

医療機器臨床試験における 例数設計のいろは

Sample size estimation for clinical trials of devices -First education-

株式会社バイオスタティスティカル リサーチ
古川敏仁

Furukawa Toshihito, Biostatistical Research

2005年9月3日

第2回 医療機器臨床試験研究会

Copyright(C) Biostatistical Research Co.,Ltd All Right Reserved.

症例数設定 ？

What is sample size estimation ?

- 医療機器臨床試験のコンサルティングで最も相談件数が多いのは「症例数の設定」

Many a need of consulting for device clinical trial is sample size estimation (SSE).

- 大きな誤解 misunderstanding of SSE

- 症例数 まず ありきではない

SSE is not first purpose of trials.

- 症例数は試験のデザイン(目的)に従って、
結果の精度を保証するためのもの

The aim of SSE is to assure the reliability of trial conclusion at study design according to study design.

例数設定相談時の問題点

Problems at SSE consultation

- 試験例数だけを求める、試験デザインやエンドポイントは重要視しない

Sample size walks alone, without study design and endpoint

- 比較試験の検定に基づいた方法しか知らないため、間違った例数や、不必要に多い例数が設定されがちである

Ignorance about SSE

You only know test based SSE, but do not know another method.

症例数設定の目的

Aim of SSE

- 試験計画時に臨床試験の主張の証明力の信頼性を保証する手段
- The aim of SSE is to assure the reliability of trial conclusion to demonstrate the claim at study design according to study design.

例数設定をする際には(Preparation) Claim 試験デザイン 仮説の明確化

1. 主張の証明 to demonstrate the claim
2. 効果の明確化 to estimate effectiveness
3. 主要評価項目 to determine endpoint
4. 試験デザイン study design
i.e. randomized control study single
arm open study
5. 仮説の設定 hypothesis
6. 統計手法の決定
to determine the statistical method

試験の目的に応じて3つの例数設定

3 Type of SSE

- 比較検定 test based

***これだけではない

In Japan, a few people know other types of SSE.

- 信頼区間 confidence interval
- ベイズ推定 bayesian inference

例数設定の原理

algorithm of SSE

有効率60% efficacy						
有効数 No. of event	対象 例数 Samples	有効率 %	95%信頼区間 C.I			
3	5	60	16.8	~	100.0	
6	10	60	29.5	~	90.5	
12	20	60	38.4	~	81.6	
60	100	60	50.3	~	69.7	
600	1000	60	56.9	~	63.1	

例数設定の原理

algorithm of SSE

$$SE = \frac{SD}{\sqrt{n}}$$

SE: 標準誤差 パラメータ θ の推定標準偏差 standards error
SD: 標準偏差 標本集団の標準偏差 standards deviation
N : パラメータ推定に用いた例数 number of samples

