

医療機器における臨床試験のあり方
必要か・不必要か
評価すべきは安全性か有効性か
: 最新のFDAガイドラインからの科学的考察

第5回 医療機器臨床試験研究会

2006年7月22日

古川敏仁

株式会社バイオスタティスティカル リサーチ

医療機器と医薬品の違い

- 機器の効果は主として物理学的特性に基づくものであり、生体との化学反応に基づくものではない
{化学反応が、機器の特性の重要な要素であるものも存在するが(例: Drug Eluting Stent)}
- 材質、機器のデザインに性能(効果)が依存する
非臨床試験、物理的なテストの情報が重要

医療機器と医薬品の違い

- 製品のライフサイクルが早い
- 逐次的な改良、改善が日常的になされる。また、個々の改善の効果は小さく、臨床試験では確認が難しいことも多いが、小さな改善の蓄積が、大きな改善となる

医療機器と医薬品の違い

- 臨床試験時のマスクが難しく、評価の客観性の保証が重要となる
- 挑戦的な課題も多い(例:人工臓器)
- 経済面からは、医療機器のマーケットは医薬品に比べ小さなものもあり、コスト-便益的な観点も臨床試験デザインでは重要となる。

Q1: 臨床試験が必要なのはヒトの体内で使用するからである？

- 回答： 科学的な意味では違います。非臨床試験や物理学的試験では証明できず、臨床試験でしか安全性・効果が証明できない場合のみ臨床試験が必要です。また、機器の特質により、臨床試験では有効性や安全性に関する情報を、試験デザイン面、倫理面から得ることができない場合もあります。このような場合、臨床試験は実施できません。
- ヒトでの使用経験（有効性、安全性の検証以外で必要とされる臨床試験の例：
First in man(FIM)試験、Feasibility試験
 - 有効性、安全性の最低限の確認
 - Learning Curve

臨床試験は常に必要か？

有効性は常に主要評価項目か？

- どのような医療機器の臨床試験においても有効性と安全性を検証することが求められるわけではない。
(機器の特質で人に用いるために欠けている情報を求めることが臨床試験の目的)
- 多くの場合、有効性の検証よりも安全性の確認が臨床試験の主要評価項目となる。
(有効性は医療機器のデザイン、非臨床試験、物理テストで担保される場合が多い)

例：安全性の確認を臨床試験の目的とするFDAガイドライン例

- FDA 21CFR870.3925
Replacement Heart Valve Guidance
- 主要評価項目：安全性
- 試験デザイン：Single-arm試験

- US Food and Drug Administration. Replacement Heart Valve Guidance. Version 4.1. Washington, DC: US FDA; 1994

Replacement Heart Valve Guidance

- 心臓の弁の説明と、人工弁の見本の図、写真 本資料に掲載許可していただける資料募集しています
- また、個々で取り上げているFDAガイドライン、ISOガイドラインの正式な入手方法ご存知の方、お教えてください。

Replacement Heart Valve Guidance

推奨する試験デザイン

- 条件：
人工弁を必要とする患者に対して、
安全性と有効性を
現在販売されている医療機器と比較して証明

ただし、FDAが確立したOPCを用いても良い

Replacement Heart Valve Guidance

推奨する試験デザイン: OPC

- 主要評価項目: OPC: Objective Performance criteria
・安全性の評価指標
- 試験デザイン: Single-arm試験
- 判定基準: 機器の特質から考えて最も重要な合併症発生率が、OPCの2倍以内であることを証明

OPC

Replacement Heart Valve Guidance

- 主要評価項目 : OPC : Objective Performance criteria
- PMA、公表論文の中から、証拠能力の高いデータ約30000弁の臨床データから確立
- 人工弁を使用した場合の標準的な合併症発生率
- 合併症発生率は機器の使用時間によらず一定と仮定

OPC

Replacement Heart Valve Guidance

Appendix K

	機械弁	生体弁
血栓塞栓症	3.0	2.5
弁血栓	0.8	0.2
すべての出血	3.5	1.4
主要な出血	1.5	0.9
すべての弁周囲逆流	1.2	1.2
主要な弁周囲逆流	0.6	0.6
心内膜炎	1.2	1.2

(単位 % 分母 弁・年)

OPC

ISO 5840

- ISO 5840 Cardiovascular implants - Cardiac valve protheses
- Dr. William N. Anderson:
1982/1/1 ~ 1999/12/31までの心臓弁置換を受けたブリティッシュコロンビア大学の患者6738名、移植総数7186件、有害事象4086件のデータより

