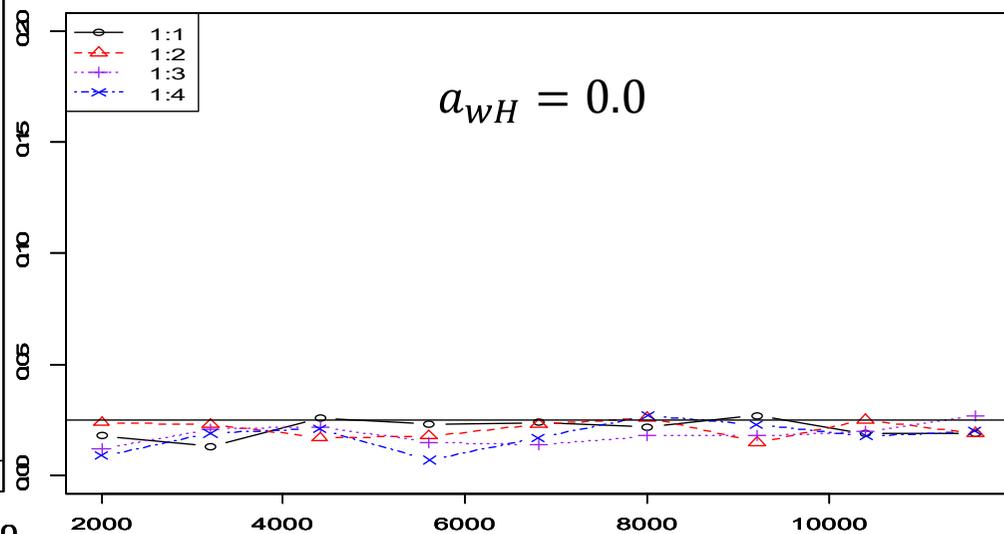
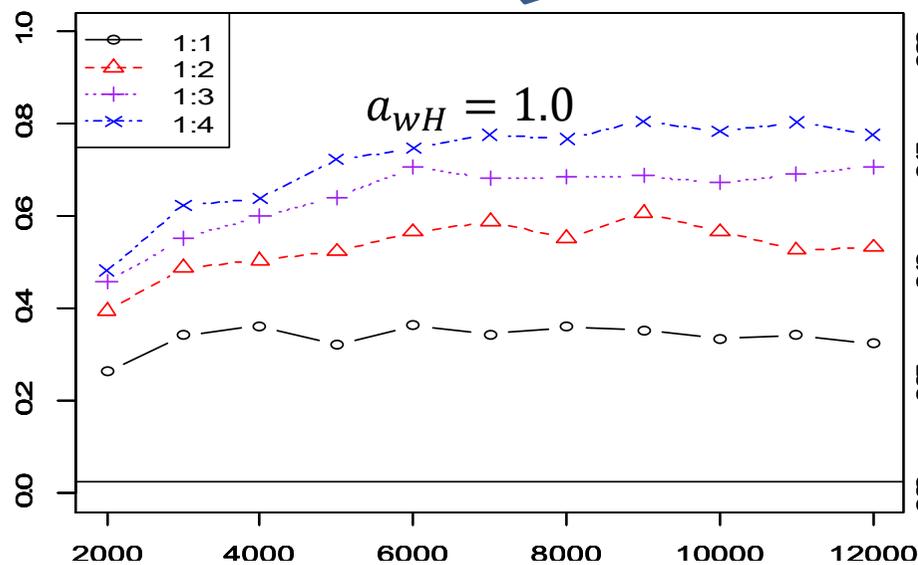
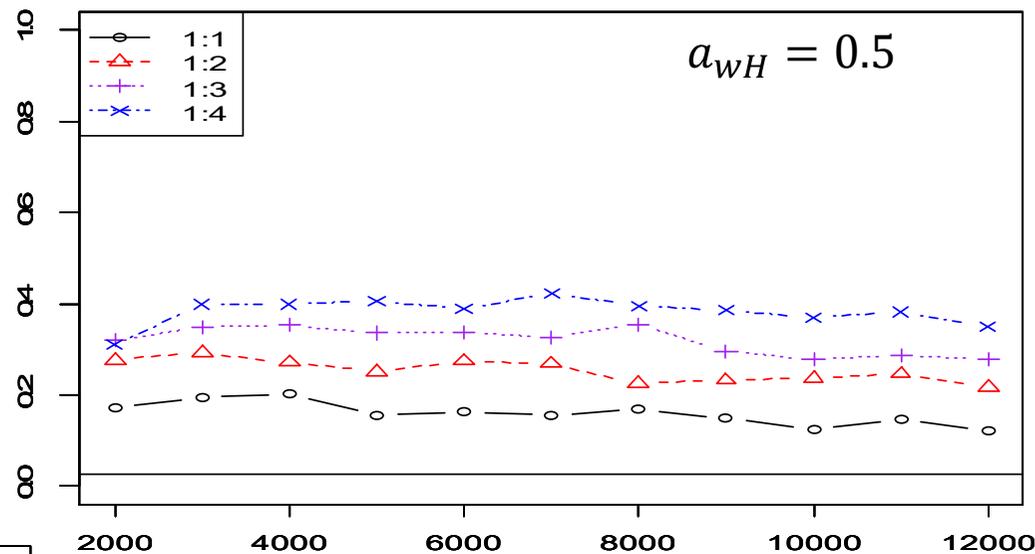