検定に見る例数設計の原理

SSE based on Test

- 検定(Statistical Test) 臨床試験を実施することで、証明したい差が、ある一定以上の確率で証明できることを保証する。

検定の一般式

Formula of Test

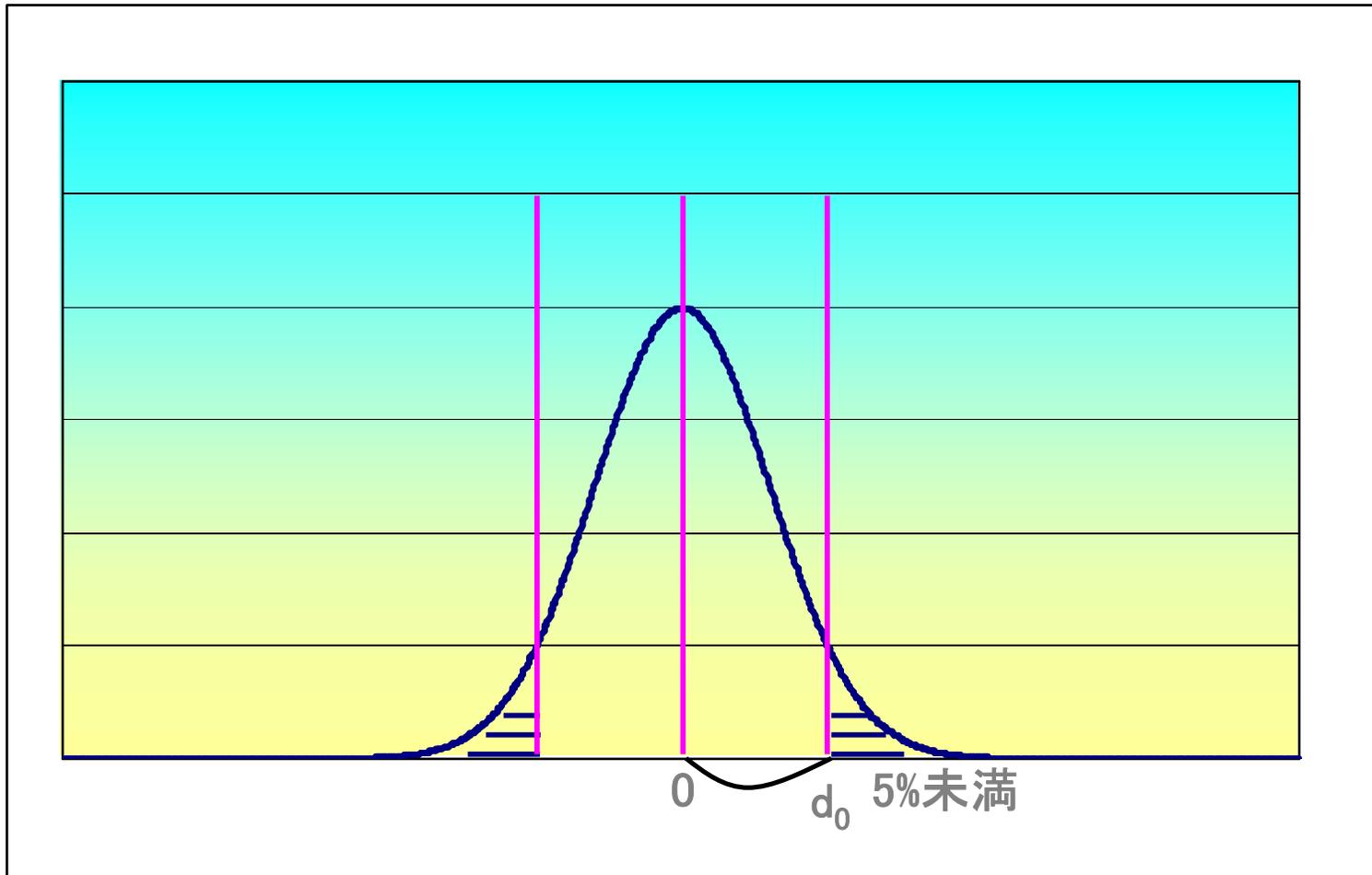
$$Z \leftarrow (\sigma^2, 0) = \frac{\theta}{SE}$$

