

第4回 DSRT
テーマ4: ベイズ統計学の医薬品の
臨床開発での活用について
【議論のまとめ】
テーブル2：午後

ベイズチームの議題

- 議題1：ベイズ統計学の導入とFDAの医療機器ガイダンスから学べること
- 議題2：p値についての議論
- 議題3：ベイズ統計学の検証試験における意思決定での利用について
- **議題4：事前分布の構成・ヒストリカルデータの利用**

本発表の対象

議題4：事前分布の構成・ヒストリカルデータの利用

ディスカッションの目標

- Assuranceという概念を概観し，その利用可能性を検討する
 - 計画中の臨床試験の成功の確信度としての利用可能性
 - 予備試験の例数設計の参考指標としての利用可能性(時間があれば)
- Assuranceの議論を通して事前分布の選択方法を検討する
 - 複数の試験や事前の知識をどう扱う方法とその解釈
- 注意点
 - 「試験の成功」の定義は議論になり得るが，ここでは主要な目的が達成されることを意図するものとする
 - 主要な解析にはベイズ統計は用いず，頻度論を用いる
 - 計画段階の試験成功の見積もりにベイズ統計の考え方をを用い，その際に**事前分布の選択の検討**を目的とする

議題4：事前分布の構成・ヒストリカルデータの利用

議題4：事前分布の構成・ヒストリカルデータの利用

[議論の導入]

Agendaとタイムラインの目安

- 架空の薬剤の開発背景とクロスオーバー試験の基礎(15分)
- Assuranceの導入(20分)
- Assuranceの架空の試験Aへの適用例(30分)
- 分散(ばらつき)に分布を仮定する(15分)
- 医薬専門家(臨床家, PK専門家委)との考察(20分)
- ベイズを使うその他の利用方法の検討(15分) [オプション]
- 予備試験の例数設計(15分) [オプション]
- まとめ(15分)

[オプション]のついているテーマは時間が足りなければ割愛させていただきます。

架空の薬剤の開発背景とクロスオーバー試験の基礎

本題に入る前に

用語および背景情報

議論の設定：これから架空の試験「計画中の試験A」の計画を考えます

- 薬剤
 - 対照薬と被検薬X
- 計画中の臨床試験
 - 申請を目的とした対照薬と被検薬Xの同等性試験(日本人対象の2群2剤4期クロスオーバー試験)
- エンドポイント
 - この議論ではAUCのみを対象とする(本来はCmaxを含む)
- 過去試験情報
 - **予備試験**：日本人合計20人を対象とした，対照薬vs被検薬Xの2群2剤2期クロスオーバー試験
- 予算的制限
 - 計画中の試験は80例までとしたい

このセッションで例数はすべて**合計例数**を表す

クロスオーバー試験のデザイン

- 予備試験のデザイン

- 2群2剤2期クロスオーバー(2x2x2)

	時期1	時期2
群1	対照薬	被検薬X
群2	被検薬X	対照薬

- 計画中の試験Aのデザイン

- 2群2剤4期クロスオーバー(2x2x4)

	時期1	時期2	時期3	時期4
群1	対照薬	被検薬X	対照薬	被検薬X
群2	被検薬X	対照薬	被検薬X	対照薬

クロスオーバー試験の解析例

- i 番目の個人の時期 j のAUCの対数の値を Y_{ij} とおく

本来は時期や群の固定効果も考慮できるが、ここでは最も簡単なモデル例を紹介した。

$$Y_{ij} = \mu + \delta \times_{ij} + b_i + \varepsilon_{ij}$$
$$\times_{ij} = \begin{cases} 1 & (\text{drug } X) \\ 0 & (\text{reference}) \end{cases}$$
$$\varepsilon_{ij} \sim N(0, \sigma_w^2) \text{ iid}, b_i \sim N(0, \sigma_b^2) \text{ iid}$$

ε_{ij} and b_i are mutually independent

- ここで、 $\exp(\delta)$ をgmean ratio(幾何平均の比geometric mean)とよび、その95%CIが0.8-1.25の間に入れば同等と判断する
- 帰無仮説は $H_0: \exp(\delta) < 0.8$ or $\exp(\delta) > 1.25$
- 個体間分散 σ_b^2 と個体内分散 σ_w^2 があり、ばらつきの指標として個体内 $\text{gCV} = \sqrt{\sigma_w^2 - 1}$ が用いられる

クロスオーバー試験の検出力

- 症例数設計・検出力の算出についてここでは詳細は議論しない
- 重要なポイント
 - 試験に必要な合計例数及び検出力は、想定される真値の gmean ratio と gCV で計算される
 - 4期の必要例数は2期の必要例数のおおよそ 1/2 の例数でよくなる

詳細は、参考文献を参照[Chen, et al. (1997)]

Assuranceの導入

試験成功の確信度の見積もり

Assurance (試験成功の確信度)

- 臨床試験の計画段階で利用される検出力は真値の予想値をもとに算出されている
- 計画段階で想定される真値の予想値の確からしさは、臨床試験毎に異なる
- ベイズ統計を用いることによって、その確からしさを考慮した試験の成功確率(Assurance)を算出する
- ただし、事前情報の信頼性が低い場合は、真値が実臨床で起こりえないような値をとる確率も加味してしまうため、成功確率が低く見積もられ、試験計画やGo/NoGo等の判断に影響を与えてしまう可能性がある

検出力=成功確率?

- 検出力は成功率として考えられるか?
 - 検出力80%の臨床試験をたくさん集めると、有意差のついた試験は8割となるだろうか？逆に80%の臨床試験を計画しているが、実際に成功する確率は(内心は)高いと感じていることはないか？
 - 例えば、USでのP3からNDAに進んだ成功確率は**58.1%**(2006-2015)であった。詳細はレポート参照[Bio, Biomedtracker, ADDPLAN]

検出力=成功確率?

- 「検出力=成功確率」となる条件
 - 検出力が真値の予想値が真値と一致していること.
- 検出力と成功確率が(大きく?)異なる理由
 - 事前の知見や過去の臨床試験の結果に基づく予想値の不確かさ
 - 不確かさを考慮するために、検出力算出に用いる真値の予想値が異なった場合を想定して検出力を算出してみることも考えられる.

事前情報の不確かさを 確率分布で表現する

- 検出力見積もりに用いる事前情報の不確かさを考慮する方法として、ベイズ統計を用いることもできる。
- ベイズを用いることによって真の群間差および、真のSDの事前情報の不確かさを確率分布で表現することができる。

注意：これ以降、「真値の分布」「真値がばらつく」類の表現が頻出します。

ベイズを使う上で重要なポイント

- 想定する (実際には真の)群間差やばらつき(SD)が **確率分布を持つことを許す**.
- 頻度論の観点では、真値を少しずつ変更して複数の検出力を算出し、『総合的に判断』することも考えられるが、ベイズ統計でも確率分布を用いて同様な考察を行うことができる。
- 真の群間差やSDに確率分布を与えたので、それらを用いて **算出される検出力も確率分布を持つ**. **その平均をAssurance**と呼ぶ。 [O'Hagan A.,et al. (2005)]