# 本試験の観測発現割合が過去試験より低い場合の検出力( $a_{WH} = 0.0$ )



# $\alpha$ エラー

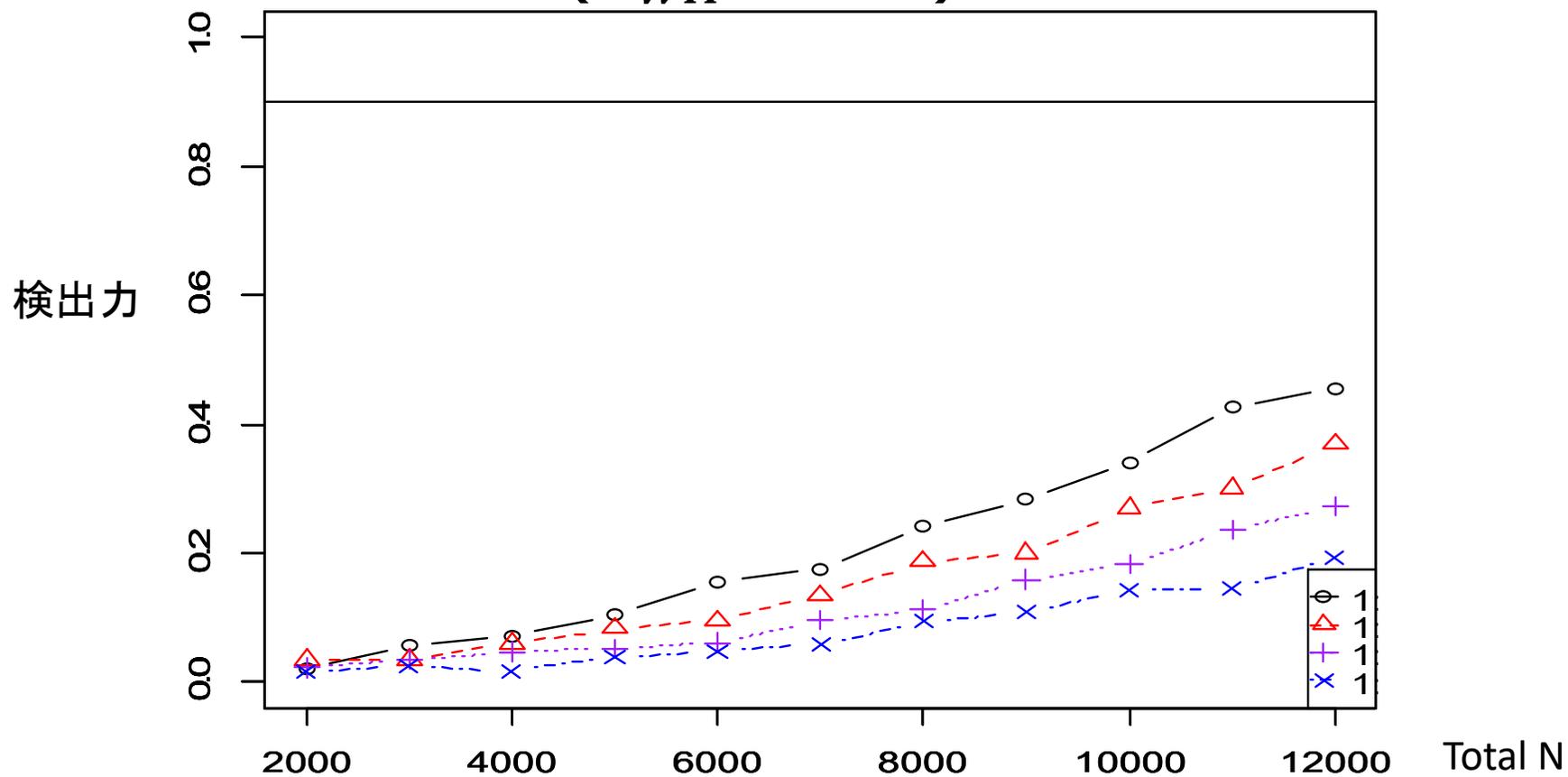
$a_{wH}$ が大きいほど $\alpha$ エラーは多くなる  
(縦軸のスケール注意)