SE: 標準誤差 パラメータ θ の推定標準偏差 standards error

θ : 検定統計量 Test statistics

Z: 帰無仮説からの正規化距離 standardize distance from 0

検定の原理



H_0

例：正規分布で有意となる $Z_{\alpha/2}=1.96$ の例数

sample size at $Z_{\alpha/2}=1.96$

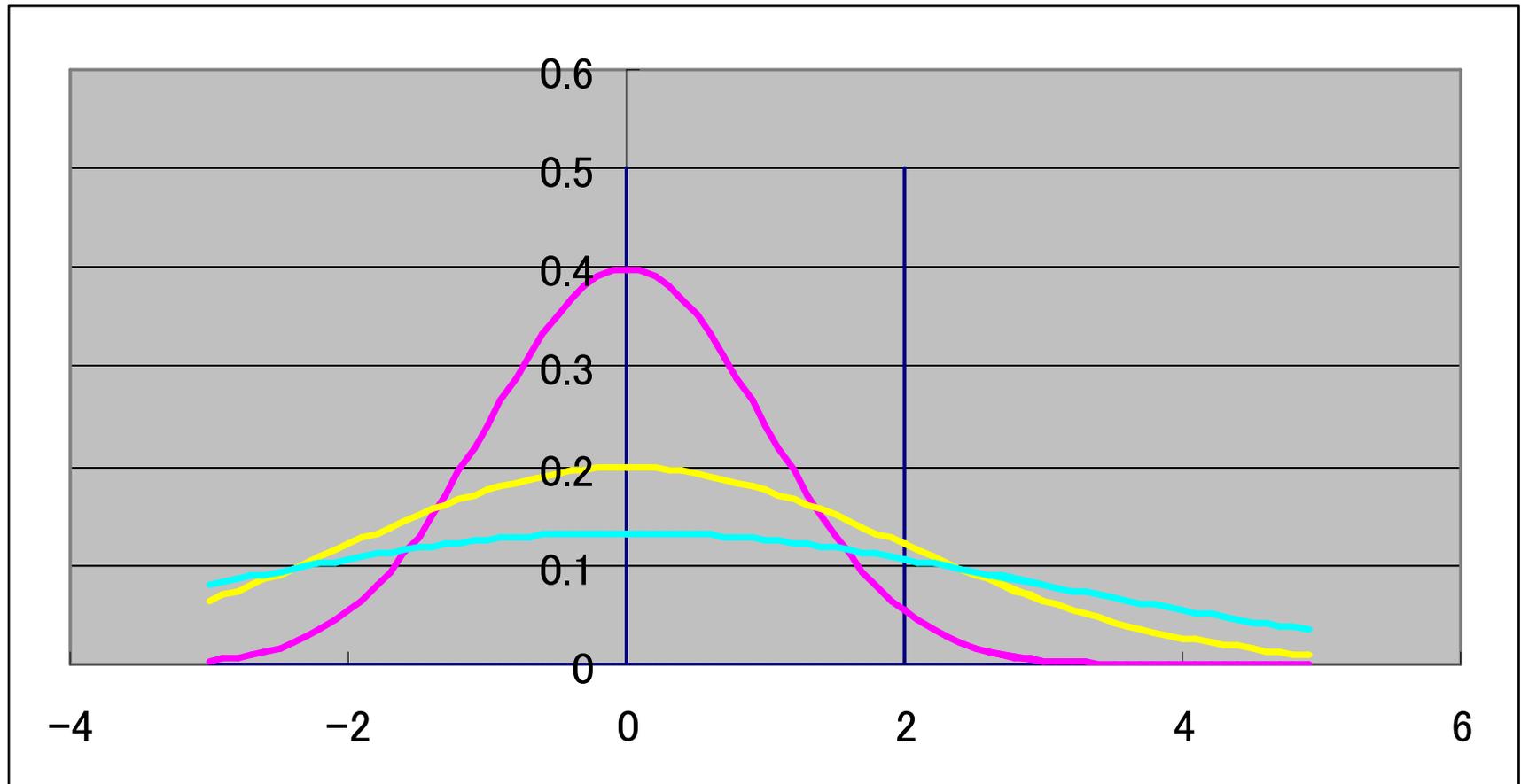
- 検出力を考えなければ if ignore power

$$\frac{d0}{se} = Z_{\alpha/2} \quad \because d0: \text{効果の差}$$

$$\frac{d0}{\sigma/\sqrt{n}} = Z_{\alpha/2} \quad \because SE = \frac{\sigma}{\sqrt{n}}$$