OPC

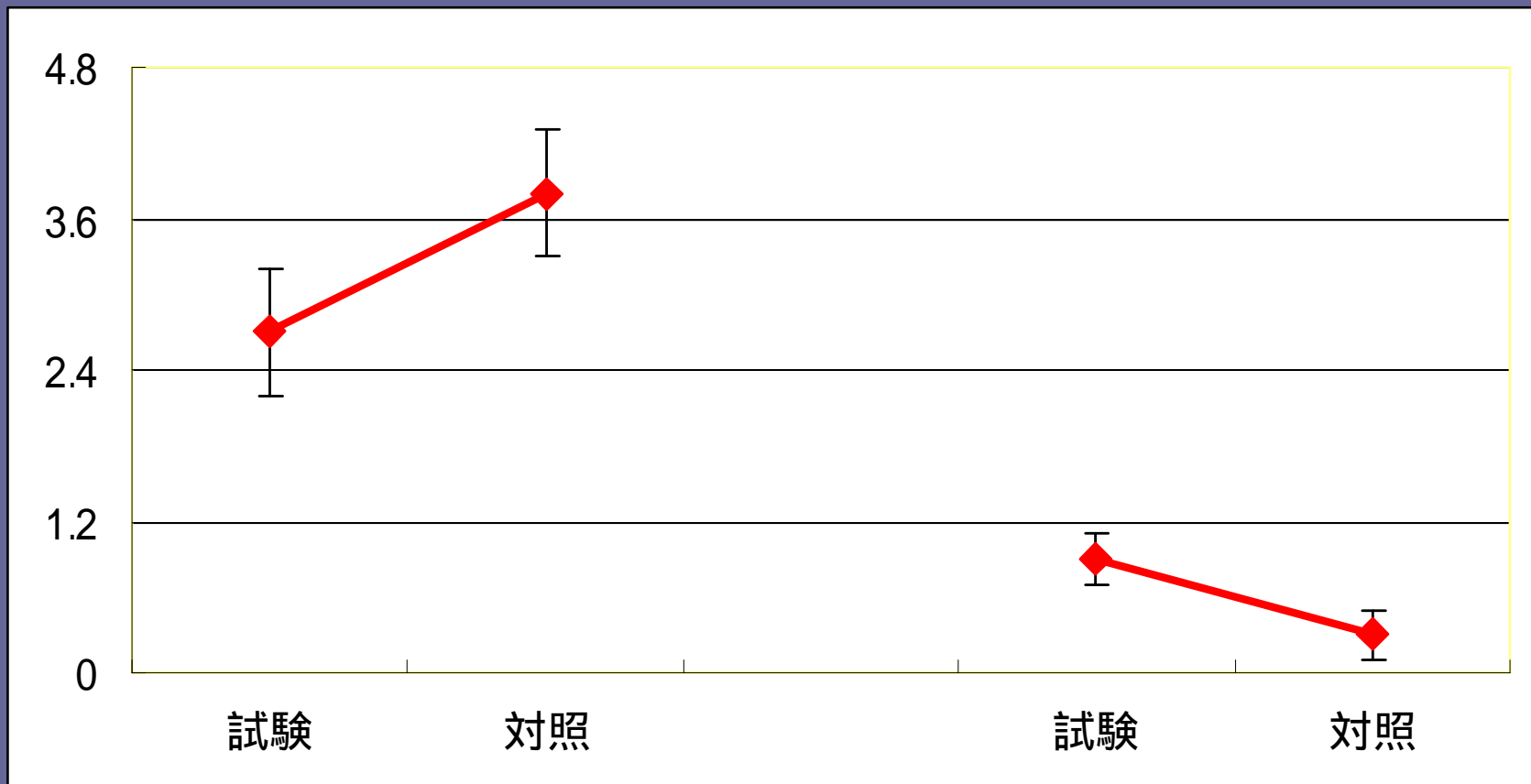
ISO 5840 Dr. William N. Anderson:

表 R.2 - 1 年目有害事象発生率 (%)

有害事象	大動脈弁 (生体弁)	大動脈弁 (機械弁)	僧帽弁 (生体弁)	僧帽弁 (機械弁)
構造的劣化	0.03	0.00	0.12	0.00
血栓塞栓症 (恒久的、RIND)	2.06	2.78	2.48	2.63
弁血栓	0.10	0.00	0.19	0.61
抗凝固関連の出血	0.45	2.44	0.80	1.95
心内膜炎	0.59	0.93	0.68	0.54
非構造的機能不全 / 弁周囲逆流	0.38	0.84	1.05	1.75
再手術	0.77	1.09	1.05	1.95

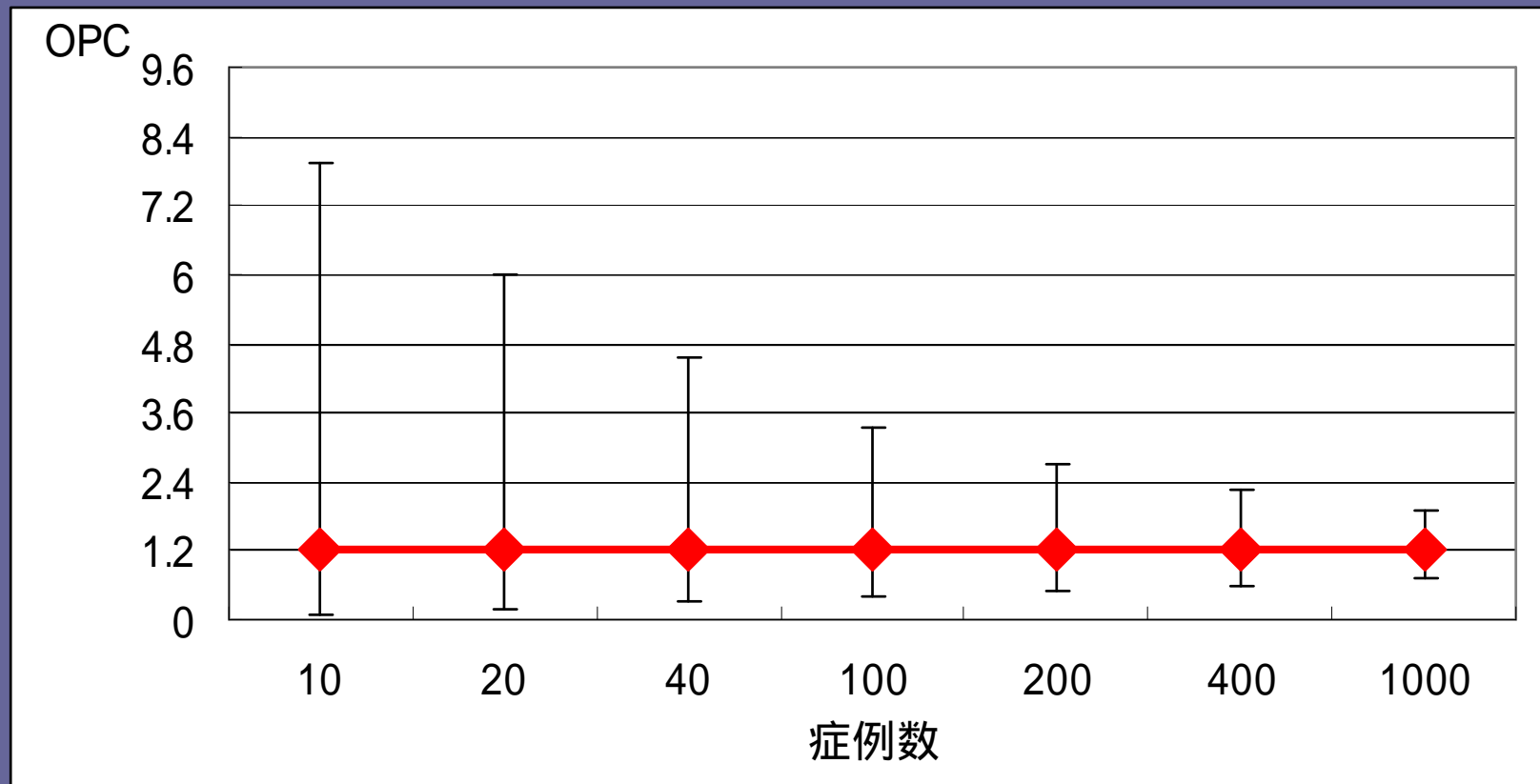
無作為化比較よりsingle arm Replacement Heart Valve Guidance