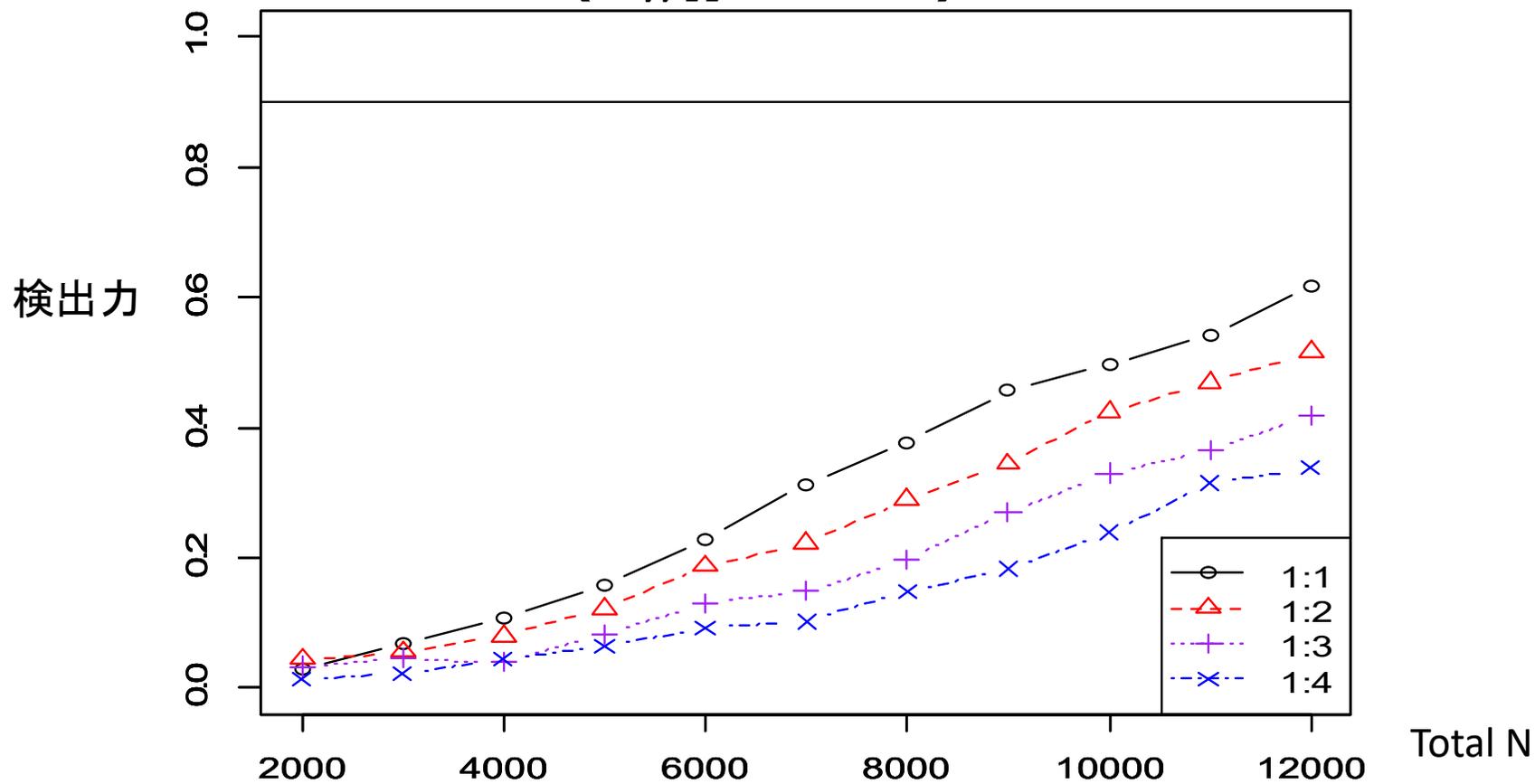
## 本試験の観測発現割合が過去試験より高い 場合の検出力と $\alpha$ エラー

- 本試験の両群の発現確率が過去試験のワルファリン群より高い3.5%
- 割り付け比( $n_w:n_z$ )は1:1 , 1:2, 1:3, 1:4の4種類
- 重みは1.0, 0.5, 0.0の3種類

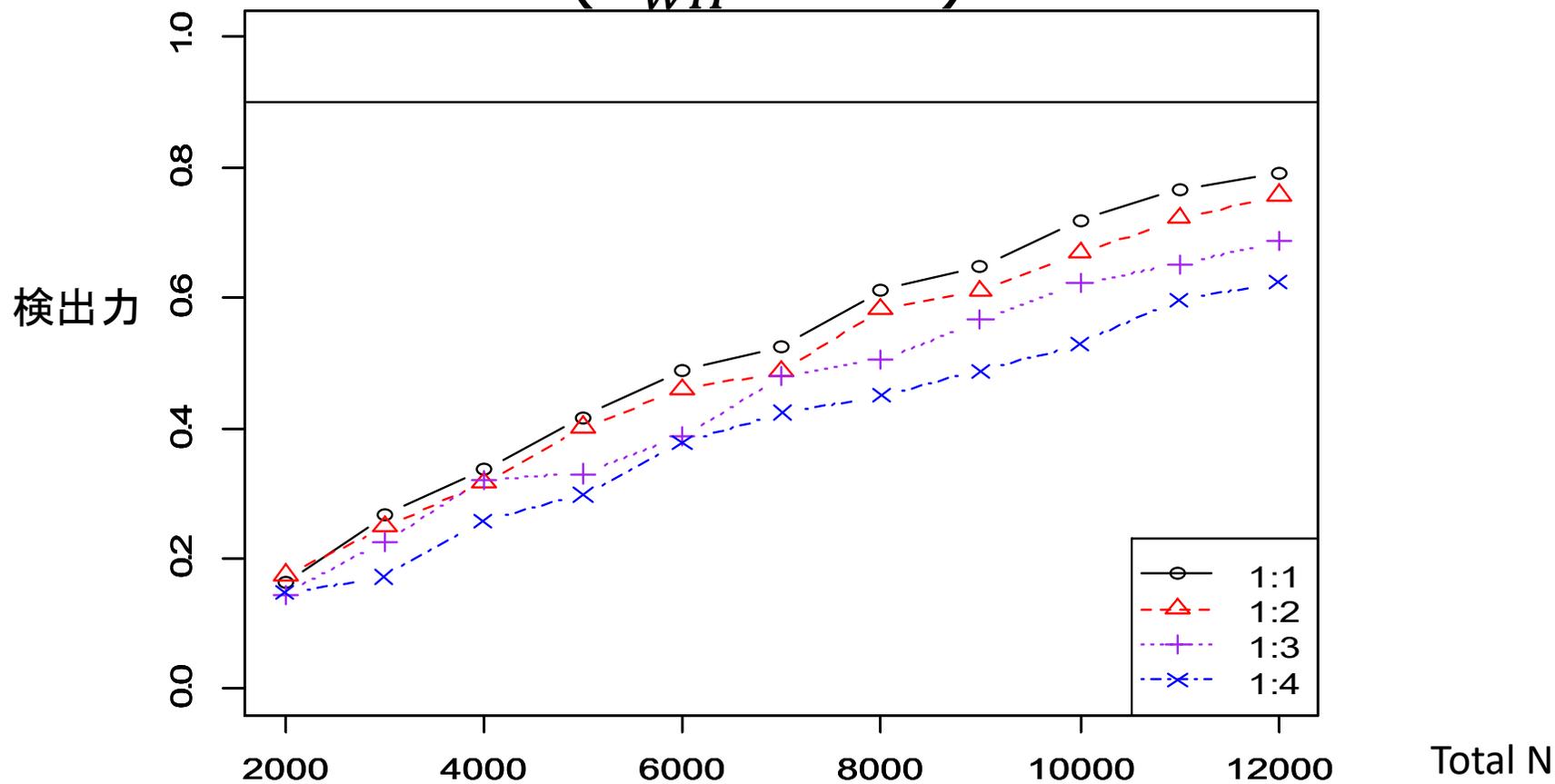
# 本試験の観測発現割合が過去試験より高い場合の検出力( $a_{wH} = 1.0$ )



