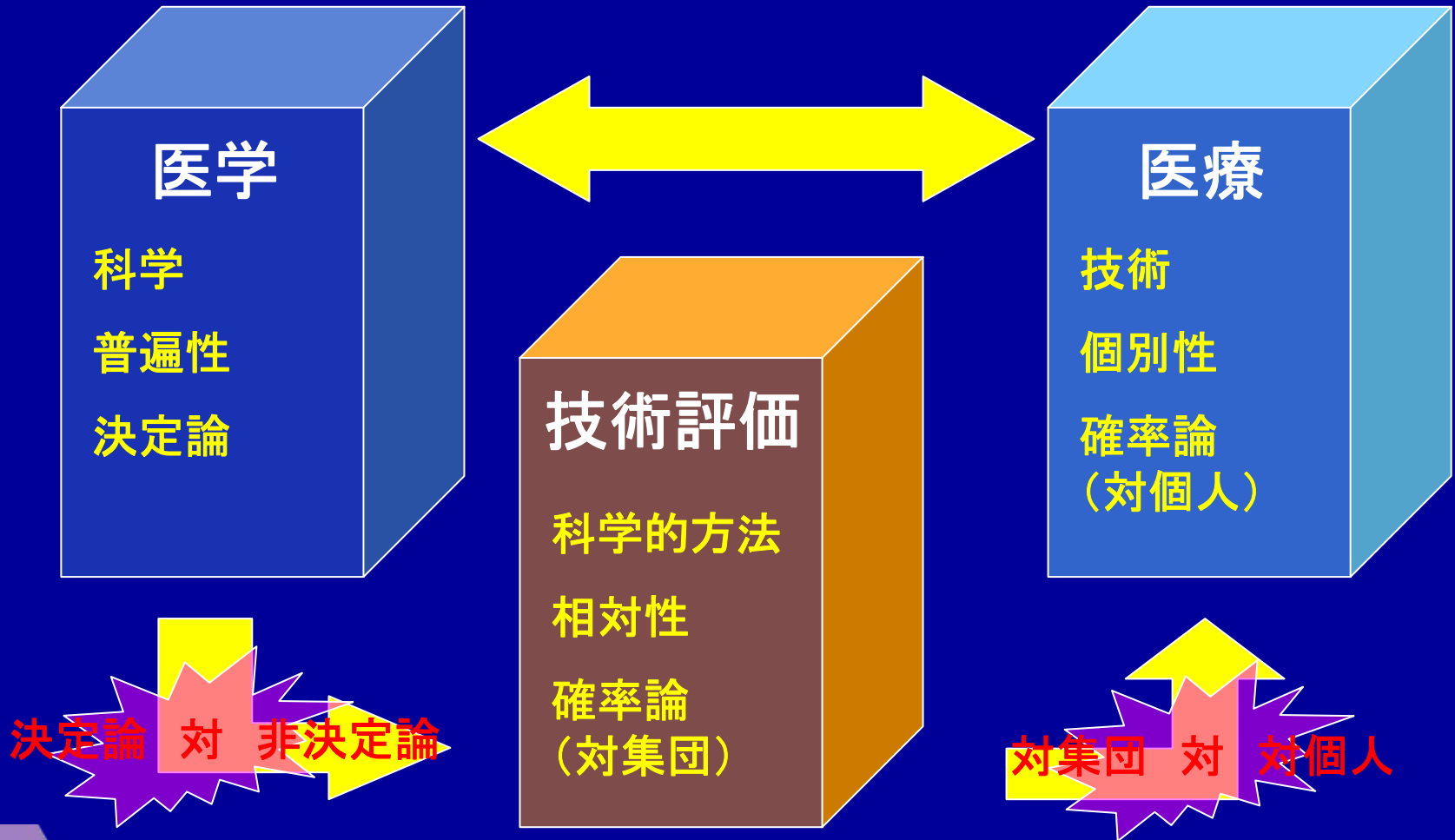


臨床試験デザイン

医学・医療と技術評価





Ronald A. Fisher
(1890-1962)