$$n = \left\{ Z_{\alpha/2} \frac{\sigma}{d0} \right\}^2$$

例数が増えればseは小さくなる



まてまて

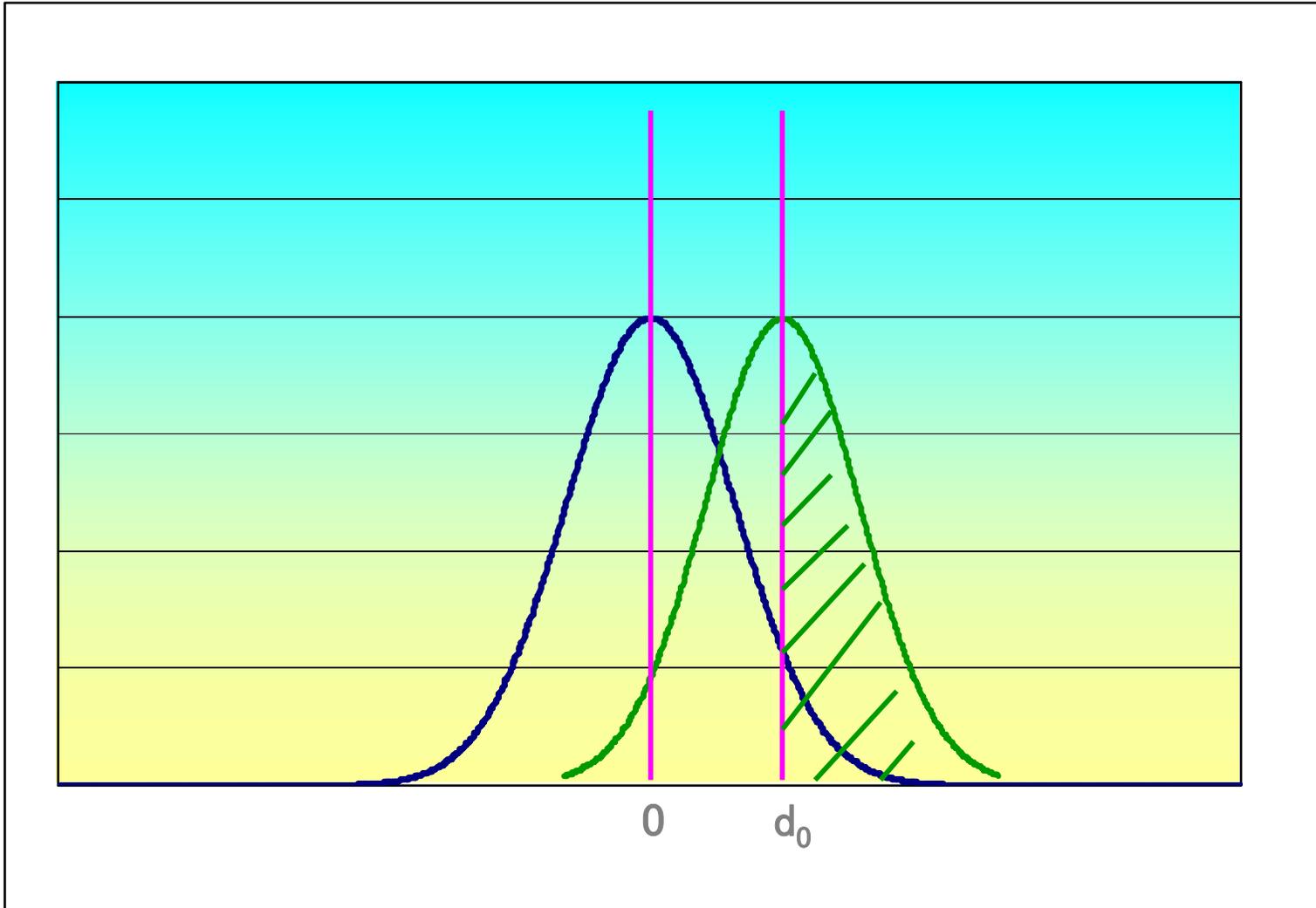
- 臨床的に意味のある差 $\delta =$ 検定で有意な差 d_0

実験を繰り返すと、実験誤差のため

仮に δ が真の差であっても、実験結果の差 d にはばらつきがある

つまり、 $d < d_0$ の結果が観測されると、実験が有意となることは保証できない

検定で有意となるとき