# 本試験の観測発現割合が過去試験より高い場合の検出力( $a_{WH} = 0.5$ )

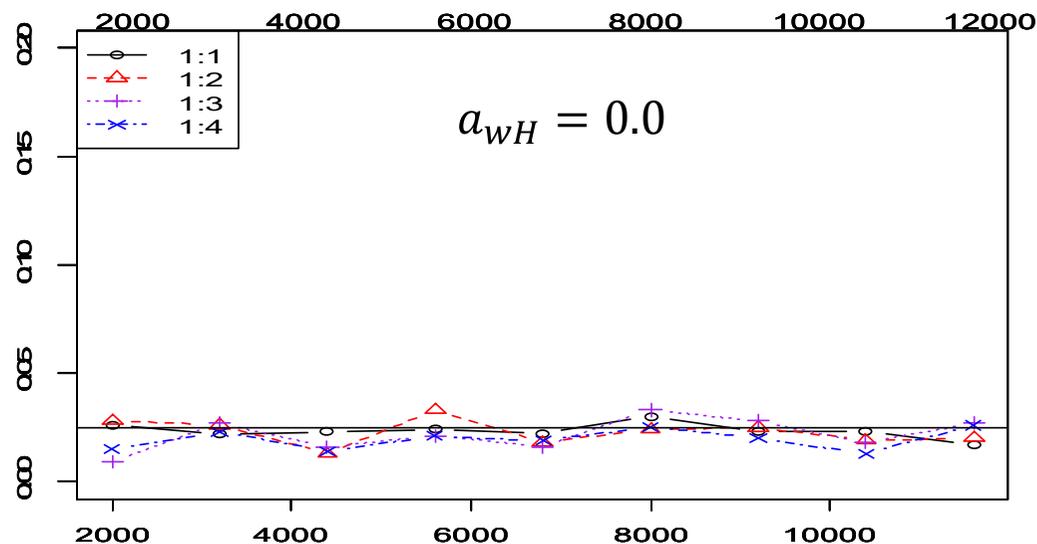
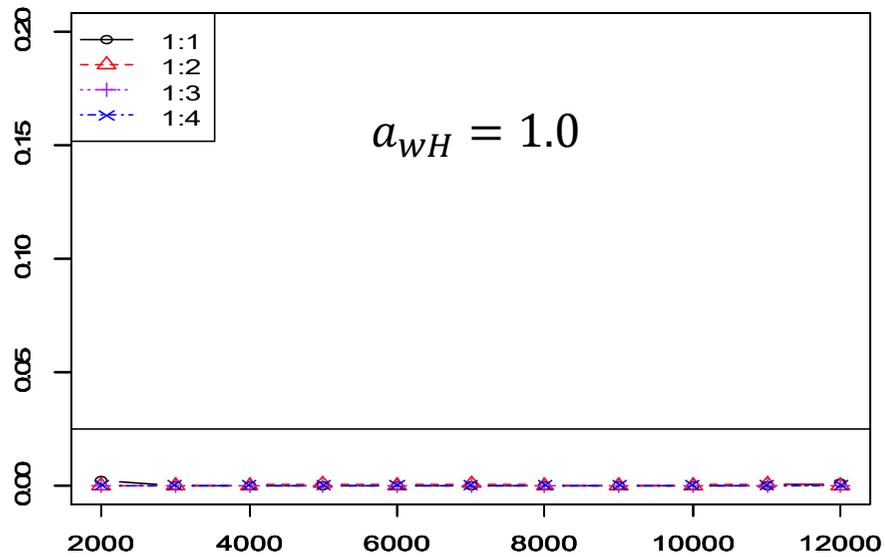
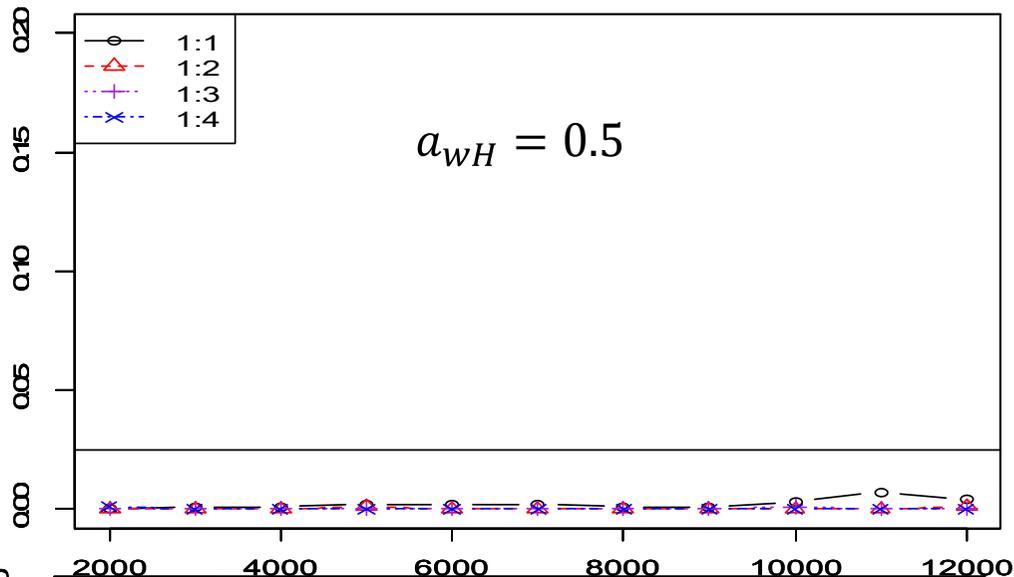