科学の実験と技術の実験

- 科学の実験
 - 人工的に作り出された純粋な条件のもとでの因果関係を確定することを目的とする
 - 科学的法則の法則性を検証する
- 技術の実験（統計的実験） by R.A.フィッシャー
 - 現実の場での因果関係を確かめることを目的とする
 - 工学、農学、医学等の現実の応用において、何をなすべきかを知る

竹内啓著「統計学的な考え方」、岩波書店、2003

臨床試験の特徴

1. 試験の場は、**臨床現場**に近い状況に設定される
2. 結果の分析には**誤差**(**個体**、**評価のバラツキ**など)の存在を前提にしなければならない
3. いくつかの**治療法**を**組み合わせ**て結果を見る必要があることがある
4. 目的は、何らかの基準によって現実の場において最も良い結果が得られるような**治療法**を求めることである

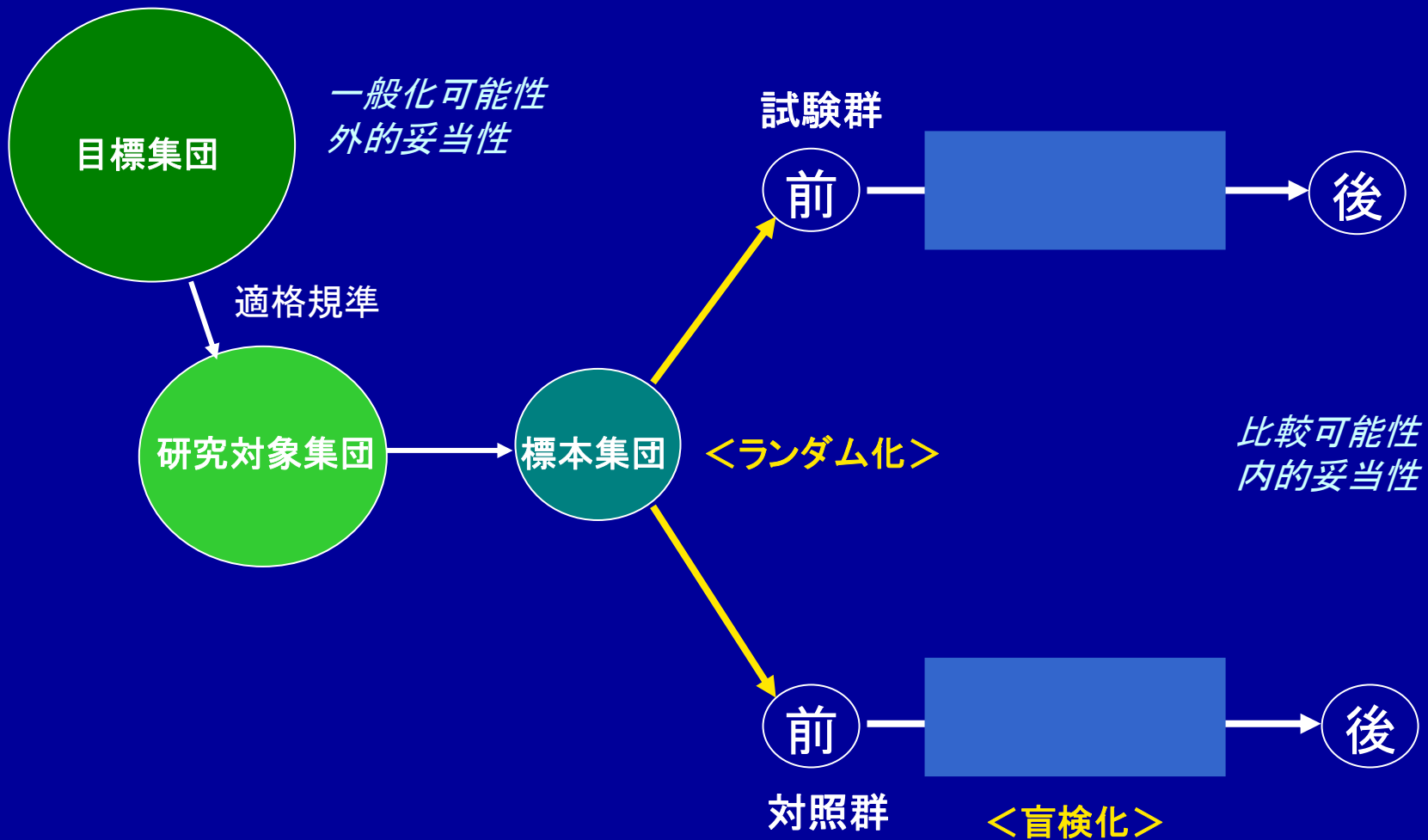
臨床試験(技術的実験)