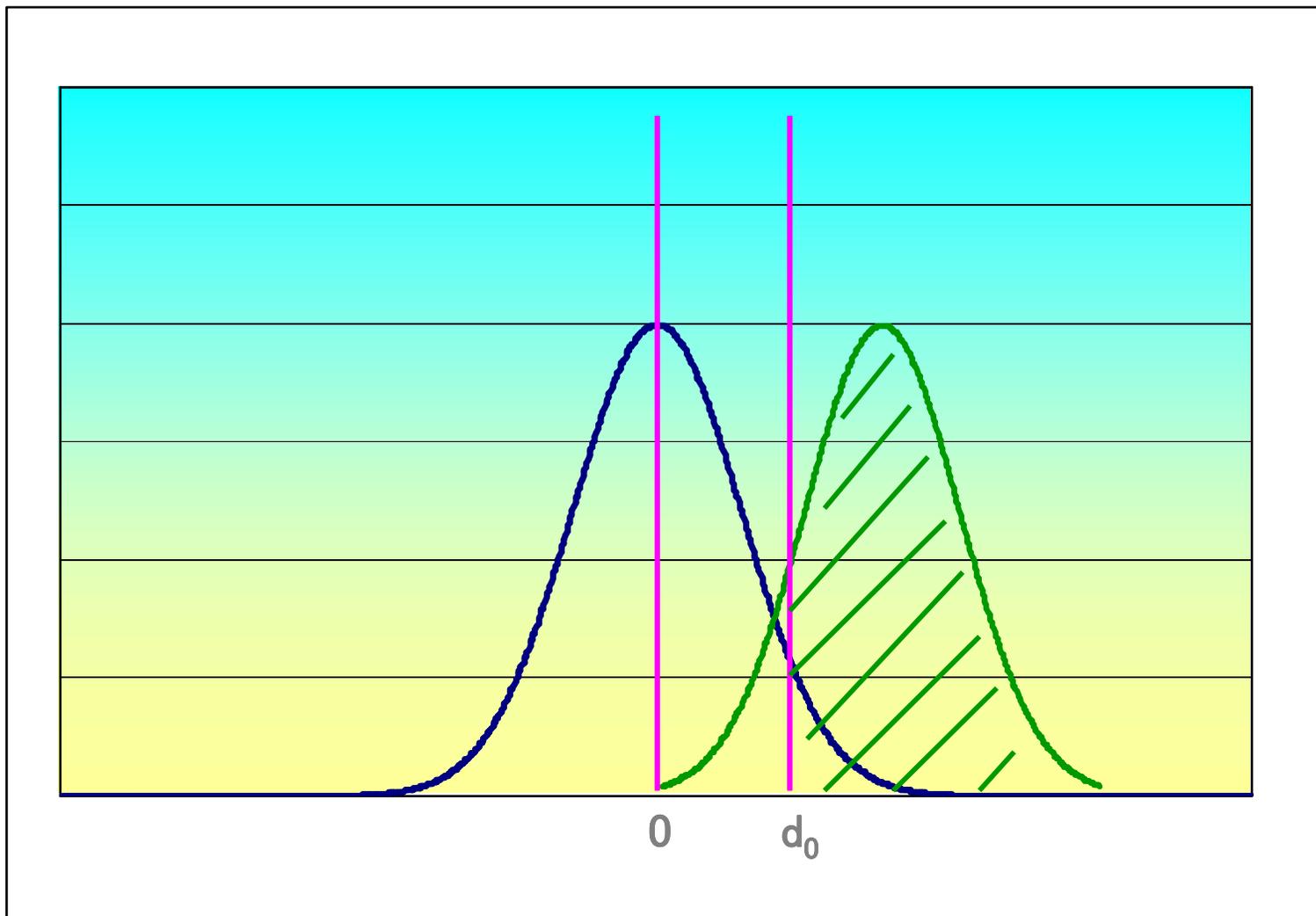
臨床的に意味のある差 $\delta =$ 検定で有意な差 d_0

- この状態(例数)で実験をしたら
- 実験結果が有意となる確率(検出力 $1 - \beta$)

$$1 - \beta = P(d \geq d_0) = 0.5$$

実験をしても、期待した結果が出る確率は
50%か……

例数設計の原理



検出力分だけ距離を広く取る必要

$$\frac{d_0}{se} = z_{\alpha/2} + z_{\beta} \quad \therefore SE = \frac{\sigma}{\sqrt{n}}$$