- 理由1: 絶対基準の重要性



無作為化比較よりsingle arm Replacement Heart Valve Guidance

- 理由2: 推定精度



無作為化比較よりsingle arm Replacement Heart Valve Guidance

- 理由2 推定精度
- 絶対基準との比較に比べ、対照群との比較は、同じ差を検定する場合
- 各群 2倍の例数、Totalで4倍の例数が必要になる。
- もし、1群例数が4倍になれば、パラメータの推定精度2倍(信頼区間幅1/2)となる。

無作為化比較よりsingle arm

Replacement Heart Valve Guidance

- 理由3 臨床試験データから得られる有効性に関する知見のほとんどは、機器の性能で説明できるため
- 機器の性能: デザイン、材質
- 有効性指標: 血行動態 (最高血流など)
- 主要評価項目が有効性ではなく、安全性なのもこの理由

無作為化比較よりsingle arm Replacement Heart Valve Guidance

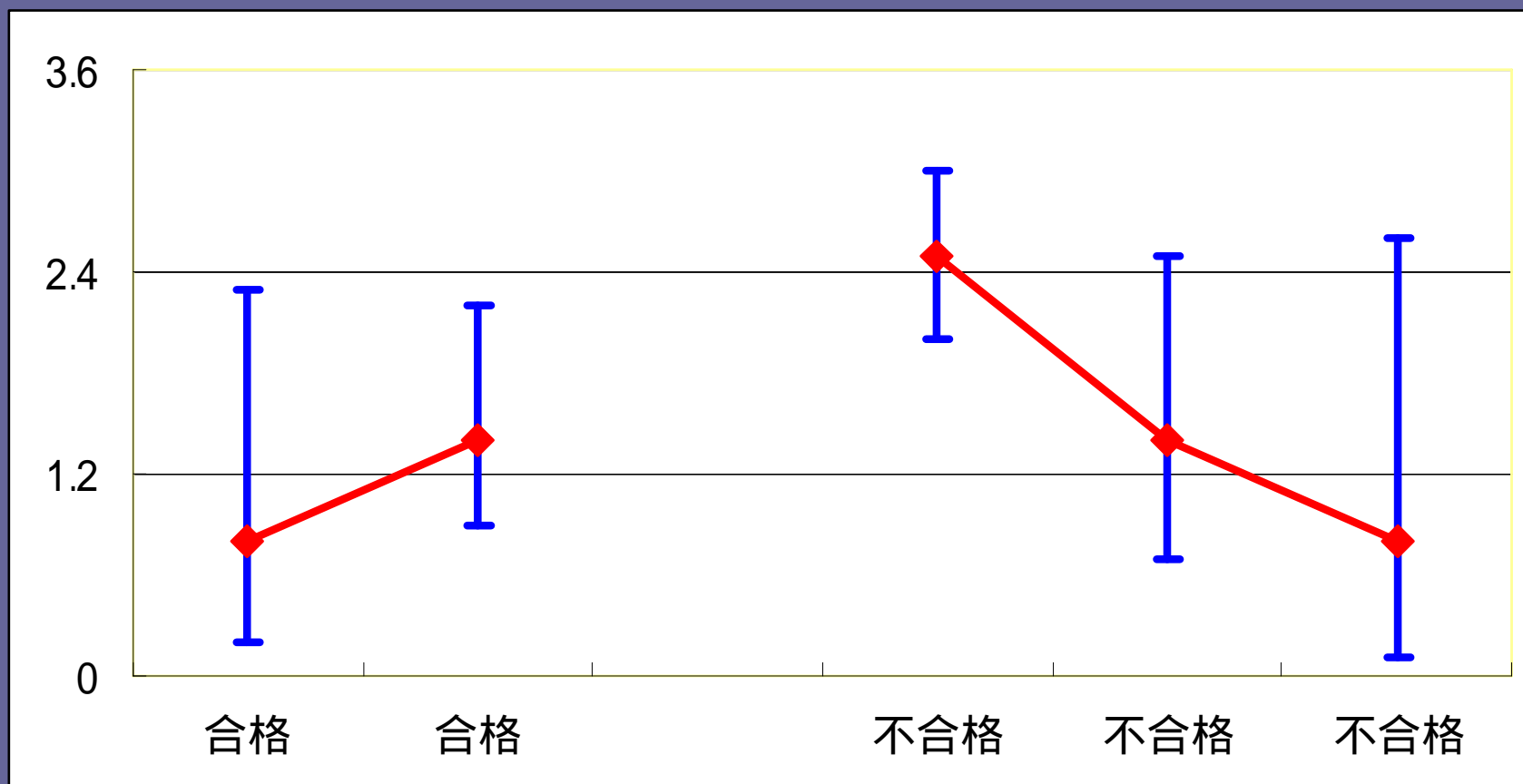
- その他の理由
- 比較可能性の保証の難しさ
- 臨床試験の途中で機器の改良
- 厳密な意味での2群比較は難しい
- ‘ Alternatives to Randomization in Surgical Study ’, Gary L. Grunkemeier and Albet Starr. J. Heart valve Dis. Vol1 No2,1992