- 理論よりもデータ
- 結果よりもプロセス
- 基礎研究(科学的実験)の延長にはない

統計学の役割

プロセス(計画～データ～検証・推測)の保証

ランダム化対照試験



登録・割付

- 登録
 - 登録とは、(通常、第三者が)参加者の意思を確認して、参加を許可すること
- 割付
 - 割付とは、各被験者の受けるプロトコル治療を(通常、ランダム化により)決定すること

客観性と独立性→登録センター(データセンター)

無作為化？ or ランダム化？

- **Randomize the people**
 - Using a method that gives them all an equal chance of being chosen
- 無作為
 - 作為を加えないで行うこと
 - 任意であること
- random variableの訳は確率変数
 - 確率化では？
- ちなみに、隨機化 in China

ランダム化の歴史

- Fisher RA, et al. (1923)
 - 農事試験に初めて適用
- Amberson JB, et al. (1931)
 - 臨床試験に初めて適用
- British Medical Research Council (1948)
 - 結核に対するストレプトマイシンのランダム化試験を発表

Fisherの3原則

－ 実験計画法 －

- 反復 (replication)
 - － 偶然誤差を評価するため
- ランダム化 (randomization)
 - － 系統誤差を偶然誤差に転化するため
- 局所管理 (local control)
 - － 大きな系統誤差を実験の誤差から除去するため

実験に伴う2種類の誤差

- 偶然誤差 (random error)
 - 測定誤差のようにある確率分布に従うと想定できる誤差
 - 繰り返し測定を行えば推定可能
- 系統誤差 (systematic error)
 - 圃場の肥沃度や日当たりの不均一性のように確率変数と見なせない誤差
 - 繰り返しには関係なく結果を歪める原因

臨床試験における良いランダム化の方法は？

- 予後因子の分布が類似したグループを作る
 - 偶然バイアス (accidental bias)を小さくする
- 次に試験に登録される被験者の治療を医師が予見できないようにする
 - 選択バイアス (selection bias)を小さくする

固定型割付

- 層別置換ブロックランダム化 -

施設1	軽症	BAAB BABA AABB.....
	重症	<u>ABBA</u> ABAB BBAA.....
施設2	軽症
	重症
施設3	軽症
	重症
.....

└───┘
ブロック (サイズ4)

適応型割付

- 最小化法 -

因子	カテゴリー	A群	B群	次の被験者
施設	1	2	1	
	2	1	3	
	3	1	0	
	4	3	4	
	5	2	2	←

重症度	軽症	8	7	
	重症	7	8	←

A群の和: $2+7=9$ < B群の和: $2+8=10$



次の被験者はA群に割り付ける

盲検化

- 治療内容を知ることによるバイアスの防止策
 - 特にエンドポイントへの影響が重要
- オープンラベル
- 単盲検(被験者)
- 二重盲検(被験者、医師)
- 三重盲検(被験者、医師、スポンサー)

エンドポイントとは？

- 試験の目的に関連する仮説を検証するうえで臨床的に意味があり、客観的に評価できる観察・検査項目またはそれらの合成指標
 - 各被験者について定義されるものであり、「発生割合」や「有効割合」など、集団について定義される指標ではない
 - 主要(primary)、副次(secondary)に分ける