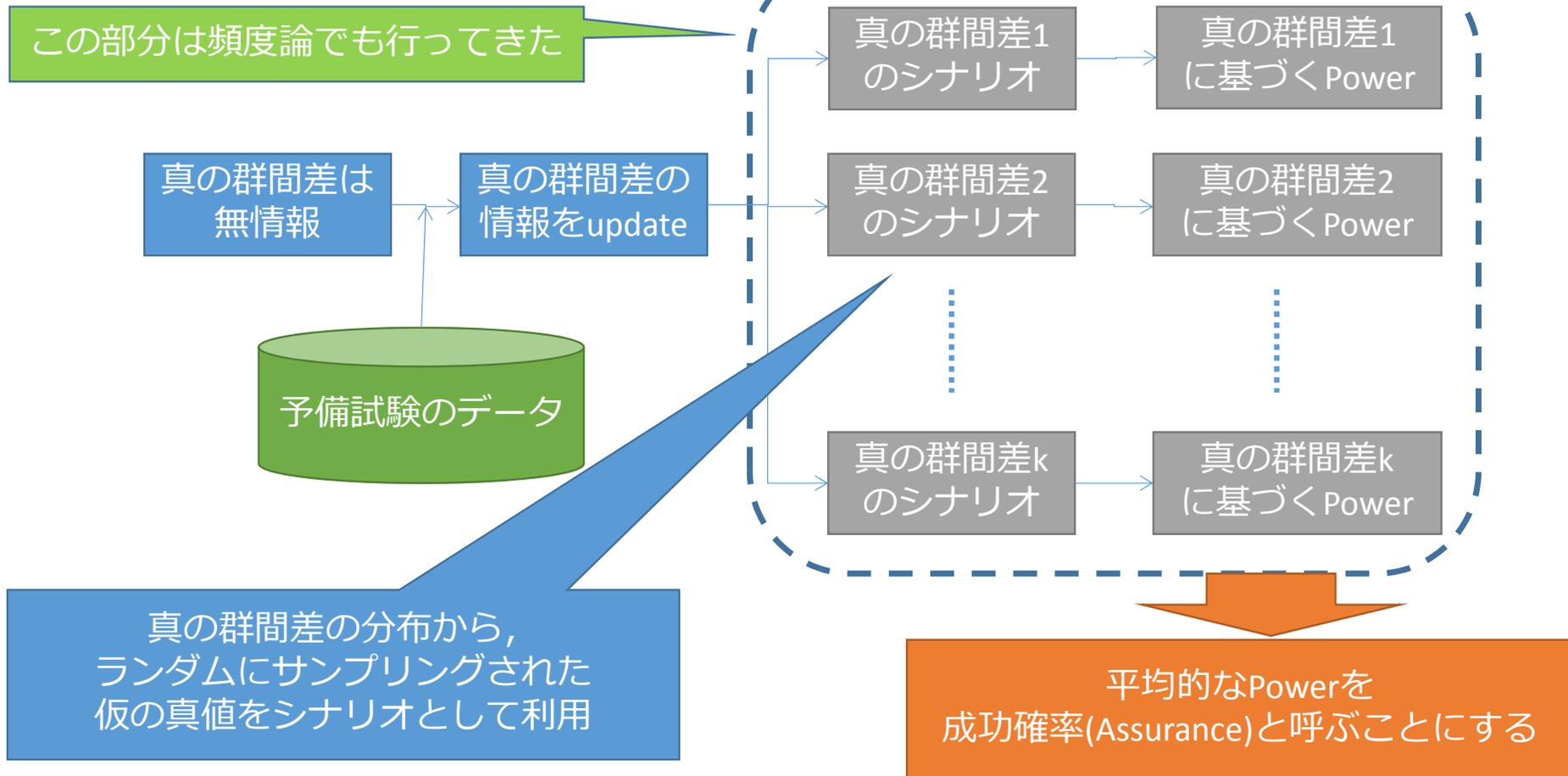
Assuranceの定義

- 試験の成功確率(Assurance)は、群間差などのパラメーターの分布が与えられた下で、次のように定義される(ここではパラメーターを δ と σ_w^2 とおいた例を示す)

$$Assurance = \int_0^{\infty} \int_{-\infty}^{\infty} P(\text{significant result} | \delta, \sigma_w^2) f(\delta, \sigma_w^2 | \text{data}) d\delta d\sigma_w^2$$

- 積分計算が解析的に難しい場合は δ と σ_w^2 の(同時)分布に基づいて、モンテカルロ積分などを行う。

Assurance算出の例



【Discussion Pointと議論の結果1】

- Assuranceの定義は理論的に受け入れ可能でしょうか？
 - 平均でよいか？ほかの指標を用いるのはどうか？
 - 考え方はOK, ただ, power算出のプロセスはすべて残したほうがよい. 例えば, 各シナリオでの δ の値と, そこでの検出力.
- Assuranceは試験の成功確率と解釈してもよいでしょうか？どのようにすれば成功確率をよりよく見積もれるでしょうか？
 - Powerの平均値を算出しているので, 成功確率に必ず一致するわけではない.
- 検出力90%の臨床試験を設定する代わりに, Assuranceが〇〇%の臨床試験を計画し, プロトコルに記載することは適切でしょうか？

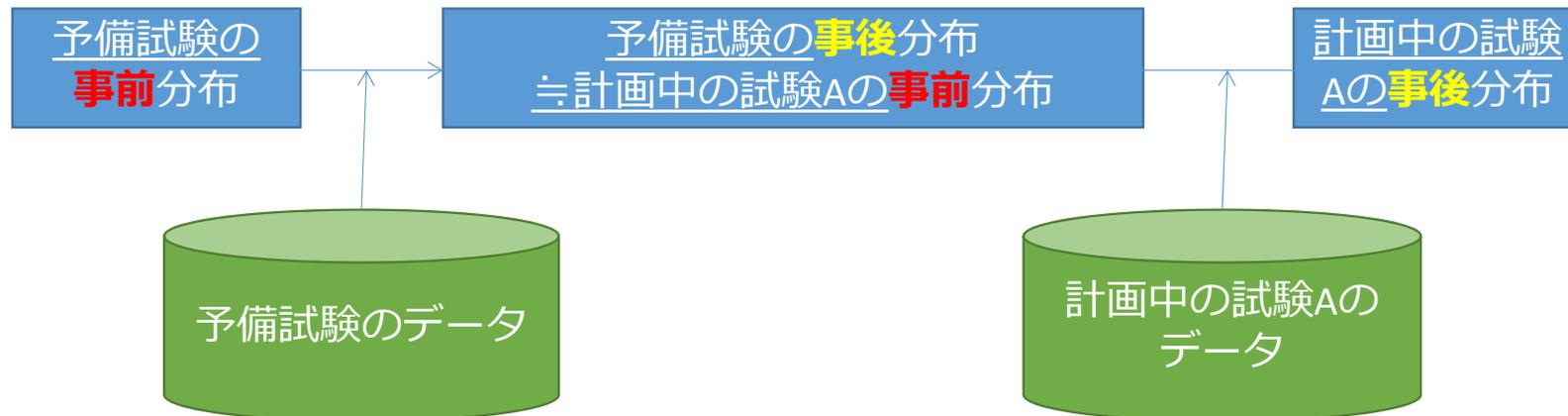
Assuranceの架空の試験Aへの 適用例

Bioequivalence studyへの応用

用語の注意点

- 2つ以上の臨床試験がある場合に、事前分布と事後分布は、注目している試験によって呼ばれ方が変わります。

事前情報と呼べるものは、計画中の試験Aの事前分布であり、予備試験の事後分布の意味で使われることがある。



(リマインド用)用語および背景情報

議論の設定：これから架空の試験「計画中の試験A」の計画を考えます

- 薬剤
 - 対照薬と被検薬X
- 計画中の臨床試験
 - 申請を目的とした対照薬と被検薬Xの同等性試験(日本人対象の2群2剤4期クロスオーバー試験)
- エンドポイント
 - この議論ではAUCのみを対象とする(本来はCmaxを含む)
- 過去試験情報
 - **予備試験**：日本人合計20人を対象とした，対照薬vs被検薬Xの2群2剤2期クロスオーバー試験
- 予算的制限
 - 計画中の試験は80例までとしたい

このセッションで例数はすべて**合計例数**を表す

予備試験の結果

- 予備試験として2x2x2クロスオーバーデザインの試験を合計20例で実施し, 結果は以下のように得られた.
 - Geometric mean ratio=**0.9**
 - 個体内Geometric CV=**0.3**(CVは各薬剤で共通とする)
- この結果から次の2群2剤4期のクロスオーバー試験Aの成功率を40例, 60例, 80例で検討したい. (通常の検出力は順に81%, 93%, 97%)

データ発生・解析モデル

解析時に使用するモデルは以下を使用する
症例*i*番の*j*番目のAUCの対数を

$$Y_{ij} = \mu + \delta x_{ij} + b_i + \varepsilon_{ij}$$

と表す。ここで

$$x_{ij} = \begin{cases} 1 & (\text{active}) \\ 0 & (\text{reference}) \end{cases}$$

$$\varepsilon_{ij} \sim N(0, \sigma_w^2) \text{ iid}, b_i \sim N(0, \sigma_b^2) \text{ iid}$$

ε_{ij} and b_i are mutually independent

$$i = 1, \dots, N, j = 1, 2$$

とする。ここまでは頻度論でも同じ

真値に分布を持たせる

- 予備試験の結果(群間差の平均)の分布

$$\bar{d}|\sigma_w^2 \sim N(\delta, \sigma_w^2/(N/2))$$

- 予備試験の事前分布に無情報事前分布 $f(\delta) \propto 1$ を用いた場合の、予備試験後の真値の分布

– 事後分布は尤度にのみ依存するようにした。この分布は「無情報事前分布」と呼ばれることがあるが、1. 確率分布ではない、2. 本当に「無情報」であるかどうかは議論の余地がある、等注意が必要

- ここではまず、 σ_w^2 は定数として取り扱う

$$\delta|\sigma_w^2 \sim N(\bar{d}, \sigma_w^2/(N/2))$$

$$f(\delta|\bar{d}, \sigma_w^2) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left[-\frac{(\delta - \bar{d})^2}{2\sigma_w^2/(N/2)}\right]$$