試験結果の判定基準

Replacement Heart Valve Guidance

- 主要な合併症の発生率が95%の確率で OPC (Historical control) の2倍以内であることを、信頼区間で示す。
- 両側有意水準 =0.05の検定で、有意になることと同値

試験結果の判定基準 Replacement Heart Valve Guidance



試験例数の設定

Replacement Heart Valve Guidance

- 主要な合併症発現率の95%信頼区間が OPC (Historical control) の2倍以内であることを、検出力80%の試験で証明すること
- OPC年間発生率 1.2%ならば 800年・弁の観察例が必要

'Sample Size Requirements for Evaluating Heart Valves with Constant Risk Events' Gary L. Grunkemeier, Diane Macculloch Johnson, and David C. Naftel. J. Heart valve Dis. Vol3 No1,1994:53-58

Table IV: Valve-years needed for a given OPC, for all OPC's given in Table 3. The values less than 2000 valve years are plotted in Figure 3. The bolded values for an OPC of 1.2%/year with a power of (at least) 80%, gave rise to the minimal size chosen by the FDA for a clinical study of a new heart valve, of 800 valve years.

OPC	Power		
	gamma 80.0%	81.6%	Poisson 90.3%
0.2% / yr	4645	4860	6375
0.6	1549	1620	2125
0.8	1162	1215	1594
0.9	1033	1080	1417
1.2	775	810	1063
1.4	664	695	911
1.5	620	648	850
2.5	372	389	510
3.0	310	324	425
3.5	266	278	365

必要例数 ISO 5840

- 少なくとも400弁・年のフォローアップ
- 最低例数
 - 人工弁の移植者150名以上
 - 弁サイズごとに15以上
- 最低例数のフォローアップ症例については
 - 柔軟性のない人工弁 5年以上
 - 柔軟性のある人工弁 10年以上のフォローアップが追加的に必要

例1:まとめ

- 比較試験ではなく、Single-arm試験を推奨
- 主要評価項目は安全性(合併症発生率)
- 試験の評価基準が設定されている。すなわち OPC : Objective Performance criteriaが、過去のデータベースより検討設定されている。
- 機器のデザインによる有効性の評価を重視(過去のデータベースの解析結果から)

第2例

統計学的側面からの進歩

Draft Guidance for Industry and FDA Staff Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials

- **This guidance document is being distributed for comments purposes only.**
- **Draft released for comment on May 23, 2006**
- **<http://www.fda.gov/cdrh/osb/guidance/1601.html>**



[CDRH](#) > [Guidance](#) > Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials

Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials - Draft Guidance for Industry and FDA Staff



DRAFT GUIDANCE

This guidance document is being distributed for comments purposes only.

Draft released for comment on May 23, 2006

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit written comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. Alternatively, electronic comments may be submitted to <http://www.fda.gov/dockets/ecomments>. All comments should be identified with docket number 2006D-0191.

For questions regarding this document, contact Dr. Greg Campbell at 240-276-3133 or greg.campbell@fda.hhs.gov.



U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health

Division of Biostatistics
Office of Surveillance and Biometrics

***Contains Nonbinding Recommendations
Draft - Not for Implementation***

Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials

- Bayesian的なアプローチ
- 医薬品の臨床試験で主として用いられている
頻度論的 (frequentist) 方法論

従来の臨床試験

- 頻度論的 (frequentist) な方法論に基づく
- 臨床試験で得られたデータのみを用いて解析する。
- 過去のデータ・証拠は、試験デザインの参考や例数設計、あるいは、解析結果の考察には用いられる。
- が、試験結果と同等のレベルで解析に共用されることはない。

Bayesian的アプローチの臨床試験

- ベイジアン (Bayesian) 的な方法論に基づく
- 過去に蓄積されたデータ・知見・証拠 (evidence) を利用、臨床試験結果と過去の証拠を結ぶ解析をする。
- 事前情報 (過去のエビデンス) を、臨床試験結果の定量的評価に活用できる。
- 試験のデザイン、解析ともに過去の証拠を利用することができる。