$$n = \{z_{\alpha/2} + z_{\beta}\}^2 \left\{\frac{\sigma}{d_0}\right\}^2$$

実際に例数を求めてみよう

- 単純な2群 (Device A vs B) 比較
- 2群の例数を等しいとする
- 事象発現率の比較

- $n = 2 \{ (z_{\alpha/2} + z_{\beta}) \cdot \sigma / \delta \}^2$

実際に例数を求めてみよう

$$n = 2 \{ (z_{\alpha/2} + z_{\beta}) \cdot \sigma / \delta \}^2$$

n : 求める1群あたりの例数

$z_{\alpha/2}$: 検定の有意水準 通常は両側0.05

z_{β} : 検出力 通常は 0.8 重要な試験 0.9

δ

σ

δ 臨床的に意義のある差

$$n = 2 \{ (z_{\alpha/2} + z_{\beta}) \cdot \sigma / \delta \}^2$$

δ : 臨床的に意義のある差

例: Device Aは B よりも 臨床的にイベント発生率が 2%は下回ることが望ましい

不明の場合: パイロット試験(それ以前の相)の試験結果から推定

σ 標本集団の標準偏差

$$n = 2 \{ (z_{\alpha/2} + z_{\beta}) \cdot \sigma / \delta \}^2$$

σ : 標本集団の標準偏差

- ・パイロット試験(それ以前の相)の試験結果から推定
- ・参照文献から推定

実際の例数の例

$1 - \beta = 0.80$

σ	δ		
	0.20	0.25	0.30
1.5	73	53	42
2	112	80	62
2.5	160	112	85

$1 - \beta = 0.90$

σ	δ		
	0.20	0.25	0.30
1.5	138	97	74
2	223	153	115
2.5	328	223	164

記載例

- 本評価項目はstent挿入後のLate Lossである。
- 新製品の性能は従来品に比して 0.25mm以上のLate Lossの減少が望まれている。
- 文献、A,B,CからLate Lossの標準偏差は2.0であると仮定した。
- **主要評価項目Late Lossの従来品と比した本製品に期待される臨床的有意差を 0.25mm、標準偏差2.0mmと想定した場合の、Student t 検定による結果を、有意水準両側0.05、検出力0.80で保証するのに必要な例数は、1群 あたり80例となる。**

比較試験の例数設定の特徴

- 検出したい差 δ が 相対的に標本の標準偏差 σ より 大きくなれば 例数は小
- 検出力(実験によって有意差が観察されるであろう確率)を大きくしようとすると 例数は大
- 検定の方式によって、例数算出の式は異なる

医療機器の臨床試験の例数設定に関する問題

- Single arm 試験が多く、比較試験の検定に基づく例数設計は使えない
(でも、現実には、無理矢理使用している例を見かけるが)