数式での表現

- 試験の成功確率(Assurance)

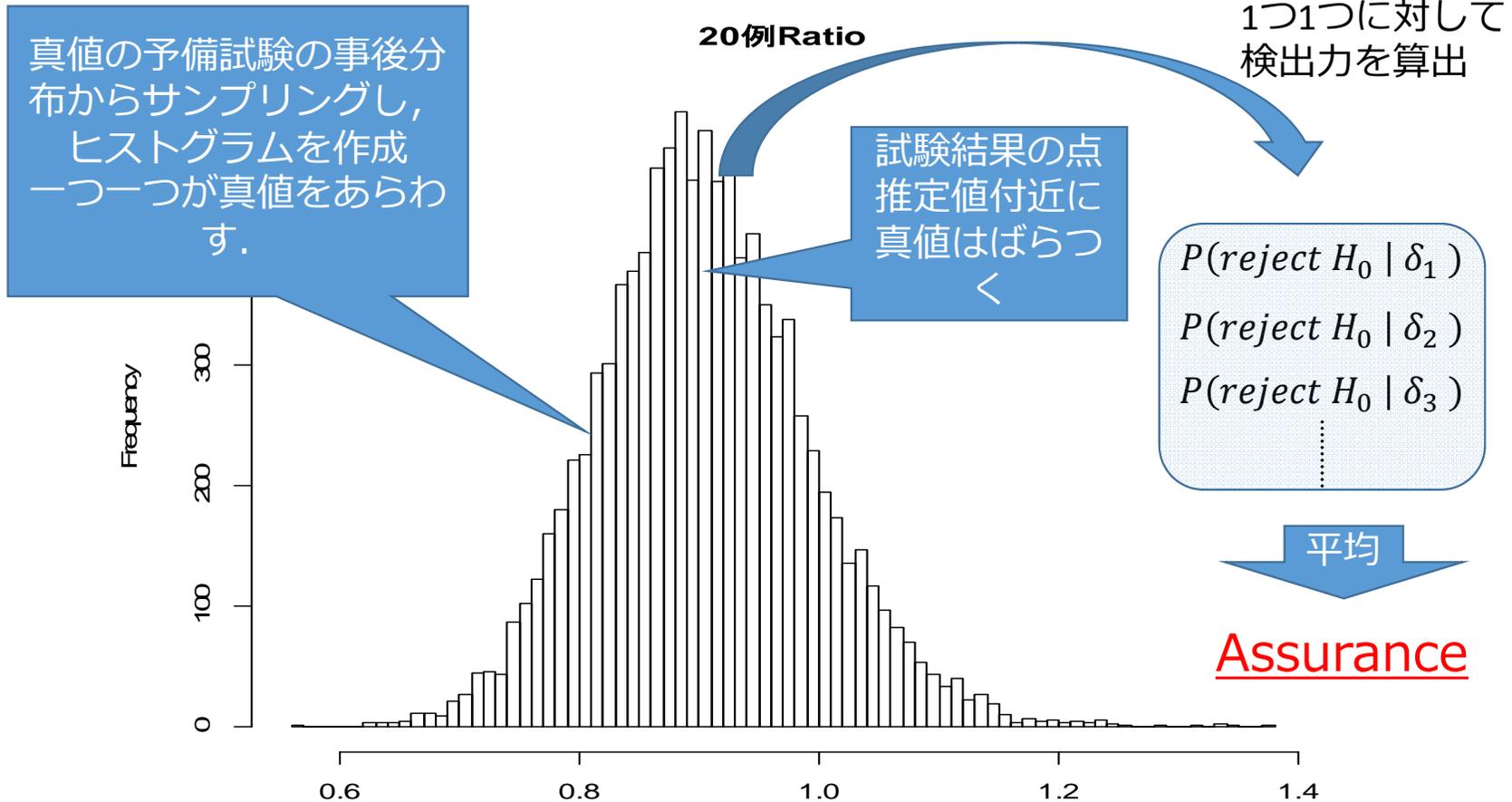
$$Assurance = \int_{-\infty}^{\infty} P(\text{significant result}|\delta) f(\delta|\bar{d}, \sigma_w^2) d\delta$$

- 積分計算が難しい場合は δ の分布に基づいて、モンテカルロ積分などを行う。
- ここでは分散 σ_w^2 (gCV)を固定していることに注意

アルゴリズム(モンテカルロ法使用)

- 以下を $i = 1, \dots, K$ まで loop する
 - $\delta_{(i)}$ を $N(\bar{d}, \sigma_w^2 / (N/2))$ からサンプリング
 - N, \bar{d}, σ_w^2 には予備試験の合計例数および対数スケールで推定された平均, 分散を用いた
 - $\delta_{(i)}$ に基づき検出力 $P(\text{reject} \mid \delta_{(i)})$ を算出する
- 最後に得られた検出力の平均を算出し, Assurance の近似とする

真値の分布：gmean ratio



例数 x Power x Assurance

合計 例数	群間差の真値が既知の場合の power (gmean = 0.85, 0.9, 0.95の順)			Assurance with 既知CV
40	36	81	98	65
60	48	93	>99	71
80	58	97	>99	74

Assuranceには上限がある?

- 予備試験の例数を固定(合計20例)し、計画中の試験Aの例数を ∞ に近づけると、0.9付近になる。(合計例数 $n=10^9$ でシミュレーション)
- これはgmean ratioの事前分布が0.8-1.25の間に入る確率(0.885)に近い。
- 通常、検出力は例数を ∞ に近づけることで1に収束するが、この意味でもAssuranceと検出力は異なる概念と言える。
- 並行群間試験の優越性試験のシミュレーションでも、同様なことが起きることが確認されている。Assuranceの上限は $1 - (\text{予備試験のp値})$ ([Carroll, K.J., 2013]参照)

【Discussion Pointと議論の結果2】

- 頻度論の検出力と、Assuranceの数値は単純に比較してよいでしょうか？理由とともに考えてください。
 - よい：4名
 - よくない：3名
 - (よくない理由) powerは1つの δ に基づく値だが、assuranceは複数の δ に対する要約。比較の土台に乗ってないのでは？
- 追加で計算したいものはありますか？
- 頻度論の方法と Assurance のどちらが「より妥当」でしょうか？
- 以上を踏まえて、よりよい方法を考えてください
 - 頻度論の方法を「ベイズ流の方法で完全に代替する」のではなく、「新たな情報を追加する」という観点の方が適切かもしれない
 - Assurance の数値1つだけでなく、途中の経過を残すことは重要
 - 横軸power, 縦軸頻度としたグラフを作ることは重要かもしれない

【Discussion Pointと議論の結果2】

- 試験Aの例数が増えてもAssuranceが100%にならないことをどのように解釈しますか？
 - δ の分布が、帰無仮説での値も含んでいるからではないか？
 - δ が帰無仮説に含まれる場合、症例数を増やせば「帰無仮説が棄却されない確率」が上がる
 - では事前分布から、帰無仮説の範囲を除外することは妥当であろうか？
 - 帰無仮説の範囲を除いてしまった場合、不確実性を減らしてしまう可能性はないか？事前分布が帰無仮説を含む場合、その可能性も考慮すべきか？
 - そもそも効果ないのであれば試験するべきではない。頻度論の検出力の計算は、常に「効果がある」状況を想定して算出する。そのため、assuranceの計算の際も、同じように「対立仮説が正しい」という仮定をおいてもよいのではないか？
 - [NOTE]事前分布に一様分布の範囲を与えることは、本資料の後半に記載あり

【Discussion Pointと議論の結果2】

- その他の論点
 - Assurance の単位は%か？何を表しているか？
 - 無単位. 複数ある検出力の重みつき平均
 - CVが小さいと assuranceは大きくなるから, CVのコントロールも忘れないようにする
 - CVの値は「1シナリオごとの検出力」「事前分布のばらつき」の両方に影響を与える