Bayesの定理

$$p(\theta | y) = \frac{p(y | \theta) p(\theta)}{p(y)}$$

$P(\theta)$: 事前確率 : 前回までの θ の確率に関する情報

$P(y)$: 今回の試験結果の確率に関する情報

$P(y | \theta)$: 尤度

$P(\theta | y)$: 事後確率 : 前回までの情報に

今回の試験結果を加味した θ の確率に関する情報

Bayesの定理

事後確率 \propto 事前確率 \times 尤度

$P(\theta)$: 事前確率: 前回までの θ の確率に関する情報

$P(y)$: 今回の試験結果の確率に関する情報

$P(y|\theta)$: 尤度

$P(\theta|y)$: 事後確率: 前回までの情報に

今回の試験結果を加味した θ の確率に関する情報

Bayesian Statistics 早分かり

- prior information : 事前情報、事前確率
- likelihood : 尤度
- posterior probabilities : 事後確率
- *credible intervals* (confidence interval, type I and II error : frequentist)
- predictive distribution

- exchangeability
- 自然共役事前分布 数值的に事後分布を計算
- Gibbs sampling Computer-intensive calculations
- MCMC Markov Chain Monte Carlo

FDAにおけるBayesian法利用の利点 Medical Device

(1) 過去の有益な情報がない場合

1. 例えば、中間解析の結果を利用して試験計画の修正

1) 症例数の変更

2) 無作為割付方法の変更

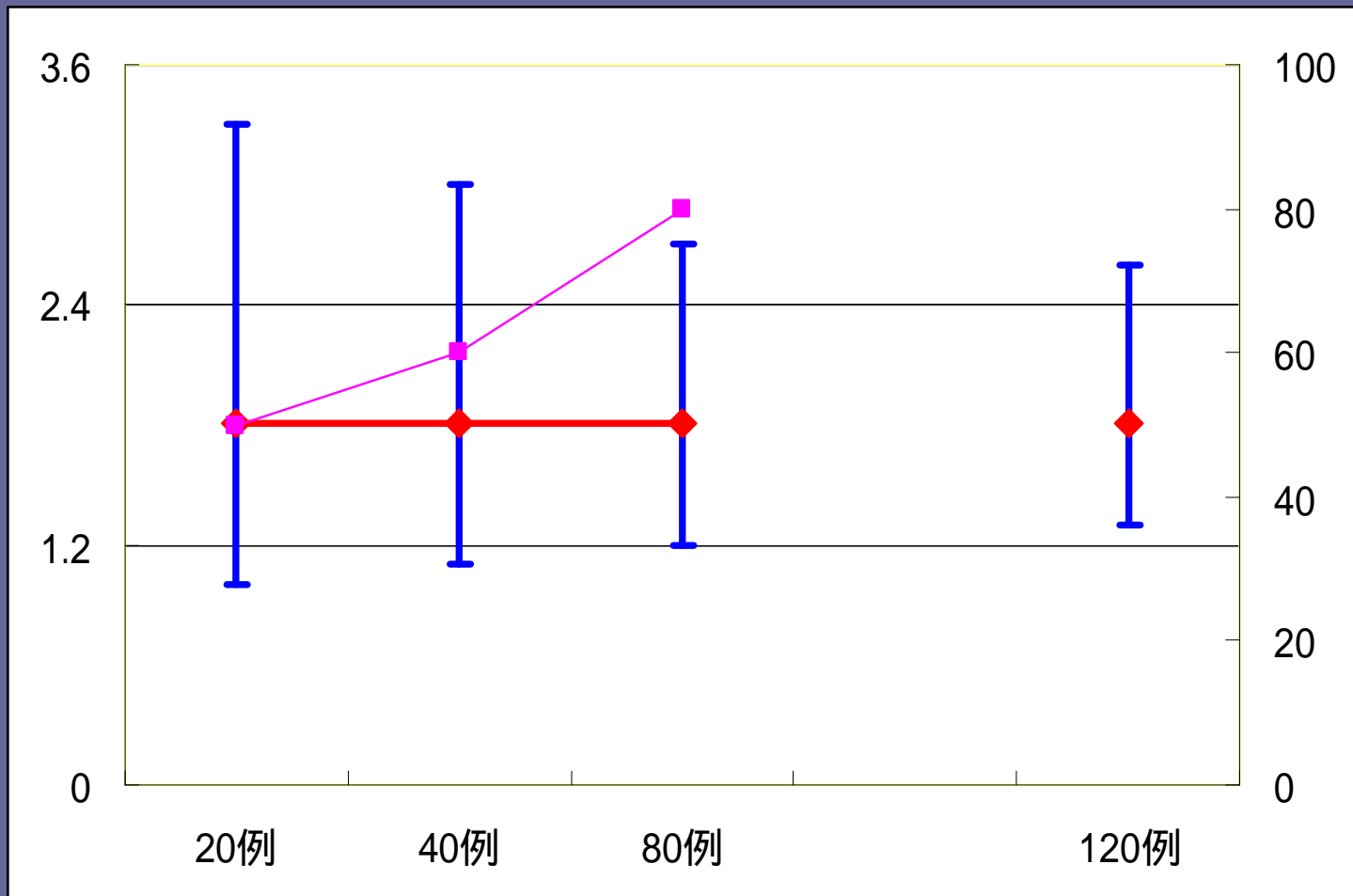
2. また、頻度論的な方法では複雑で取り扱えないような統計モデルに適用できる

predictive distributionを利用した 中間解析

- 中間解析の結果で、試験続行の可否を判定
- 例：試験続行した場合、試験が失敗する確率を計算 確率が一定以上(80%)であれば、試験中止

120例 中間解析3回の試験回の試験

最終的な合併症発生率の90%信頼区間がOPCの2倍以内
試験が成功しない確率が80%以上となったとき試験中止



Hierarchical modele

- PMA P980048(FDA CDRH), BAK/Cervical Interbody Fusion System:
- For Multi-center Trial
 - » physician training
 - » technique
 - » experience with the device
 - » patients management
 - » patient population