# 本試験の観測発現割合が過去試験より高い場合の検出力 ( $a_{WH} = 0.0$ )



# $\alpha$ エラー

ほとんど棄却されないため $\alpha$ エラーは低い  
重み( $a_{WH}$ )が0の時は $\alpha$ エラーを2.5%付近にできる



# 意思決定の動作特性の検出力と $\alpha$ エラーの計算 (再掲)

- 検出力を計算するとき
  - 本試験の真の発現確率を $p_w = X_w/n_w, p_z = X_z/n_z$ と設定
  - $X_w \sim B(n_w, p_w)$ および $X_z \sim B(n_z, p_z)$ からサンプリング
  
- $\alpha$ エラーを計算するとき
  - 本試験の真の発現確率を $p_w = X_w/n_w, p_z = X_z/n_z + \Delta$ と設定
  - $X_w \sim B(n_w, p_w)$ および $X_z \sim B(n_z, p_z)$ からサンプリング

# ディスカッションポイント

- ここで定義した $\alpha$ エラーおよび検出力の定義はこの意思決定方法の動作特性を特徴づける目的で妥当でしょうか？
- 他の 検出力の定義はあるでしょうか？
- 上の2つの議論を踏まえ、ベイズ流で考えた『検出力』はどのように解釈できるでしょうか？
- ベイズを検出力で評価することをどのように解釈しますか？計算するとしたら、どの部分の不確実性を考慮したいですか？非劣性マージンや事前分布の下となった測定値は確率変数としてシミュレーションすべきでしょうか？
- $\alpha$ エラーを検討する場合はどうでしょうか？

応用編(当日の時間次第では一部を割愛します.)

# 試験計画段階における例数設計

# ベイズ統計学に基づく サンプルサイズの決定

- 本試験のサンプルサイズは、以下を考慮し、サンプルサイズを薬剤Z群6000例、ワーファリン群2000例と決定した
  - 事前分布の重み( $a_{wH}$ )は1とする
  - 発現確率の非劣性マージンを0.94%とする
  - 真の薬剤Zおよびワーファリンの1.5年後のイベント発現確率を2.54%と仮定した場合、薬剤Zのワーファリンに対する非劣性を示す頻度は93%であった
  - 1:2, 1:3, 1:4の割り付け比( $n_w:n_z$ )のとき合計例数8000例で90%以上の検出力を持つと観測され、1:2割り付けよりも1:3, 1:4割り付けの方が検出力が大きかった。また1:3と1:4では検出力にあまり差がなかったため、1:3割り付けが採用された

# ベイズ統計学に基づく サンプルサイズの決定(つづき)

- 中間解析時点のデータに基づき、サンプルサイズの再設計を行う
- 独立モニタリング委員会を組織し、試験開始後に非盲検下でワーフアリン群のイベント発現確率が過去試験と乖離があると考えられる場合は、プロトコルを修正し重みおよび例数を変更するか頻度論をプライマリとすることも検討する
  - 例えば、本試験のワーフアリン群の発現確率が3.5%の時は頻度論を用い、6000例/群と変更する
  - 詳細な閾値や事前分布の重み( $a_{WH}$ )を調整する方法については本日のディスカッションの対象外であるが、考えるべきポイントは整理する



# 頻度論に基づく の決定(比較のため)

- サンプルサイズを6000例/群とすると,
  - 発現確率の非劣性マージンを0.94%
  - 真の薬剤Zおよびワーファリンの1.5年後のイベント発現確率を2.54と仮定した場合, 検出力は90% (ICON社の例数設計ソフトウェアADDPlanを利用)

# ディスカッションポイント

- 例えば以下のようなprior-data conflictが起こることを想定して，計画段階で何をしたらよいか？
  - 発現確率は事前情報で2.5%を想定していたが，本試験では1.5%であった場合
- 薬剤Zにimproperではなく他の分布を用いるべきでしょうか？その他の事前分布を用いる場合，どのような分布が望ましいか？



# ディスカッションポイント

- 独立モニタリング委員会による中間解析について
  - 閾値をどう決めるか？（数値を事前規定するか，手順を事前規定するか？）
  - このような変更を行う際の $\alpha$ エラーは適切にコントロールされるか？
  - これ以降の割り付け比率が，実薬群にかなり偏るが，問題ないか？



# ディスカッションポイント

- 試験デザイン(サンプルサイズ, 意思決定の閾値 0.975, 独立モニタリング委員会の設置, 割り付け割合)は受け入れ可能でしょうか? 現時点で, さらに計画しておいた方がいいことは何が考えられますか?