分散(ばらつき)に分布を仮定する

分散は未知？ 既知？

分散は未知？ 既知？

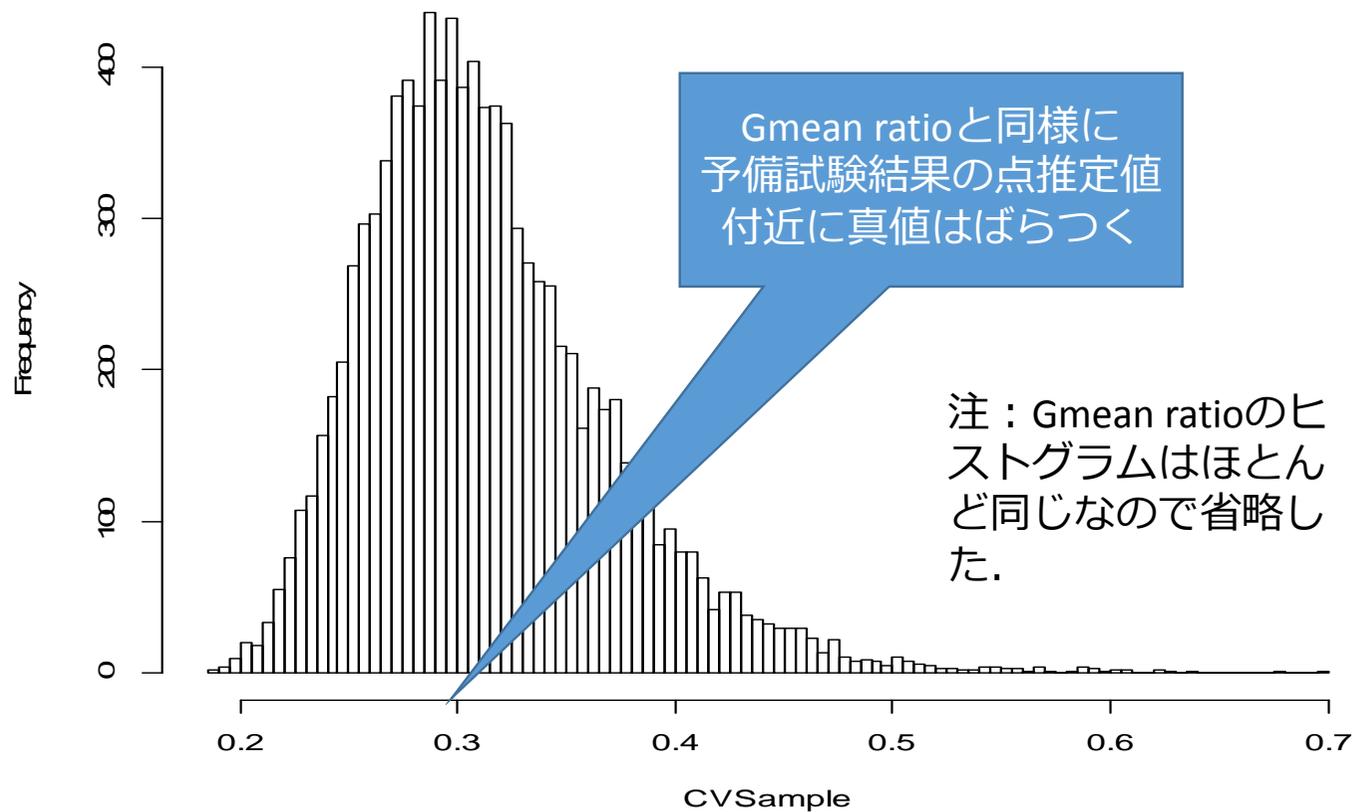
- ばらつきが**既知**，**未知**である話は，統計の初等的な話でもよく出てきたように(正規分布かt分布か等)ここでも話題に挙げられる。
- 実際にearly phaseなどで得られる分散の推定値は，通常（少なくともいくらかは）真値と異なる
- ここでは，さらに分散がばらつくことを考慮して，検討してみる。

アルゴリズム(モンテカルロ法使用)

- 以下を $i = 1, \dots, K$ まで loop する
 - $\sigma_w^2(i)$ を $\chi^{-2}(N - 1, s^2)$ からサンプリング
 - N, s^2 には予備試験の合計例数および対数スケールで推定された分散を用いた
 - $\delta_{(i)} | \sigma_w^2(i)$ を $N(\bar{d}, \sigma_w^2(i)/(N/2))$ からサンプリング
 - N, \bar{d} には予備試験の合計例数および対数スケールで推定された平均を用いた
 - $(\delta_{(i)}, \sigma_w^2(i))$ に基づき検出力 $P(\text{reject} | \delta_{(i)}, \sigma_w^2(i))$ を算出する
- 最後に得られた検出力の平均を算出し, Assurance の近似とする

真値の分布 gCV

20例CV



例数 x Power x Assurance

合計 例数	Power	Assurance with 既知CV	Assurance with 未知CV
40	81	65	63
60	93	71	70
80	97	74	73

ばらつきの既知・未知はAssuranceにあまり影響を与えなかった。
(この試験では)

医薬専門家(臨床家, PK専門家) との考察

統計家以外の視点から見た場合

(リマインド用)用語および背景情報

議論の設定：これから架空の試験「計画中の試験A」の計画を考えます

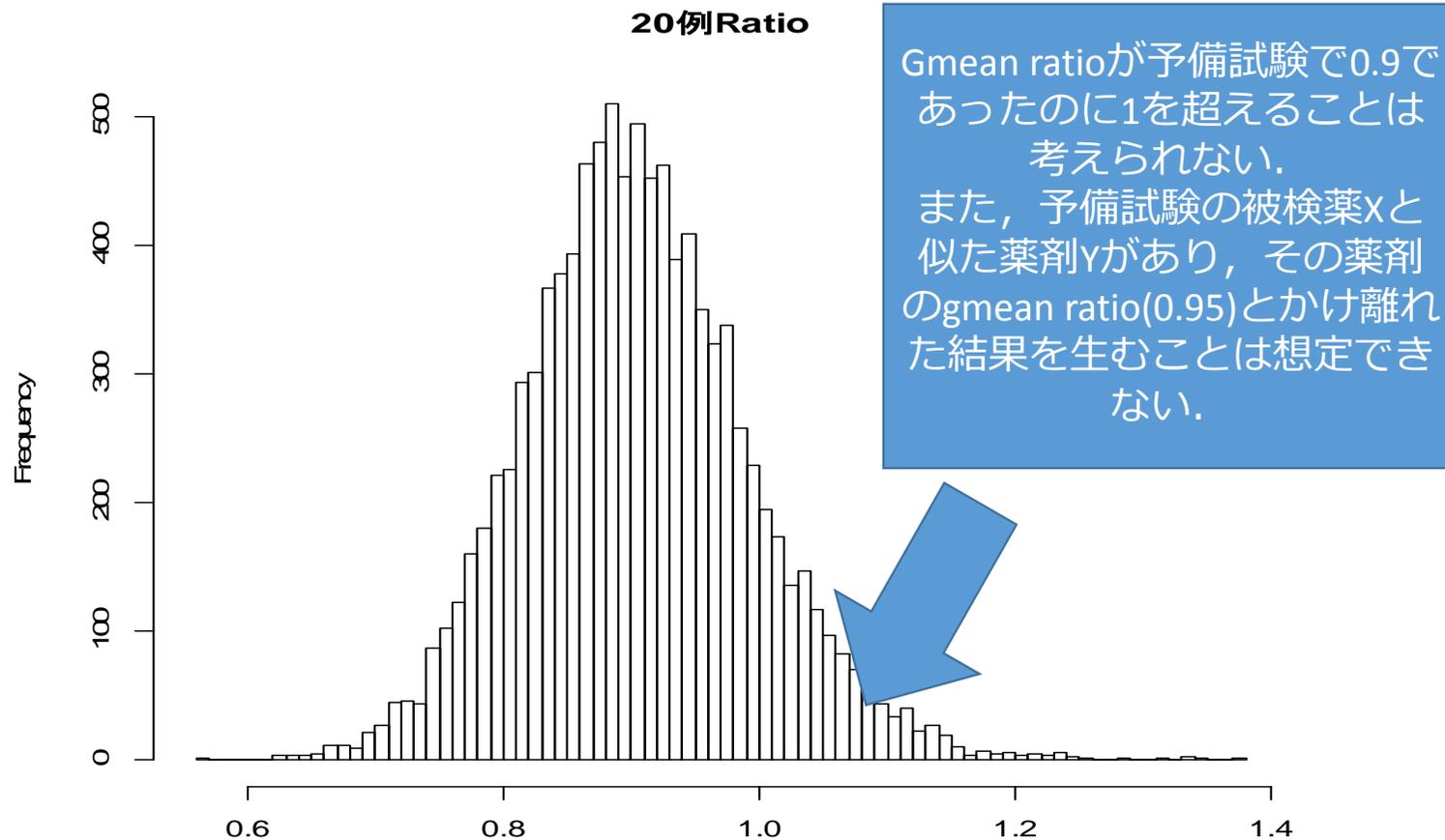
- 薬剤
 - 対照薬と被検薬X
- 計画中の臨床試験
 - 申請を目的とした対照薬と被検薬Xの同等性試験(日本人対象の2群2剤4期クロスオーバー試験)
- エンドポイント
 - この議論ではAUCのみを対象とする(本来はCmaxを含む)
- 過去試験情報
 - **予備試験**：日本人合計20人を対象とした，対照薬vs被検薬Xの2群2剤2期クロスオーバー試験
- 予算的制限
 - 計画中の試験は80例までとしたい