Hierarchical model PMA P980048

Table 7: Long-term¹ Fusion – Bayesian Analysis Results (Restricted² Cohort)

Success Measure	95% Credible Interval for Difference in BAK/C and ACDF Rates ³	Multivariate Longitudinal Bayesian Analysis Conclusion
Fusion	(+3.2, +6.1)	BAK/C is superior to ACDF

¹ Follow-up examination conducted 24 months post-operatively or later

² See *Analysis Cohorts* in Section 11.e. for an explanation of the Restricted and Unrestricted Cohorts

³ Predictive distributions calculated from Long-Term multivariate longitudinal Bayesian results

Table 8: Long-term¹ Fusion - Classical Analyses Results (Restricted² Cohort)

	BAK/C		Control		P
	x/n	%	x/n	%	
6 month	119/120	99.2	77/94	81.9	<0.001
12 month	123/126	97.6	86/97	88.7	0.010
Long-Term	86/86	100.0	66/69	95.7	0.086

¹ Follow-up examination conducted 24 months post-operatively or later

² See *Analysis Cohorts* in Section 11.e. for an explanation of the Restricted and Unrestricted Cohorts

Hierarchical model PMA P980048

- 従来の頻度論的な方法では有意ではない
P=0.088
- Bayesian 有意
信頼区間 +3.2 ~ +6.1
- Bayesian法のほうが得？
- No: 施設、背景因子で調整しただけ？
- 非線形混合効果モデルでも同じ結論？
- 申請当時、非線形混合効果モデルは利用できなかった？