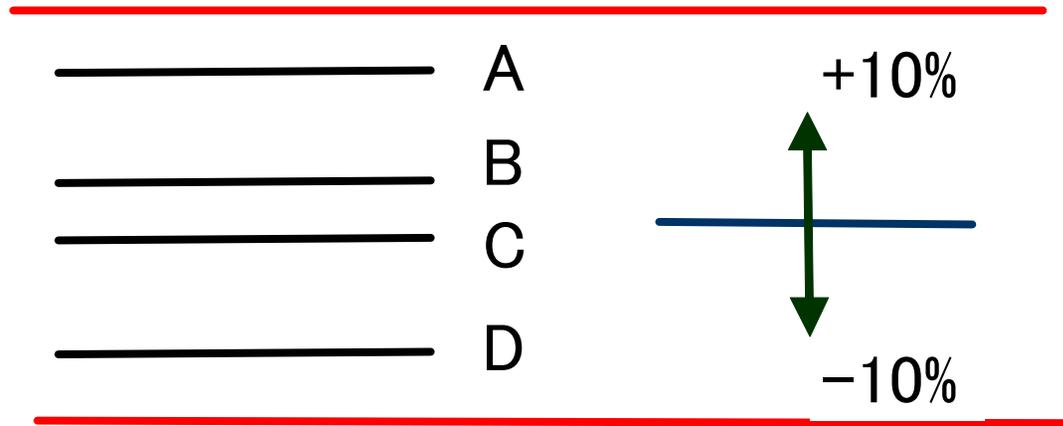
試験の目的に応じ、以下の2つを活用

- 信頼区間 confidence interval
- ベイズ推定 bayesian inference

シングル アーム試験

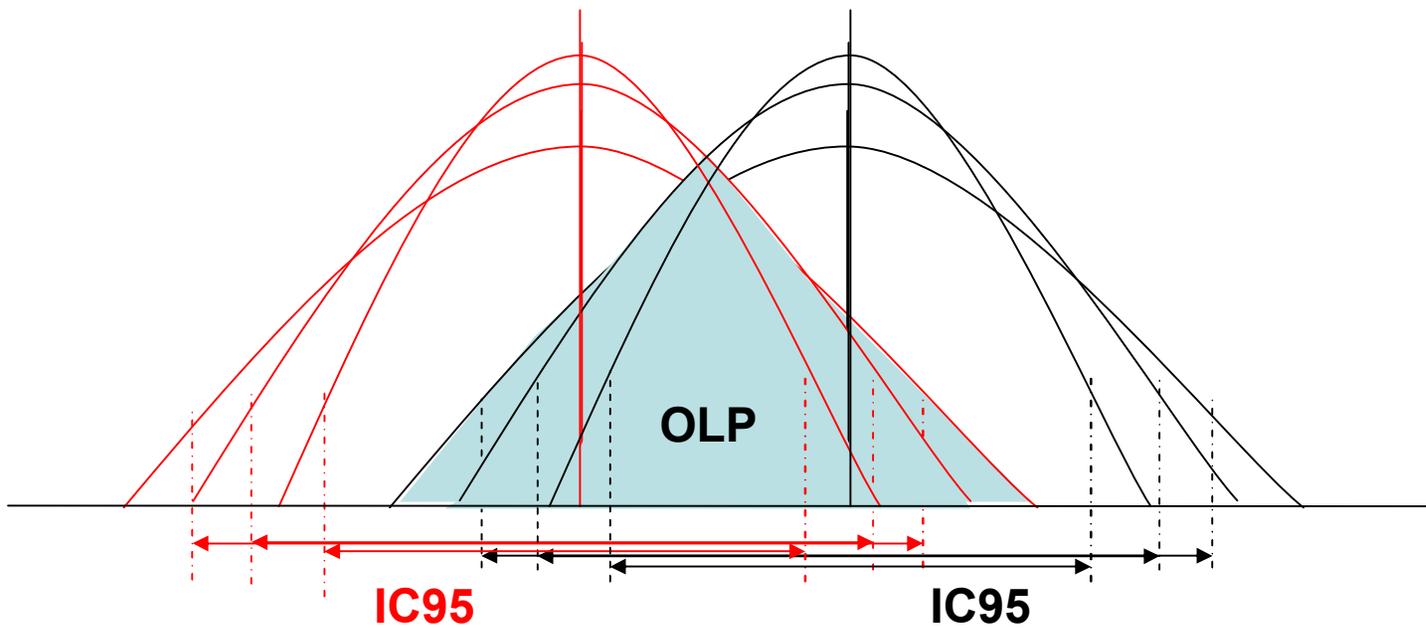
- 例1 : 試験結果を過去の試験成績と比較
- 有効率が、過去の試験成績と同等であることを示す
- 試験結果の有効率 過去の試験成績の中に入る

有効率



Single arm 試験

- 例2: Historical Controlとの比較
- 検定ベース？
例: 非劣性試験
- Historical Controlとの類似性の検討
評価項目推定値の重なり overlapping
例: 本試験結果は、Historical Controlと比較して、
75%以上の確率で、優れていると推測される。



OLP=Over Lapping Parameter

信頼区間の例数設定

SSE based on confidence interval

$$SE = \frac{SD}{\sqrt{n}}$$

$$\text{求める信頼区間幅} = 1.96SE = \frac{SD}{\sqrt{n}} 1.96$$

$$n = \left\{ \frac{1.96SD}{\text{求める信頼区間幅}} \right\}^2$$

シングル アーム試験

有効率 (π)	信頼区間幅 (ε)					
	10	12	13	14	15	20
75	72	50	43	37	32	18
60	92	64	55	47	41	23
50	96	67	57	49	43	24

逐次的なFeasibility試験の場合

- 試験結果が逐次的に蓄積 Stepwise Studys
- 大きな試験例数は難しい
difficulty of large sample size study
- 過去の試験結果を利用して、今回の試験結果と合わせて評価

To use a prior information to derive for a posterior conclusion.

ベイズ推定 bayesian inference

信頼区間の例数設計例

- 本製品に類似する既認可ステントの臨床試験のデータベースによると
- 12カ月のMACEの成績は、メタアナリシス平均8.4% 範囲は[6.4~10.4]であった。
- 本製品の予想MACEが従来品と変わらないと想定すると、範囲 6.4-10.4の間に少なくとも試験100回のうち、95回はこの範囲に入るような試験を計画したい。
- そこで、**予想MACEを 8.4%とし、2項分布のもと、この期待値の95%信頼区間が +2.0%に収まるのに必要な例数は 86例となる。**