このセッションで例数はすべて**合計例数**を表す

医薬専門家との考察

- 医薬専門家とAssurance・Powerおよび予備試験の事後分布のグラフを用いて、予備試験の評価と開発計画について、次のようなコメントがあったと想定する
 - 60%などAssuranceが低すぎて + 予備試験のgmeanの事後分布のばらつきが大きすぎてhopelessな薬剤に見える
 - そもそも、gmean事後分布のヒストグラムに経験上に起こり得ない値が含まれることに違和感がある
 - 海外試験では対照薬のgCVは0.4程度だと分かっていたが、予備試験では対照薬のgCVが小さく推定されている
 - 80%のAssuranceを持つ例数で設計するとどうなるのか？ その場合、プロトコルにはどう記載するか？

Gmeanの事前分布の考察



予備試験のみ用いたベイズ統計の問題点

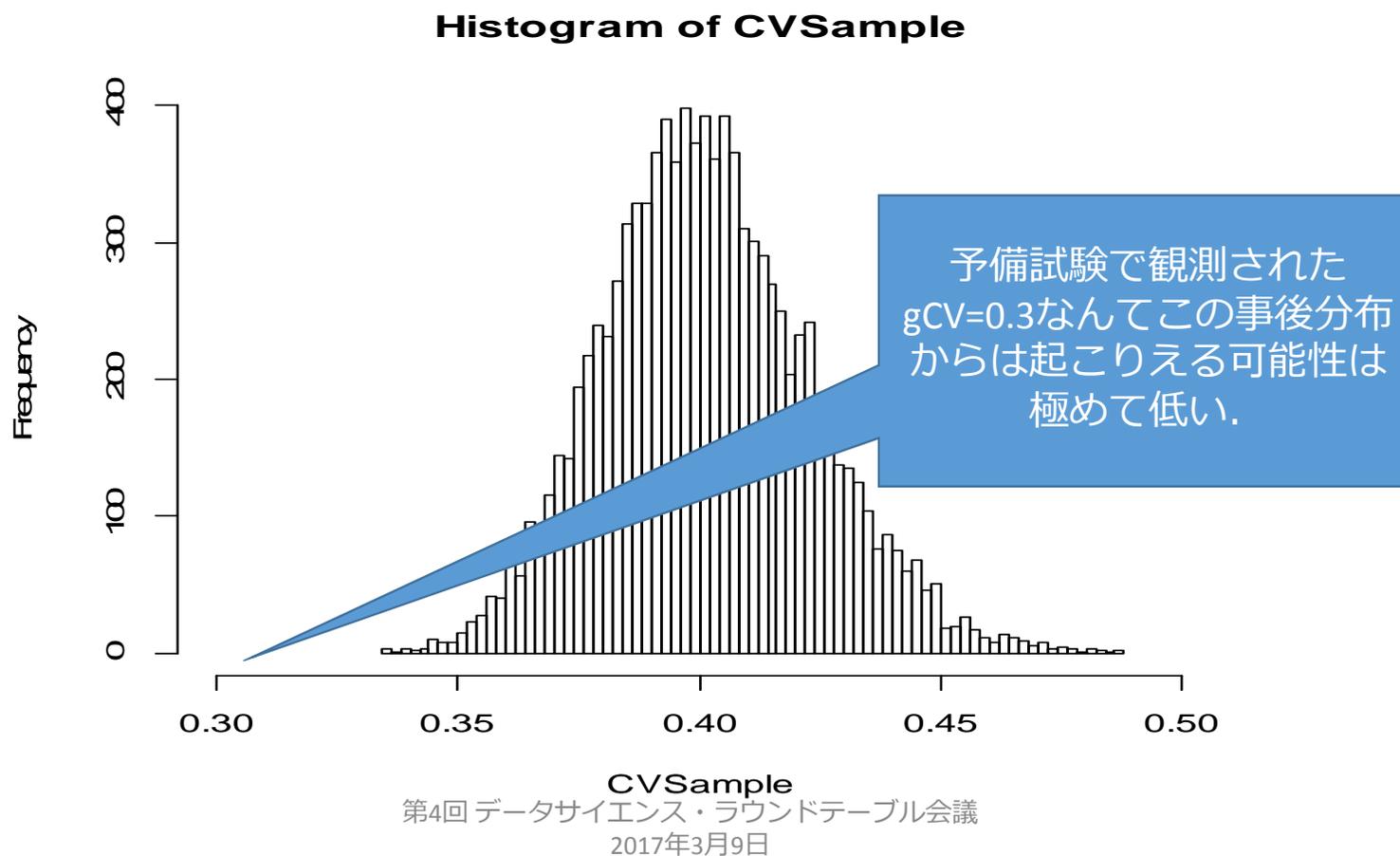
予備試験の事前分布に無情報を用いた場合,

- 例数が多くない予備試験の事後分布(本試験計画段階にとっては事前分布)では, 経験的に起こり得ない真値の確率を加味してしまう可能性がある。(分布の両裾が長い)
- 予備試験の点推定値は事後分布(次の試験の事前分布)のピークになってしまう。予備試験の結果がデータが偶然偏った場合, 1試験のみでは事後分布の中心を補正できない。(1つの経験の情報による偏り)

追加情報(海外試験の事後分布)

- 医薬専門家から次のような追加情報が得られた
 - **海外試験**：人種不特定の合計100人を対象とした、計画中の試験Aと同じ対照薬 vs 別の被検薬Zの2群2剤4期クロスオーバー試験
 - 日本人は含まれていない(と思われる)
 - 対照薬の個体内gCVは**0.4**であったと推定された
 - ばらつきが大きく推定された理由は外れ値が稀に観測されることも考えられてる。(外れ値が出るメカニズムは分かっていない)
 - 外れ値が観測された場合はgCVは0.4程度である
 - 外れ値が観測されなかった場合はgCVは0.3程度である
 - この海外試験の例数が最も多く、また4期であるため、1人あたり2回の測定があるため、ある程度信頼される推定だと考えられる
 - 海外試験のgCVの事後分布に基づきサンプリングを行った

追加情報(海外試験の事後分布)



今回の題材に対する 解決アプローチのProposal

- 今回の方法では予備試験の事前分布を無情報にしていたが、臨床的な判断に基づいて、**予備試験の事前分布**を操作する。
 - 定性的に分かっているものは事前分布として利用する.
- 予備試験の点推定値は事後分布のピークになってしまう。予備試験の結果が(もし)たまたま偏った結果を生んでしまった場合、1試験の事後分布の偏りを補正できない。
- 海外試験の結果も加味して、**計画中の試験Aの事前分布**を設計する。