FDAにおけるBayesian法利用の利点

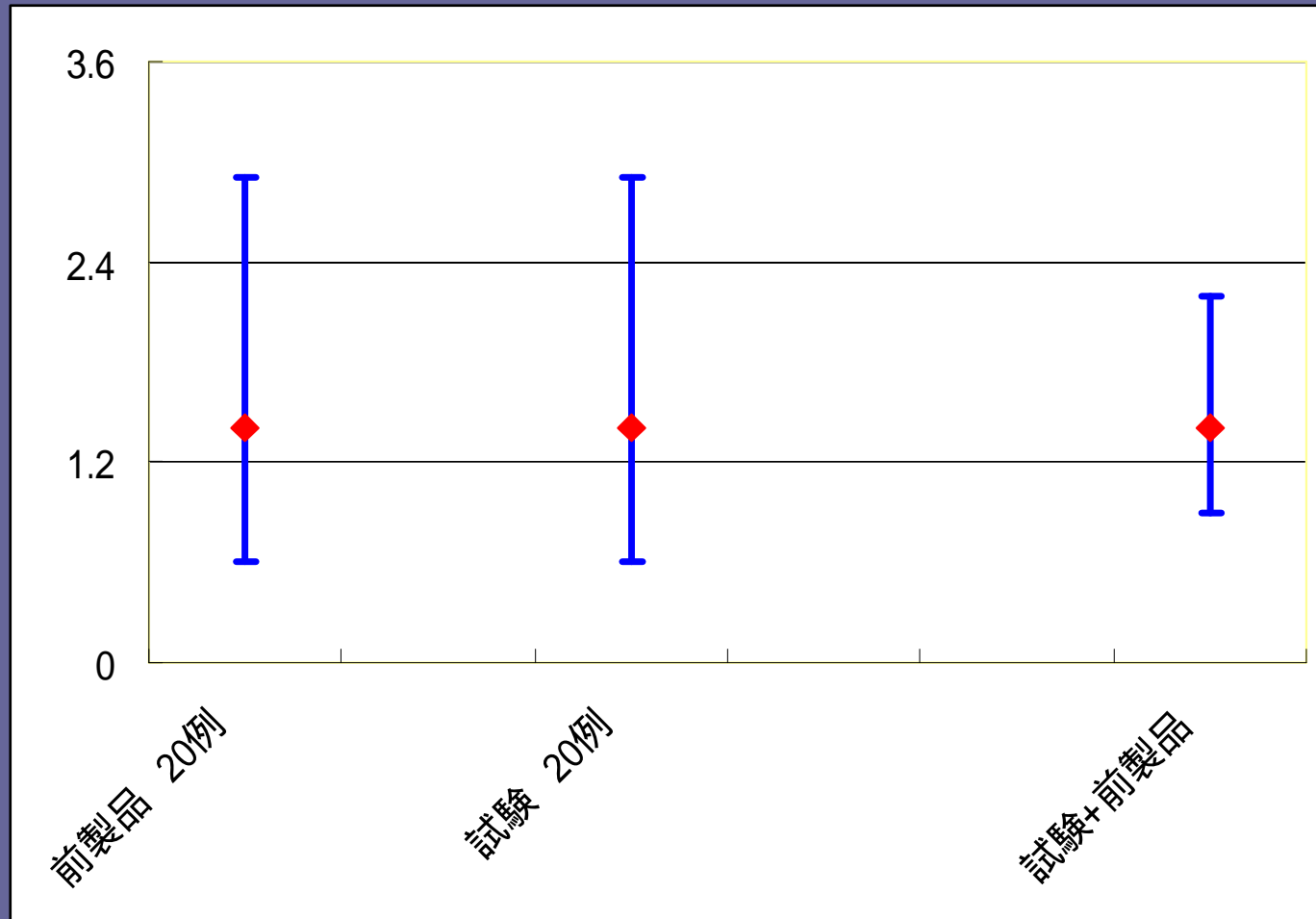
(1)過去に行われた申請のデータや、許認可の知見を活用し、

- 1) より小さな規模の臨床試験
- 2) より短な観察期間
- 3) 検定の感度を向上

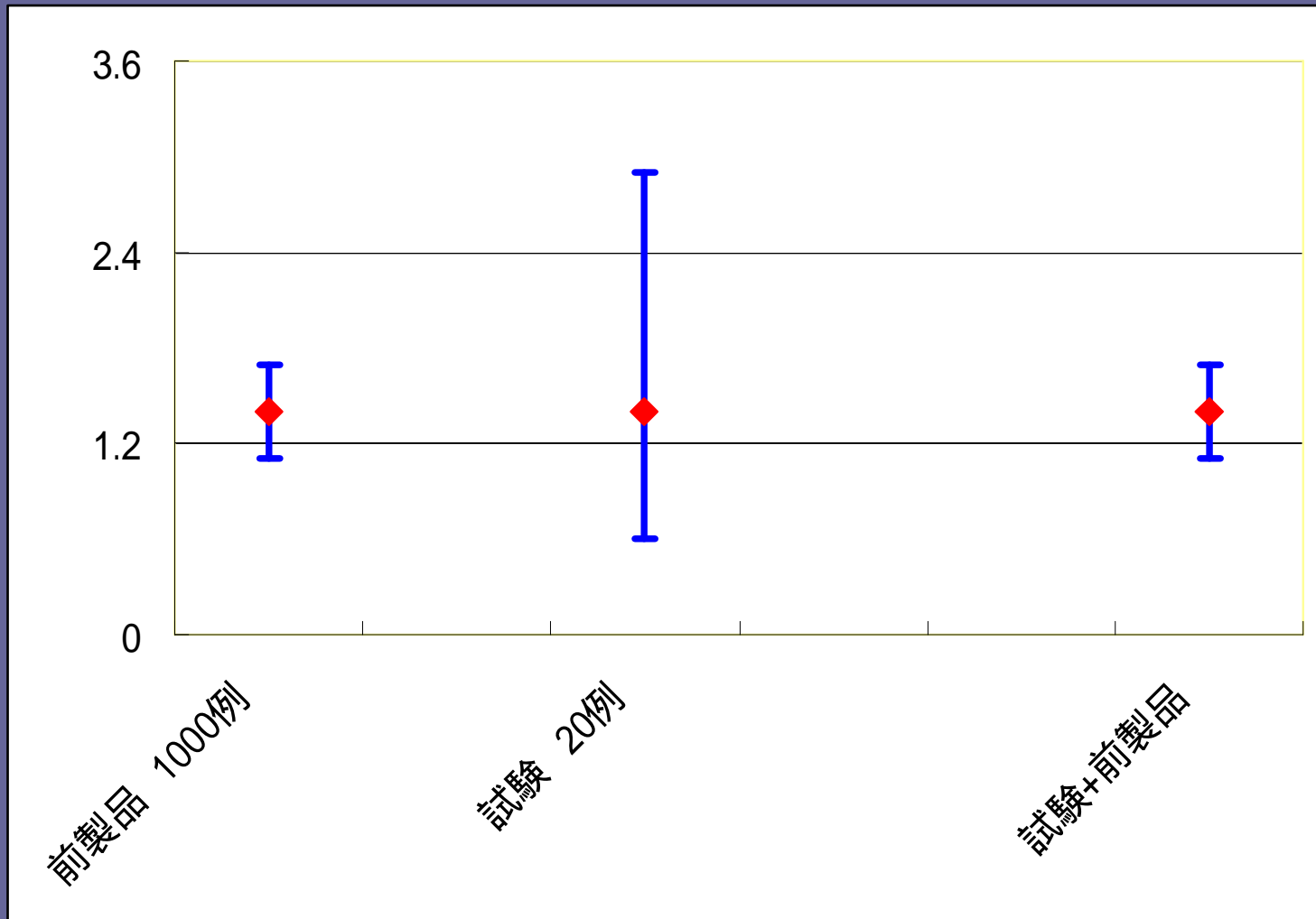
でのPivotal臨床試験を可能にする。

Borrowing strength from other studies

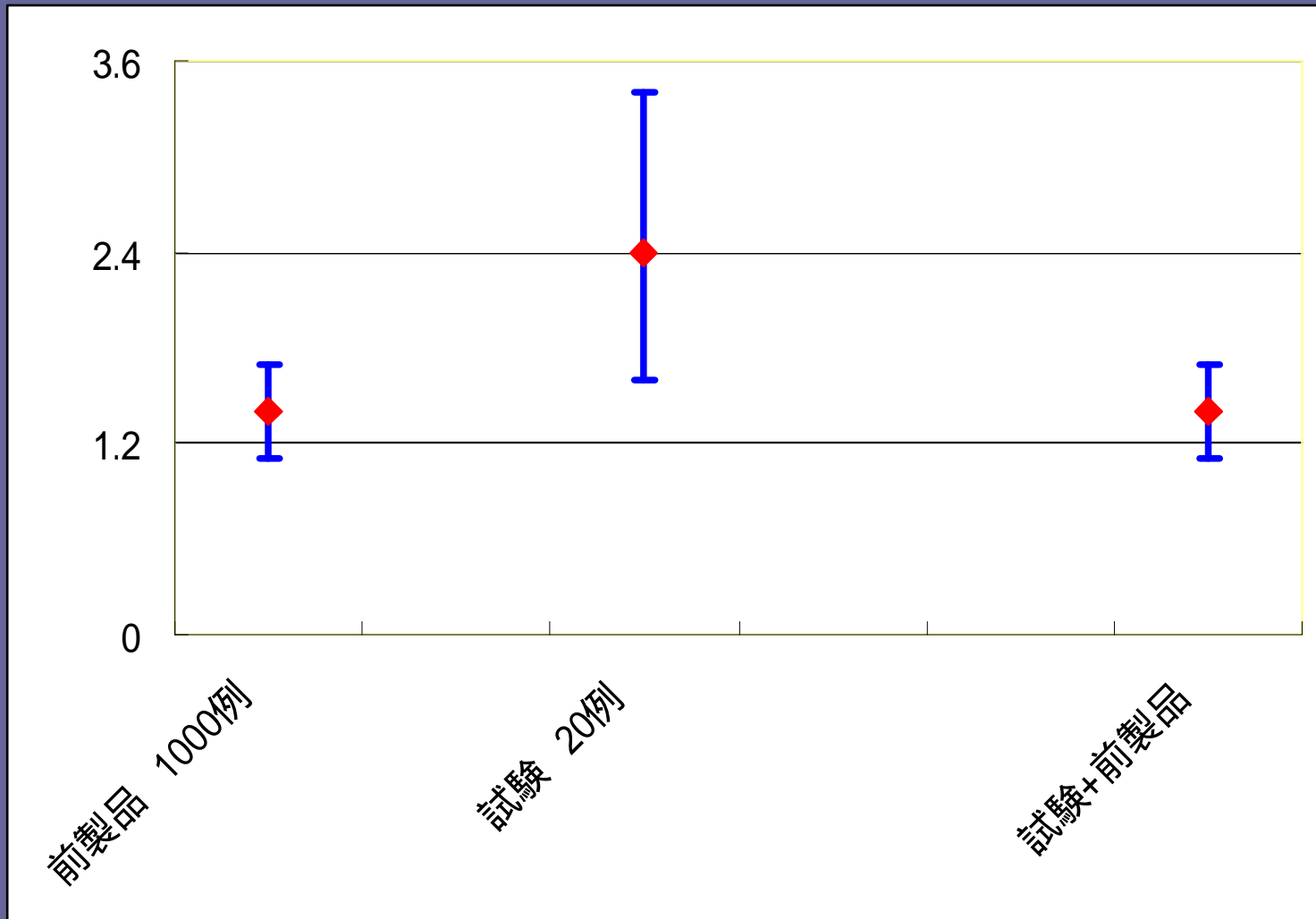
Borrowing strength from other studies



Borrowing strength from other studies ?



Borrowing strength from other studies ?



Borrowing strength from other studies ?

- 過去の試験データの重みが大きければ、臨床試験の結果に関係なく、事後確率が求まる
- すなわち、試験をしても意味がなくなる
- 過去の試験の重みを妥当な大きさに減じることが重要
- FDA (PMDA) と相談が一番重要

例2: より学ぶもの

- 頻度論的 (frequentist) 方法論もBayesian的なアプローチも試験デザインの重要性は同じ
 - 無作為化の重要性
 - 比較対照の重要性
 - 後知恵解析にならないよう、事前の解析計画の重要性
 - FDAとの協議

例2: より学ぶ最も重要なこと 自分で判断すること

- Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

臨床試験再考 無意味な臨床試験

- 統計学的評価 パラメータの分布の記述
- 評価項目のばらつきとパラメータ推定値との相対的な比較