# 結果とその解釈



# キーオープン後

- 試験は計画通りに終了しキーオープン後、プライマリ解析( $a_{WH} = 1$ で)次のような結果であった

	N	Event	割合	非劣性確率
薬剤Z	6000	163	2.72%	99.1%
ワーファリン	2000	48	2.43%	
事前分布(ワーファリン)	3665	93	2.54%	

- 事前に計画されていた通り非劣性を結論とした
- 比較のため、 $a_{WH} = 0$ とした場合の非劣性確率を計算すると93.9%であった
  - 参考：非劣性の事後確率が97.5%に近くなるような $a_{WH}$ は0.4付近であると観測された
- また頻度論においても有意ではなかった(差の95%信頼区間[-0.6%, 1.1%])



# ディスカッションポイント

- この試験の結果について、さらに知りたいこと・懸念点がありますか？

# 非劣性試験におけるベイズ統計学の メリット・デメリット

# 非劣性試験におけるベイズ統計学の メリット・デメリット

- メリット

- 結果の解釈が分かりやすい
- 今回の設定で、事前情報と本試験のイベント発現割合に乖離がない場合においては
  - 例数の効率的な配分が行うことができ、結果的に必要症例数減少につながる
  - 動作特性の上での $\alpha$ エラーは片側2.5%程度に収まる

- デメリット

- 本試験のエンドポイントの何らかの傾向が事前分布と異なる場合、プロトコル改定や事前分布の重み( $a_{wH}$ )に対する工夫をしなければ、誤った判断をしてしまう

# ディスカッションポイント

- 本日のディスカッションの結論として、メリット・デメリットを整理したいと思います。他にもメリット・デメリット等が考えられますか？
- 改めて、この開発において事前情報はどの程度入れることが適切だと考えますか？事前分布の利用割合の設定( $a_{wH}$ )はどのような方法で決められそうでしょうか？

ご参加いただきありがとうございました。

$$P(\textit{You are a Bayesian} | 5^{\textit{th}} \textit{ DSRT}) = ?$$

バックアップ

# ワーファリンの過去試験と非劣性 マージンの導出

# ハザード比のマージン

- FDAガイダンスに基づき、生存時間解析によって次の検定が行われることが本来期待される。

$$\text{帰無仮説 } H_0: \lambda_Z / \lambda_W \geq \Delta_{HR}$$

$$\Delta_{HR} = 1.378$$

# 割合のマージンの導出

- 想定されるワーファリンのイベント発現率(%/年)を $\lambda_w$ とおき, 18か月後のイベント発現確率に対応する非劣性マージンを次のように算出した. ( $\lambda_w$ の値は後のスライドで紹介)

$$\Delta = \exp(-\lambda_w \times 1.5) - \exp(-\Delta_{HR} \times \lambda_w \times 1.5)$$

- ここでは計算の都合上このマージンを用い, このマージンが臨床的にも受け入れられているものと仮定する.

# ワーファリンの過去の情報

- ワーファリンvsプラセボ(1989 - 1992)[1]

**Table 1. Placebo-Controlled Trials of Warfarin in Nonvalvular Atrial Fibrillation**

Study	Summary	Events/Patient Years		Risk Ratio (95% CI)
		Warfarin	Placebo	
AFASAK	open label. 1.2 yr follow-up	9/413 = 2.18%	21/398 = 5.28%	0.41 (0.19, 0.89)
BAATAF	open label. 2.2 yr follow-up	3/487 = 0.62%	13/435 = 2.99%	0.21 (0.06, 0.72)
EAFT	open label. 2.3 yr follow-up patients with recent TIA	21/507 = 4.14%	54/405 = 13.3%	0.31 (0.19, 0.51)
CAFA*	double blind. 1.3 yr follow-up	7/237 = 2.95%	11/241 = 4.56%	0.65 (0.26, 1.64)
SPAF I	open label. 1.3 yr follow-up	8/260 = 3.08%	20/244 = 8.20%	0.38 (0.17, 0.84)
SPINAF	double blind. 1.7 yr follow-up	9/489 = 1.84%	24/483 = 4.97%	0.37 (0.17, 0.79)

\* CAFA was stopped early because of favorable results observed in other studies.

メモ: ガイダンス記載のマージンはここから算出されている

# 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%
<b>合計</b>	<b>3665</b>	<b>93</b>	<b>1.6</b>	<b>2.54%</b>

- 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確率(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確率の事前情報として使  
いる