アプローチ①

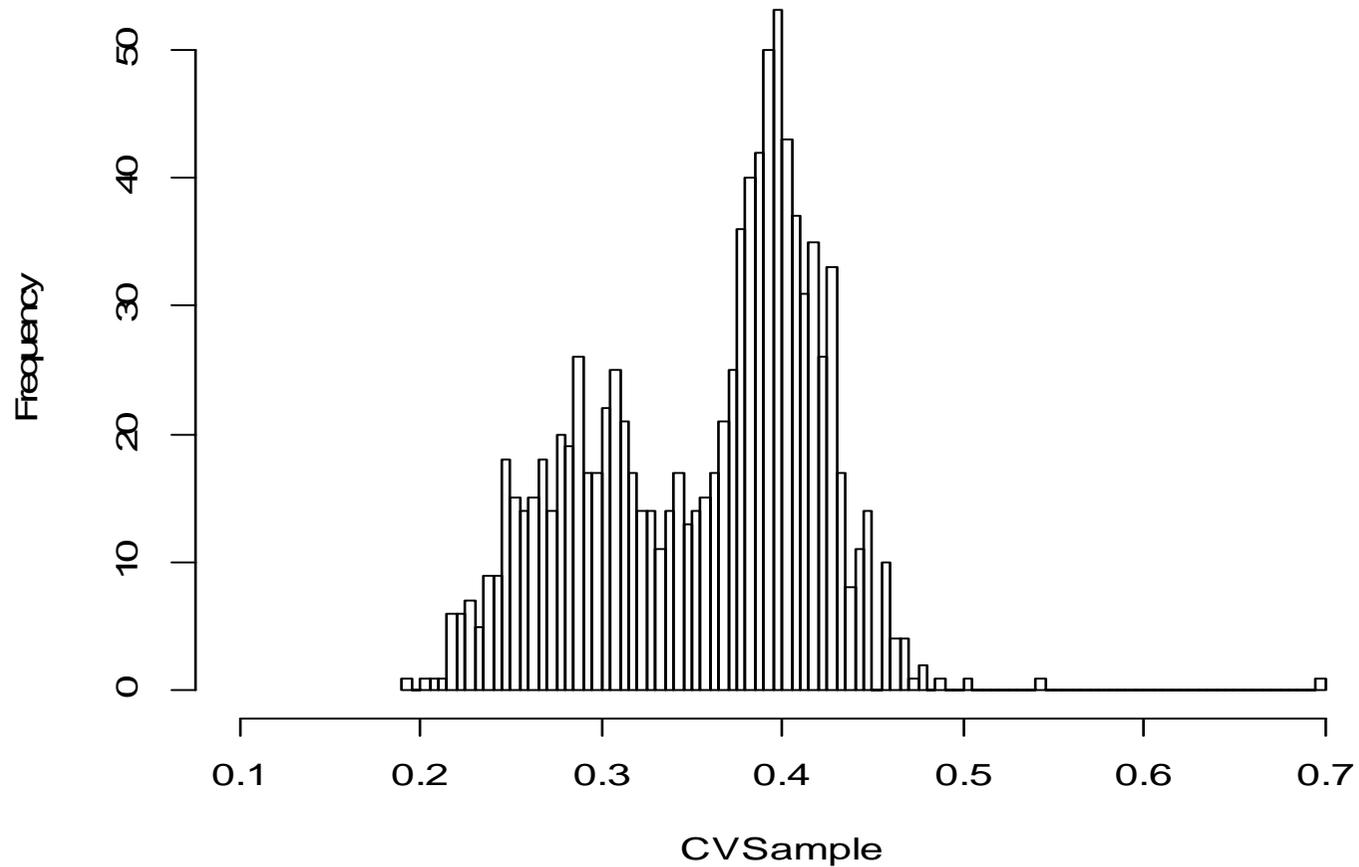
- 海外試験のgCVと予備試験のgCVの違いのみに着目したgCVの事前分布を考える.
- 検討の候補となったモデルは以下であった.
 - 2つのgCVの違いを2つの χ^2 の混合分布で表す
 - 2つのgCVの違いは試験間差によるものと考え階層モデルを用いる
- 医薬専門家によると、2つの試験のgCVの違いは、試験間差や人種差によるものではなく、稀に測定される外れ値が含まれていたかどうかであった.
- 外れ値は対照薬にも被検薬 x のどちらにも出る可能性があり、gmean ratioには影響を与えないものと考えられた.
- 外れ値が計画中の試験Aで観測される確率は1/2と仮定した.
- 従って、gCVの分布は2つの試験のgCVの推定値に基づく χ^2 の混合分布で表して、Assuranceを算出した.

アルゴリズム(モンテカルロ法使用)

- 以下を $i = 1, \dots, K$ まで loop する
 - $\sigma_w^2(i)$ を確率0.5で $\chi^{-2}(N_1 - 1, s_1^2)$ から確率0.5で $\chi^{-2}(N_2 - 1, s_2^2)$ サンプルング
 - N_1, s_1^2 には予備試験の合計例数および対数スケールで推定された分散を用いた
 - N_2, s_2^2 には海外試験の合計例数x2および対数スケールで推定された分散を用いた
 - $\delta_{(i)} | \sigma_w^2(i)$ を $N(\bar{d}, \sigma_w^2(i) / (N_1/2))$ からサンプルング (N_1, \bar{d} には予備試験の合計例数および対数スケールで推定された平均を用いた)
 - $(\delta_{(i)}, \sigma_w^2(i))$ に基づき検出力 $\text{Power}(\delta_{(i)}, \sigma_w^2(i))$ を算出する
- 最後に得られた検出力の平均を算出し, assurance の近似値とする

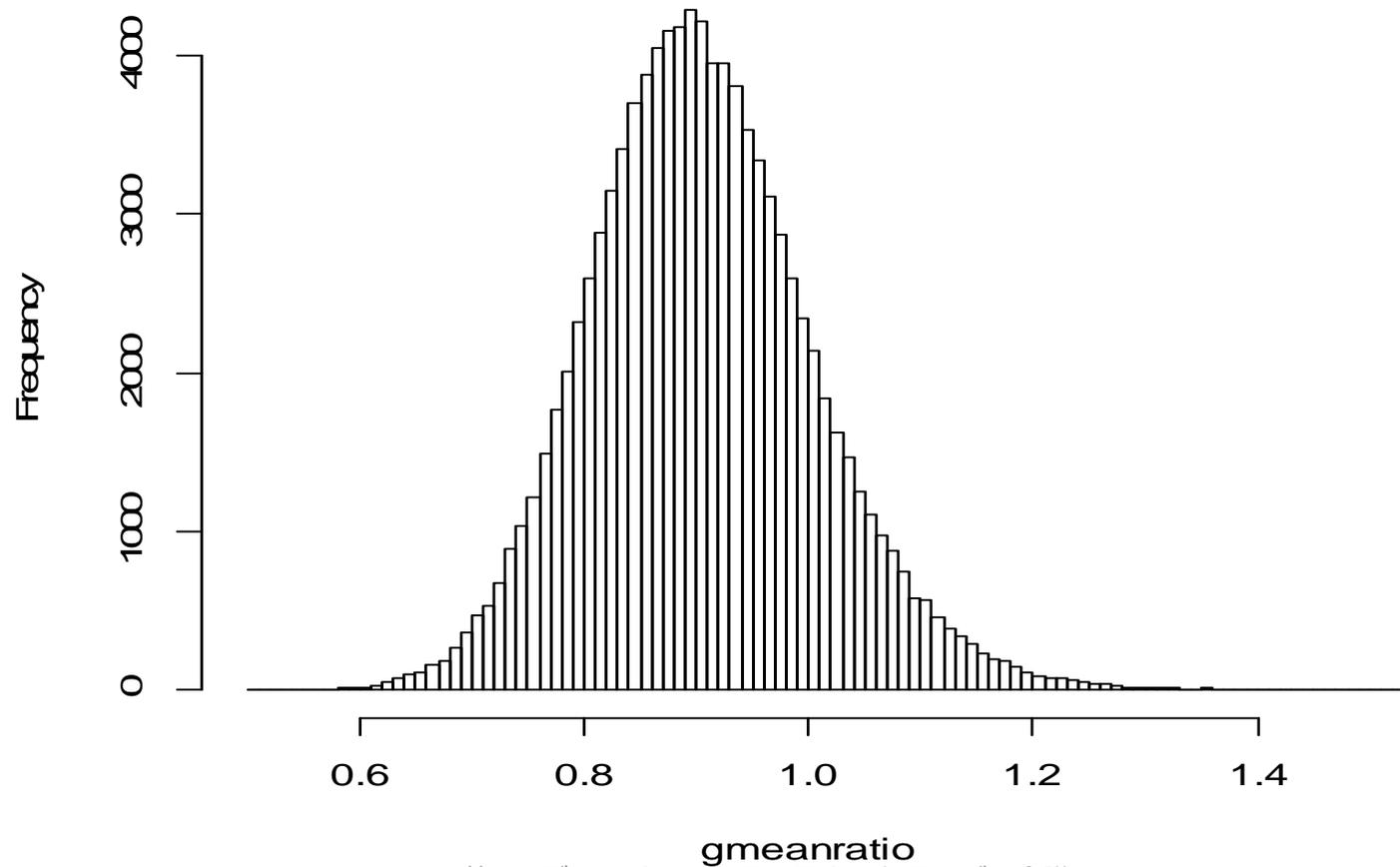
真値の分布：gCV

mixture CV



真値の分布：gmean ratio

20例Ratio



第4回 データサイエンス・ラウンドテーブル会議
2017年3月9日

例数 x Power x Assurance

合計 例数	Power	Assurance with 既知CV	Assurance with 未知CV	Assurance with mix CV
40	81	65	63	57
60	93	71	70	63
80	97	74	73	67

アプローチ②

- 予備試験の結果のみに基づく事後分布(本試験計画段階にとっては事前分布)はシンプルで初心者でも扱いやすいが、それ以外の「gcvの事前情報」や「gmeanの経験的に合理的な範囲」を反映していない。
- この要件を満たす分布を数学的に導くことも考えられるが、ここでは行わない
- そもそもやりたいことは、**開発のリスクを加味した成功率を見積もること**であったため、想定することのできる真値の分布の範囲は、定性的に決めることとした。

試験Aの事前分布に一様分布を利用

- 予備試験の結果よりもいい想定 ($g\text{mean ratio} \uparrow$ $g\text{CV} \downarrow$)に基づく見積もりは、powerが高くなるのは自明であるため興味がない。
- 一方、悪い想定 ($g\text{mean ratio} \downarrow$ $g\text{CV} \uparrow$)を見積もる場合にどの程度powerが小さくなりAssuranceに影響を与えるかが興味のあるところ。
- 従って、事前分布に予備試験よりも悪い想定のみを範囲とした一様分布を用いることとする。