主要・副次エンドポイント

- 主要エンドポイント (primary endpoint)
 - その試験で最も重要な1つのエンドポイント
 - 2つ以上設定した場合は、検定の多重性を考慮する
- 副次エンドポイント (secondary endpoint)
 - 主要エンドポイント以外のエンドポイント

エンドポイントの型

- 連続型
 - 単回測定、繰り返し測定
- 分類型(二値)
- 分類型(順序あり)
- 分類型(順序なし)
- 回数型
- 時間ーイベント型

連続型

- 多くの臨床検査値、身体的特徴など
- 位置(平均など)とばらつき(分散など)で特徴付けられる分布を持つ
- 通常は集団での平均を評価する
- 統計手法は、t検定、Wilcoxon検定、分散分析など
 - 交絡因子の効果を調整するときには、共分散分析、重回帰分析など

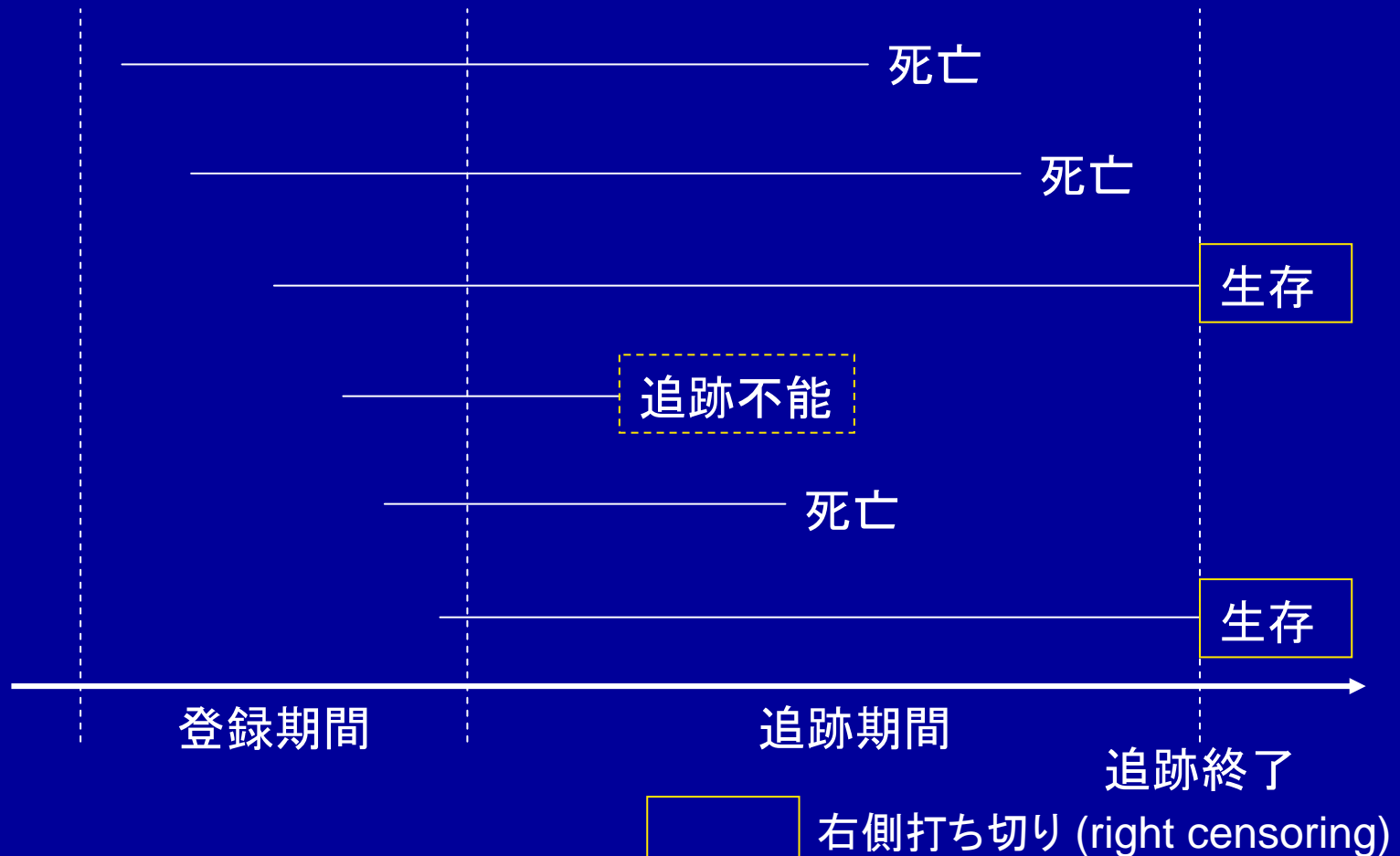
二値分類型

- 腫瘍反応、感染の有無など
- 通常は集団での割合（成功あるいは失敗）を評価する
- オッズ、オッズ比を評価することもある
- 統計手法は、 χ^2 検定、Fisher正確検定など
 - 交絡因子の効果を調整するときには、logistic回帰分析など

時間－イベント型

- 治療開始、診断などから重要な臨床イベントまでの時間
 - 例えば、生存時間、退院までの時間など
- 連続型との決定的な違いは、打ち切り (censoring) の存在

時間-イベント型



時間-イベント型

- 通常、集団での生存時間分布の代表値（中央値、5年生存確率など）を評価する
- ハザード率、ハザード比も有用
- 打ち切りの発生がイベントと独立という仮定が必要
- 生存曲線の推定はKaplan-Meier法、検定はlogrank検定、交絡因子の効果を調整するときは、Cox回帰分析など