# 割合スケールのマージンの計算

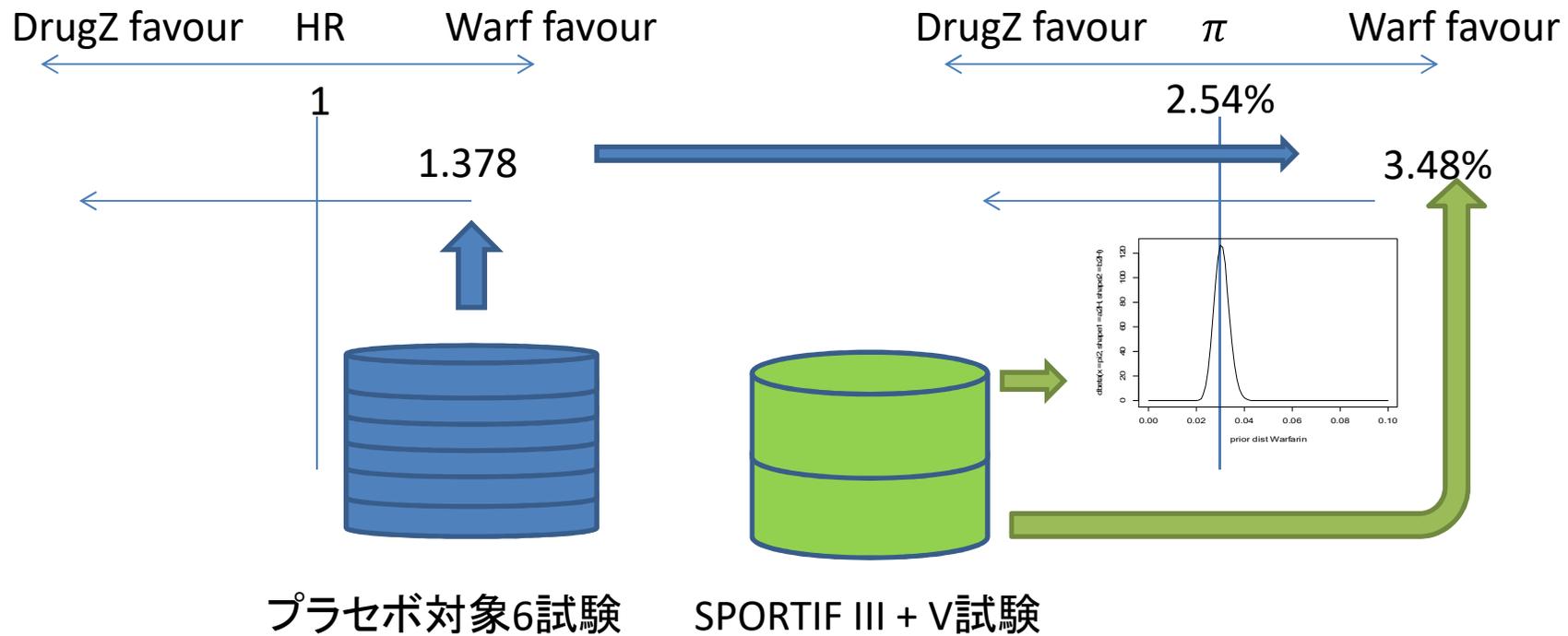
- 過去のデータに基づき、以下のように発現レートを算出し、割合スケールにおけるマージンを計算した。

$$\lambda_w = -\frac{\log\left(1 - \frac{X_{wH}}{n_{wH}}\right)}{1.5} = 0.17135$$

$$\begin{aligned}\Delta &= \exp(-\lambda_w \times 1.5) - \exp(-\Delta_{HR} \times \lambda_w \times 1.5) \\ &= 0.00942\end{aligned}$$

- 従って、割合スケールの非劣性マージンは0.94%を用いる。

# ワーファリン関連情報の整理



# 過去試験の背景因子

# SPORTIF III [3]

	<b>Ximelagatran (n=1704)</b>	<b>Warfarin (n=1703)</b>
<b>Characteristic</b>		
Men	1158 (68%)	1196 (70%)
Age (years, mean [SD])	70.3 (8.6)	70.1 (8.6)
Bodyweight (kg, mean [SD])	80.7 (16.8)	81.7 (16.9)
Ethnic origin		
White	1494 (88%)	1500 (88%)
Asian	201 (12%)	196 (12%)
Other	9 (1%)	7 (1%)
Aspirin at entry	345 (20%)	359 (21%)
Vitamin K antagonist at entry	1267 (74%)	1235 (73%)
Systolic blood pressure (mm Hg, mean [SD])	139 (18)	139 (18)
Atrial fibrillation onset less than 1 year	368 (22%)	347 (20%)
Paroxysmal atrial fibrillation	160 (9%)	124 (7%)
Risk factors		
One*	511 (30%)	545 (32%)
Two	614 (36%)	579 (34%)
Three or more	579 (34%)	579 (34%)
Previous stroke, TIA, or both	417 (24%)	405 (24%)
Previous non-CNS embolism	74 (4%)	77 (5%)
Age 75 years or older	581 (34%)	565 (33%)
Hypertension	1229 (72%)	1230 (72%)
Left ventricular dysfunction	574 (34%)	584 (34%)
Age 65 years or older and coronary artery disease	581 (34%)	558 (33%)
Age 65 years or older and diabetes mellitus	288 (17%)	290 (17%)

Data are number of patients (%) unless otherwise stated. TIA=transient ischaemic attack. \*Includes five patients in each group incorrectly randomised without risk factors.

# SPORTIF V [4]

**Table 1.** Characteristics of Randomized Patients According to Randomized Treatment Assignment\*

	Ximelagatran (n = 1960)	Warfarin (n = 1962)
Men, No.(%)	1365 (70)	1353 (69)
Age, mean (SD), y	71.6 (9.2)	71.6 (9.0)
Body weight, mean (SD), kg	90.1 (21.9)	89.1 (21.3)
Body mass index, mean (SD)†	30.0 (6.6)	29.6 (6.2)
Race, No. (%)		
White	1875 (95.7)	1888 (96.2)
Black	67 (3.4)	58 (3.0)
Asian	15 (0.8)	10 (0.5)
Aspirin at entry, No. (%)	352 (18)	367 (19)
VKA at entry, No. (%)	1617 (83)	1661 (85)
Systolic BP, mean (SD), mm Hg	133 (18)	132 (18)
Atrial fibrillation onset <1 y, No. (%)	311 (16)	304 (15)
Paroxysmal atrial fibrillation, No. (%)	282 (14)	270 (14)
Risk factors, No. (%) ‡		
1	490 (25)	509 (26)
2	600 (31)	597 (30)
≥3	867 (44)	852 (44)
Previous stroke and/or TIA, No. (%)	369 (19)	348 (18)
Previous non-CNS embolism, No. (%)	92 (4.7)	85 (4.3)
Age ≥75 y, No. (%)	838 (43)	820 (42)
Hypertension, No. (%)	1584 (81)	1582 (81)
CHF/LV dysfunction, No. (%)	735 (38)	788 (40)
Age ≥65 y + CAD, No. (%)	822 (42)	803 (41)
Age ≥65 y + diabetes mellitus, No. (%)	389 (20)	373 (19)

Abbreviations: BP, blood pressure; CAD, coronary artery disease; CHF, congestive heart failure; CNS, central nervous system; LV, left ventricular; TIA, transient ischemic attack; VKA, vitamin K antagonist.