本試験の事前分布に一様分布を利用

- 以下のような一様分布を定義し、Assuranceを求めた。

パラメーター	予備試験の結果	事前分布の一様分布の範囲
Gmean ratio	0.9	[0.85-0.90]
gCV	0.3	[0.3-0.4]

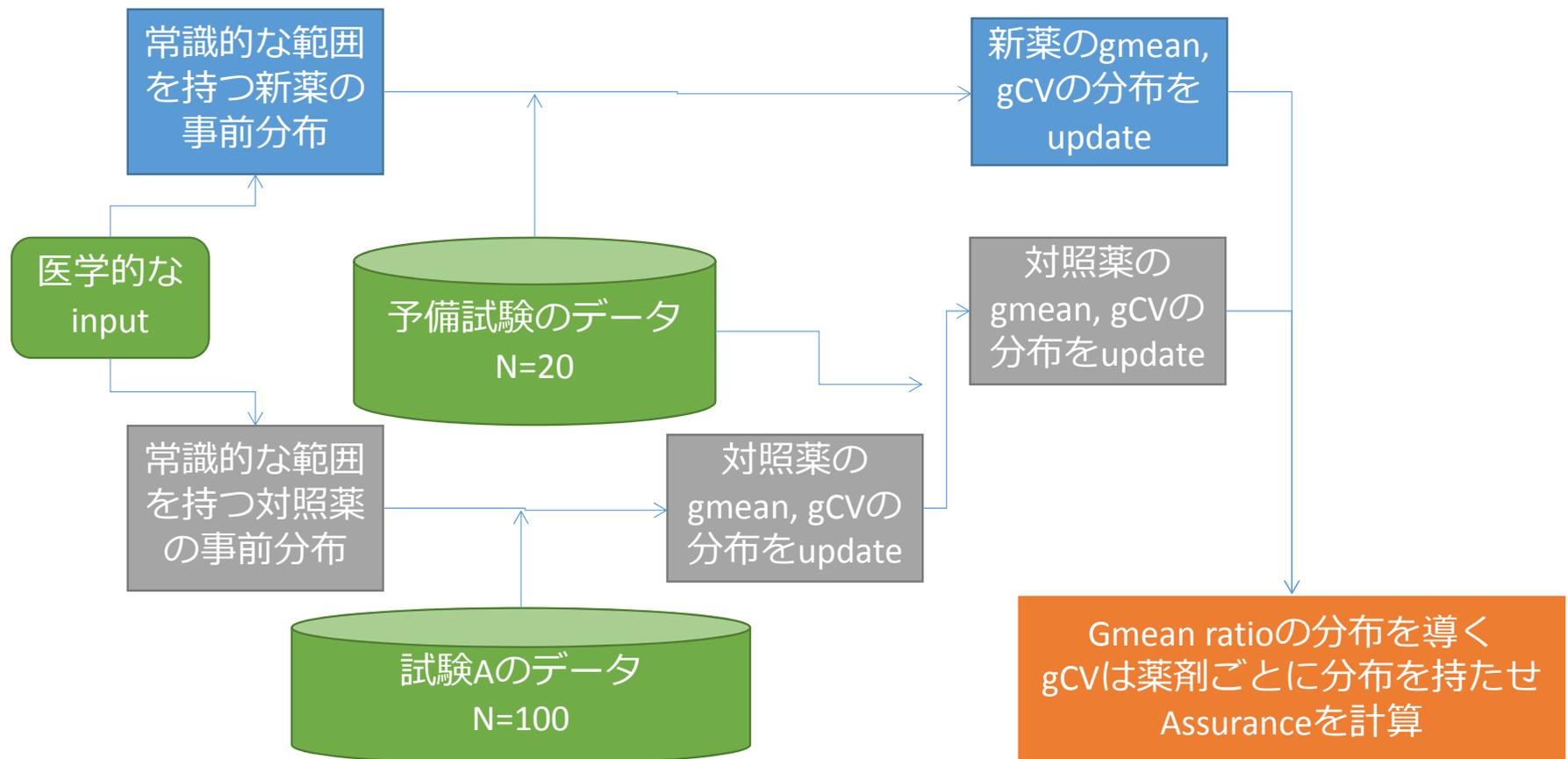
例数	Power	Assurance with一様分布
40	81	44
60	93	57
80	97	66

第4回 データサイエンス・ラウンドテーブル会議
2017年3月9日

例数 x Power x Assurance

合計 例数	Power	Assurance with 既知CV	Assurance with 未知CV	Assurance with uniform	Assurance with mix CV
40	81	65	63	44	57
60	93	71	70	57	63
80	97	74	73	66	67

選択されなかったアプローチ③



ここから学べること

- 少ない例数の予備試験から導かれた事前分布はあまり情報を与えてくれない。また、情報が偶然偏ることがあるため、その**事前分布がそのまま使えるとは限らない**。
- 試験計画時の事前分布を構成する場合、複数の情報・臨床的な知見をまとめることは高度なテクニックを要する。

【Discussion Pointと議論の結果3】

- Assuranceは例数設計の話か？
 - Assurance の値をもとにした例数設計はできるのではないが、ただし、検出力と同様の80%, 90%を基準とすべきか、については検討の余地がある
 - Assurance がばらばらすぎて、基準がないため例数決められないか？
- 予備試験計画段階、または本試験計画段階で、gmean ratioの分布はxx~xxのような範囲が想定される場合、どのような事前分布の構成ができるでしょうか？
 - 事前分布は得られた情報から保守的にみるべき。
 - 頭打ちになる現象があるので、より事前分布の詳細な検討が必要である。
 - 事前分布の設定に柔軟性があるので、恣意的な操作ができてしまいそうな点は、注意が必要であろう
- その他の論点
 - ベイズ流の方法を使用しているのが試験計画段階のみであり、試験終了後の判断には全く用いていないため、受け入れられ易いかもかもしれない

ベイズを使うその他の利用方法の検討

Bioequivalence studyへの応用

後発品医薬品～ガイドラインより

- 後発品医薬品の生物学的同等性ガイドラインでは、血中濃度のばらつきの大きな薬剤のために、**gmean ratioの点推定**による判定([0.9-1.11]の区間に点推定が入るかかどうか)の記載がある
- ここでは、臨床試験評価のみを概観する。(その他の溶出などの条件は満たされていると仮定)
- 事前情報の分布に基づき、得られるgmean ratioの点推定が0.9-1.11の区間に入るかどうかを検討する。
- ここではgmean ratioの95%CIが0.8-1.25の区間に入るかは簡単のため考慮しない

後発品医薬品の生物学的同等性ガイドライン抜粋

2) 同等性の判定

試験製剤と標準製剤の生物学的同等性評価パラメータの対数値の平均値の差の90%信頼区間が、 $\log(0.80) \sim \log(1.25)$ の範囲にあるとき、試験製剤と標準製剤は生物学的に同等と判定する。

なお、上記の判定基準に適合しない場合でも、試験製剤と標準製剤の生物学的同等性評価のパラメータの対数値の平均値の差が $\log(0.90) \sim \log(1.11)$ であり、試験製剤と標準製剤の平均溶出率の比較において、第3章、B. IV. 4. に従った溶出試験で溶出挙動が同等と判定された場合には、生物学的に同等と判定する。ただし、この規定が適用されるのは、本試験で総被験者数20名（1群10名）以上、あるいは本試験及び追加試験を併せて総被験者数30名以上が用いられた場合に限られる。

参考パラメータの評価については、経口即放性製剤及び腸溶性製剤の項に準じる。

P(gmean in [0.9 – 1.11])

- これまで同様に δ, σ_w^2 をサンプリングし, 以下を算出した
い

$$\int_0^{\infty} \int_{-\infty}^{\infty} P(\text{Satisfy requirement} | \delta, \sigma_w^2) f(\delta, \sigma_w^2 | \text{data}) d\delta d\sigma_w^2$$

- パラメーター δ, σ_w^2 をサンプリングし, 計画中の試験から得られる新たな平均 \bar{d}_{new} は以下の分布に従うので,
 $\bar{d}_{new} | \delta, \sigma_w^2$ の分布から, $[\log(0.9), \log(1.11)]$ に入った回数を算出し, 割合を算出する

$$\bar{d}_{new} | \delta, \sigma_w^2 \sim N(\delta, \sigma_w^2 / n_{new}) \quad \longrightarrow \quad P(\log(0.9) \leq \bar{d}_{new} \leq \log(1.11) | \delta, \sigma_w^2)$$

アルゴリズム(モンテカルロ法使用)