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

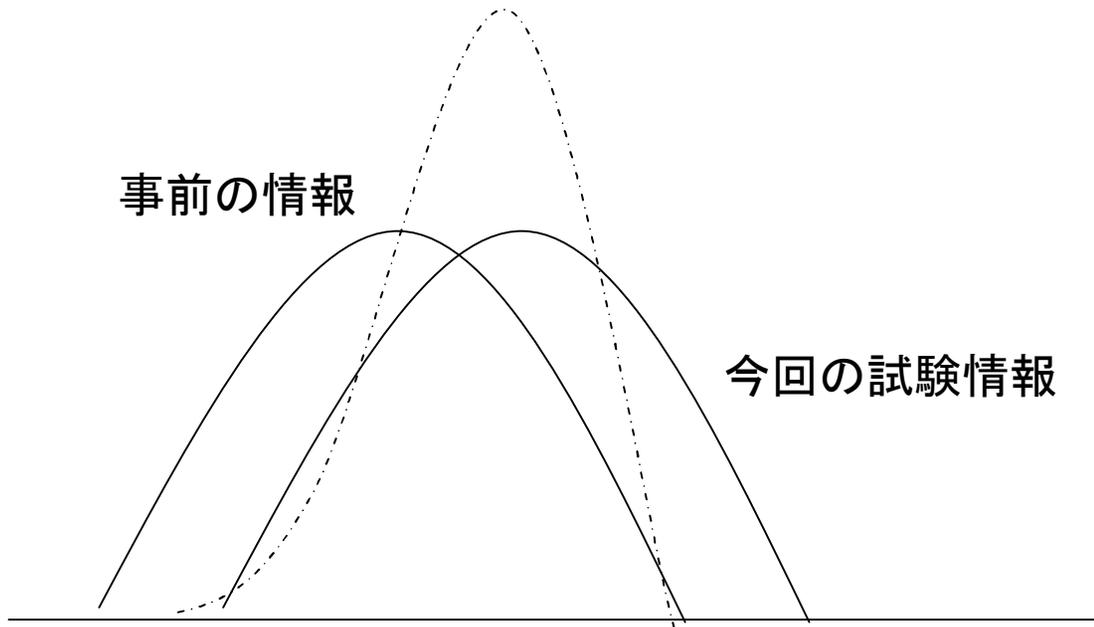
$P(\theta | y)$: 事後確率: 前回までの情報に

今回の試験結果を加味した θ の確率に関する情報

逐次的に推測した情報

事前の情報

今回の試験情報



Bayesの定理の報告例

- 本製品の事前の試験Aにおける有効率は20%であった。目標とする有効率15%よりも本製品が優れている確率は0.55であった。

新たに本試験により症例を3例追加した結果、bayes推定有効率は22%、目標有効率15%よりも本製品が優れている事後確率は0.85であった。

事後確率が試験設定時の目標0.80を上回ったので、本製品は、目標有効率15%よりも優れていると結論した。

信頼区間 confidence interval ベイズ推定 bayesian inference に基づく例数設定の利点

利点 Pros

- ① 柔軟な試験例数に対応できる。
flexible SSE

注意点 Cons

- ① 結論を導く論理性が非常に重要
The logic is very important to derive the conclusion.
- ① 利用した統計モデルの妥当性の証明が必要(特にベイズ推定では)
To assure the validity of Bayesian Model.

結論 Conclusion

1. 例数設計は、試験の目的に応じた試験デザインのもとで、試験の結論の精度を保証するために実施するもの

The aim of SSE is to assure the reliability of trial conclusion at study design according to study design.

2. 決して、まず例数ありき あるいは、例数だけ単独で設定するものではない。

SSE is not first purpose of trials and determined alone.

結論 Conclusion

3. 現実には、比較試験の検定原理に基づく例数設定のみが利用される傾向にあり、試験デザインを保証するものではないことが多い

Not only test based SSE, but also confidence interval based and bayesian inference based SSE are very useful.

4. 信頼区間 confidence interval
ベイズ推定 bayesian inference

を利用すれば、より合理的な例数設定は可能

結論 Conclusion

5. 最も重要なことは、信頼のできる統計家と、試験デザインの検討段階から相談をしながら進めること

Most importance is to have a team including an able statistician.

重要なのは能力とコミュニケーション

We need to have will, vision and communication.