\*The treatment groups did not differ with respect to any of the listed characteristics.

†Body mass index is calculated as weight in kilograms divided by the square of height in meters.

‡Risk factors are listed under "Patients" in the "Methods" section.

# POCOCKの条件

## Pocockの6条件 [11]

1. 詳細に定義された, 新規試験の対照治療群と同じ治療を受けている
2. 新規試験と同じ適格基準を用いた最近実施された臨床試験である
3. 治療の評価方法が新規試験と同じである
4. 重要な患者背景の分布が新規試験と同様である
5. 新規試験とほぼ同じ組織で実施されている
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. 例えば, 新規試験において症例登録スピードが予想よりも上がらないとき, 適格基準をみたすぎりぎりの症例が登録されることも起こり得る. そういった場合, 新規試験と既存試験で共通の適格基準を用いたとしても登録患者の廃棄が二つの試験間で異なってくる可能性が~~ある~~本語訳は武田ら(2015)より

# **ADAPTIVE POWER PRIORS**

# Adaptive power priors with empirical Bayes [12]

## 2.3 | Empirical Bayes

An alternative to the fully Bayesian methodology is that of EB, which estimates hyper-parameters based on the data. Here, we choose the weight parameter with an EB approach, maximising the marginal likelihood of  $\delta$ ,

$$\hat{\delta}(x_0, x_\star) = \arg \max_{\delta \in [0,1]} L(\delta; x_0, x_\star), \quad (8)$$

where

$$\begin{aligned} L(\delta; x_0, x_\star) &= \int L(\theta; x_\star) p(\theta | \delta, x_0) d\theta \\ &= \frac{\int L(\theta; x_\star) L(\theta; x_0)^\delta p(\theta) d\theta}{\int L(\theta; x_0)^\delta p(\theta) d\theta}. \end{aligned} \quad (9)$$

Note that  $p(\theta | \delta, x_0)$  is properly normalised (as in Equation 4) and that Equation 9 does not have a prior on  $\delta$ . Using the EB-type methodology, we define a new power prior based on the conditional power prior (Equation 1) with the estimate (Equation 8) of  $\delta$  inserted, which we will refer to as the EB power prior. The EB estimate eliminates the need for a prior on  $\delta$ , and by using the marginal likelihood based on  $x_0$  and  $x_\star$ , it adapts to the current and historical data. We still require a prior on  $\theta$  to complete the specification. The EB power prior is still a posterior distribution based on the historical data, but now conditional on  $\hat{\delta} = \hat{\delta}(x_0, x_\star)$ . In the normal case, the EB estimate of  $\delta$  is available in closed form, see Appendix A.

# 参考文献

# 参考文献

1. Food and Drug Administration. 2016. *Non-Inferiority Clinical Trials to Establish Effectiveness Guidance for Industry*. USA: Food and Drug Administration
2. Food and Drug Administration. 2010 , *Guidance for Industry and FDA Staff : Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials*. USA: Food and Drug Administration
3. Olsson, S.B., Executive Steering Committee of the SPORTIF III Investigators. 2003. Stroke prevention with the oral direct thrombin inhibitor ximelagatran compared with warfarin in patients with non-valvular atrial fibrillation (SPORTIF III): randomised controlled trial. *The Lancet*. Nov 22;362(9397):1691-8.
4. Albers, G.W., et al. 2005. Ximelagatran vs warfarin for stroke prevention in patients with nonvalvular atrial fibrillation: a randomized trial. *The Journal of the American Medical Association*. Feb 9;293(6):690-8.

# 参考文献

5. Connolly, S.J., *et al.* 2009 . Dabigatran versus Warfarin in Patients with Atrial Fibrillation. *The New England Journal of Medicine*. September 17, 361:1139-1151
6. Granger, C.B., *et al.* 2011. Apixaban versus Warfarin in Patients with Atrial Fibrillation. *The New England Journal of Medicine*. September 15, 365:981-992
7. Patel, M.R., *et al.* 2011. Rivaroxaban versus Warfarin in Nonvalvular Atrial Fibrillation, *The New England Journal of Medicine*. September 8, 365:883-891
8. Giugliano, R.P. *et al.* 2013. Edoxaban versus warfarin in patients with atrial fibrillation. *The New England Journal of Medicine*. November 28;369(22):2093-104

# 参考文献

9. Doi, M., Takahashi, F., Kawasaki, Y. 2017. Bayesian noninferiority test for 2 binomial probabilities as the extension of Fisher exact test. *Statistics in Medicine*. **36**(30), pp. 4789–4803
10. Morita, S., Thall, P. F., Muller, P. 2008. Determining the effective sample size of a parametric prior. *Biometrics*, **64**(2), pp.595–602,
11. Pocock, S. J. 1976. The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases*, **29**(3), pp.175–188.
12. Isaac G., Leonhard H., On behalf of the COMBACTE-Net consortium. 2017. Adaptive power priors with empirical Bayes for clinical trials. *Pharmaceutical Statistics*. **16**(5) 349–360.