- 以下を $i = 1, \dots, K$ まで loop する
 - $\sigma_w^2(i)$ を $\chi^{-2}(N-1, s^2)$ からサンプリング
 - N, s^2 には予備試験の合計例数および対数スケールで推定された分散を用いた
 - $\delta_{(i)} | \sigma_w^2(i)$ を $N(\bar{d}, \sigma_w^2(i)/(N/2))$ からサンプリング
 - N, \bar{d} には予備試験の合計例数および対数スケールで推定された平均を用いた
 - $N(\delta_{(i)}, \sigma_w^2(i))$ から, $d_j (j = 1, \dots, K)$ をサンプリングし, $d_{(i)}$ を算出する
- 最後に $d_{(1)}, \dots, d_{(K)}$ のうち, $[\log(0.9) - \log(1.11)]$ に入った割合を計算する

例数 x Prob in [0.9-1.11]

予備試験の 合計例数	計画中の試験A の合計例数	Prob in [0.9-1.11]
10	40	43
10	60	43
10	80	43
20	40	48
20	60	48
20	80	48

計画中の試験Aの例数依存的でない。
(試験Aの例数を増やしても傾向は変わらない)

予備試験の例数設計

予備試験の事後分布を構成することは、計画段階の試験の成功にどれだけ確信度を与えるか？

予備試験の例数設計

- 予備試験の計画段階で、その後の検証試験を実施すべきかどうかを判断するのに十分な例数を計画しておくのも、Assuranceの利用方法の一つと考えることもできる。
- ここでは、予備試験の例数がその後の臨床試験に進むべきかどうかの判断材料としてのAssuranceを算出するのに十分な例数かどうかを考える。
- 利用可能な予算(例数)は予備試験とその次の臨床試験の合計例数が100例までとする。
- シンプルにするために、その他の情報は利用可能でないことを想定する。 (海外試験や医学専門家のインプット)

予備試験の例数設計の例

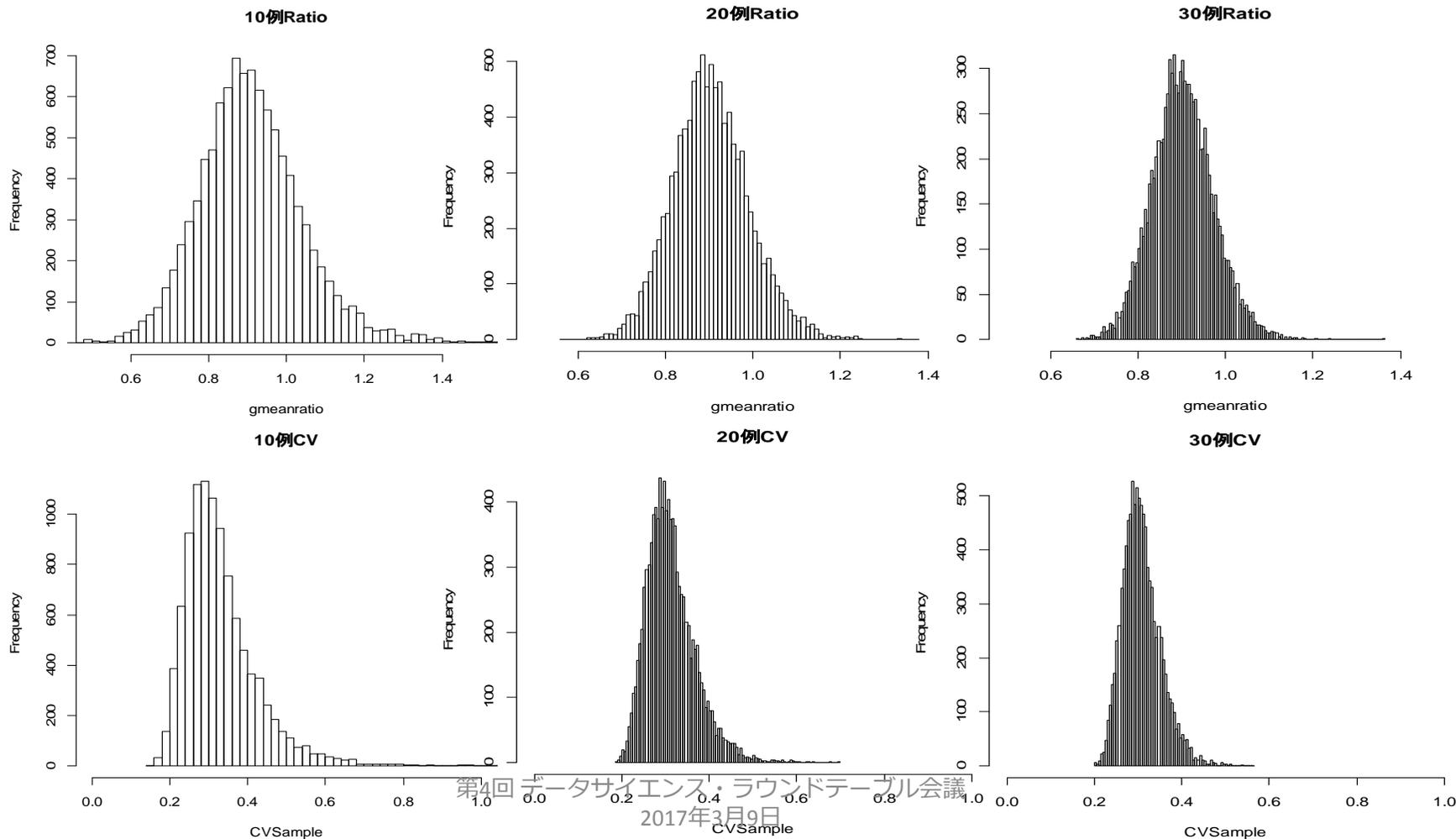
予備試験，次の試験の合計例数を60,80,100例で固定し，予備試験の例数を動かすと...

次の試験の例数 Statistical Power Assurance	シナリオ1 予備試験 N=10例	シナリオ2 予備試験 N=20例	シナリオ3 予備試験 N=30例
2試験合計の例数=60	次の試験の例数=50 Power=88% Assurance=58%	次の試験の例数=40 Power=81% <u>Assurance=63%</u>	次の試験の例数=30 Power=70% Assurance=61%
2試験合計の例数=80	次の試験の例数=70 Power=96% Assurance=62%	次の試験の例数=60 Power=93% Assurance=70%	次の試験の例数=50 Power=88% <u>Assurance=71%</u>
2試験合計の例数=100	次の試験の例数=90 Power=98% Assurance=64%	次の試験の例数=80 Power=97% Assurance=73%	次の試験の例数=70 Power=95% <u>Assurance=76%</u>

予備試験の例数と事後分布

- 目的の検証的な臨床試験のAssuranceを上げるためには、よりよい事前分布を計画的に設計することが一つの目標になり得る.
- 臨床試験全体の利用可能なコストは固定されていると仮定すると、Assuranceをできるだけ大きくなるような各試験への例数の割り振りを考えることは、より効率的な開発計画を与えると考ええる.

予備試験の事後分布(N=10,20,30)



【Discussionの結果(まとめ)】

- ベイズ統計を用いると、「過去のデータを利用できる」ため、一見「頻度論よりも適切な」結果が得られそうです。本当にそうでしょうか？
 - **事前情報を考慮することができる、という点は正しそう。一方で、assurance は、ばらつきを考慮することで「真値が得られているという想定のもとでの」検出力より値が小さくなりがち。**

【Discussionの結果(まとめ)】

- 頻度論の例数設計で、平均・分散に複数の値を割り振って「総合的に判断」する方法は、ある意味ではベイズ流の考え方ともとらえられないでしょうか？その場合、今回の方法との違いは何でしょうか？
 - 頻度論では、「総合的に判断」する部分に恣意性が入り得るが、assurance を用いる場合、事前分布等の状況を与えたもとでは、自動で計算される。一方で、事前分布等の設定に恣意性が入り得る。

【Discussionの結果(まとめ)】

- 全体的なまとめ

- Assurance を算出すること自体は意味がありそうである。また、計画段階で ベイズ流の方法を用いているが、実際の試験終了後の判断は頻度論の手法で行うため、規制当局等にも受け入れられ易いのではないか。
- 検討事項は、事前分布の設定を含め多くあるため、使用の際には十分な検討が必要である
- 頻度論の方法を「置き換える」と考えるよりは、「情報を追加することを目指す」とよいかもしれない
 - Assurance だけでなく、途中経過の δ の値や、そこでの検出力、等も出力することで、よりよい判断が行える可能性があるのではないか
 - 事前情報に様々な試験の状況があるときは、事前分布に組み込むことでよりよい判断ができるかもしれない