- 「ばらつき」が考えられない評価項目を目的とした臨床試験は統計学的に意味がない

無意味な臨床試験例：義足

- 従来の義足では、材質のもろさから、30cm以上長い義足はできなかった。
- 主張：材質の改良により、60cmの義足ができた。
- 主要評価項目：足の長さ

- 結果：？

従来品 $30 \pm 1\text{cm}$

試験機器 $60 \pm 1\text{cm}$

$p < 0.0000001$

無意味な臨床試験例：義足 問題点

- 主要評価項目：足の長さ 機器の性能から明らか
- 材質の頑健性：臨床試験で調査すべきは倫理的に許されない
機械的なテストデータ重要
- 結論：臨床試験より、機器デザイン、物理テスト重要

臨床試験再考 Single Arm試験

- 絶対的な評価指標のあるHistorical Controlが必要 例：OPC
- 評価方法 信頼区間の推定
- Historical Controlから設定した閾値と信頼区間の比較
- 他の試験成績との単純な比較は無意味
 - 比較可能性が保証されない
 - 被験者背景や年代、医療技術環境の違いからくるバイアス
- 絶対的な評価指標がない場合
 - 分布の記述と、他試験データとの相対的な比較
 - 分布の重なりなどによる評価
 - 論理的妥当性により保証できる範囲の証拠
 - 「検証」はできない

臨床試験再考

検証は絶対に必要か あるいは できるのか

- 重要なのは、試験結果の判断ルール
- 判断と検証とは違う
- Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials参照

Bayesian法、厳密な意味で、頻度論的な検定、検証手法ではない。credible intervals で許容する範囲を決定

判断は誰がするのか

- PMDA ? メーカー ?
- 機器の開発に関与するもの全員
- 重要なのは、合理的に結論を導くこと = ひとつの科学