例：進行乳癌に対するランダム化試験

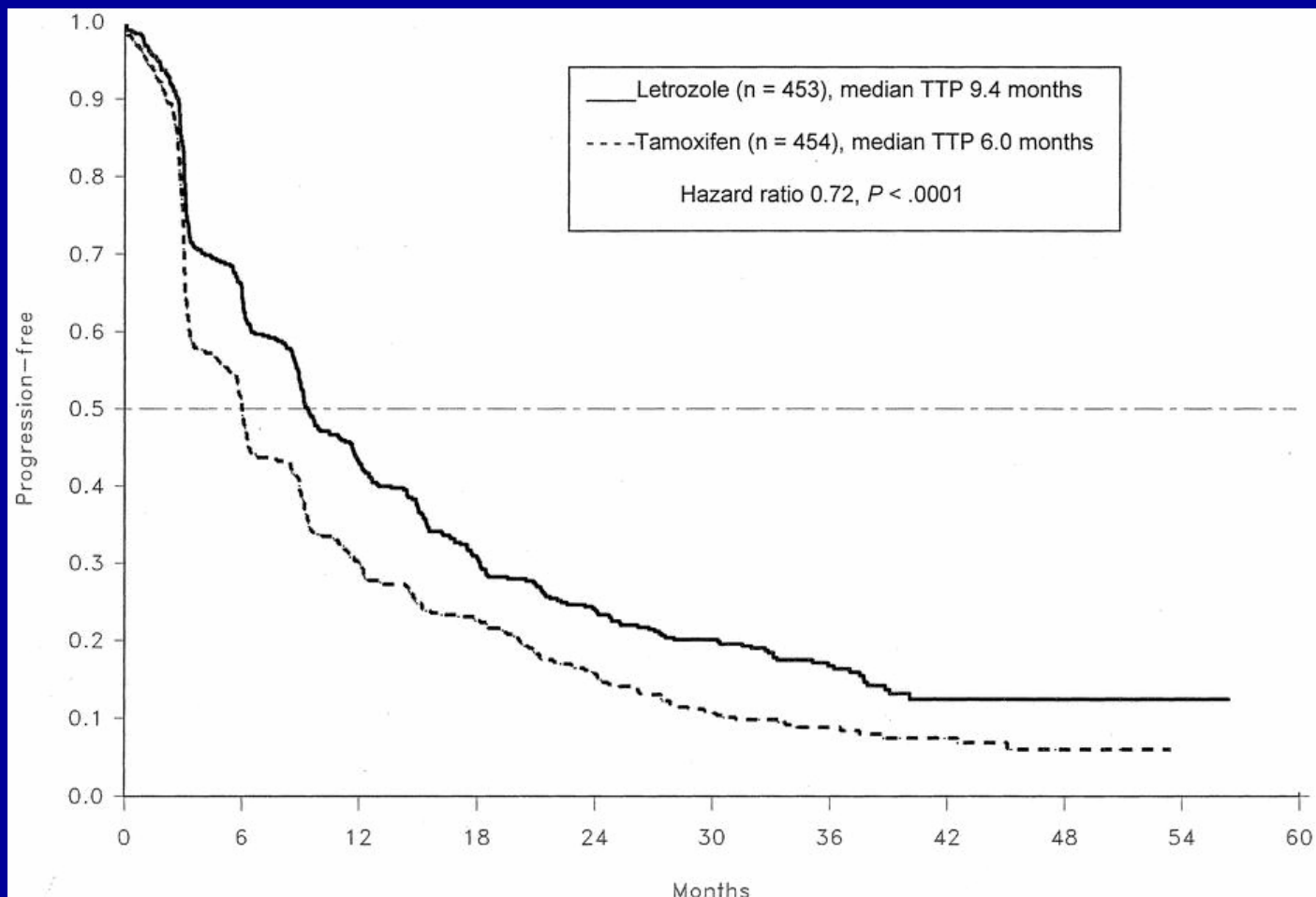
- 主要エンドポイント

- 進行までの期間 (TTP : time to progression)

- ランダム化の日から原病の進行が最初に認められた日までの期間
- 進行とは、
 - 測定可能病変の25%以上の増大、測定不能病変の同程度の増大、または新病変の出現
 - 乳癌の悪化に伴う治療中止、または治療期間中あるいは治療中止後6週以内の乳癌による死亡

Mouridsen H, et al. JCO 2003;21:2101

例：進行乳癌に対するランダム化試験



Mouridsen H, et al. JCO 2003;21:2101

例：進行乳癌に対するランダム化試験

- 副次エンドポイント

- 治療成功期間 (TTF : time to treatment failure)
- 反応までの期間 (time to response)
- 全生存期間 (OS : overall survival)
- 腫瘍反応 (ORR : overall tumor response rate)
- 臨床的利益 (rate of clinical benefit)
- 安全性 (safety)

Mouridsen H, et al. JCO 2003;21:2101

例：進行乳癌に対するランダム化試験

- 反応 (response)
 - 3か月後の再評価で完全反応 (CR: complete response) または部分反応 (PR: partial response) が確定されること
 - 完全反応は腫瘍の完全な消失、部分反応は腫瘍の50%以上の縮小 (WHO規準)
 - 注: 最近ではRECIST (response evaluation criteria in solid tumor) ガイドラインが標準

Mouridsen H, et al. JCO 2003;21:2101

真・代替エンドポイント

- 真のエンドポイント (true endpoint)
 - 臨床イベントとほぼ同義
- 代替エンドポイント (surrogate endpoint)
 - 臨床試験の計画や解析において臨床イベントの代わりに使われる検査、計測、スコア等の変数である。その変数が関心のある臨床イベントと相関があると信じられており、治療差を検出するのに簡便であるので用いられる。(Meinert CL, 1986)

代替エンドポイントの例

代替エンドポイント	真のエンドポイント
血中コレステロール値	心疾患の発生
血圧値	心疾患、脳血管障害の発生
血糖値	糖尿病合併症の発生
腫瘍縮小効果	癌死
不整脈	心臓死
心機能指標	心不全の発生
血管狭窄度	狭心症、心筋梗塞
CD4リンパ球数	AIDS死
眼圧	緑内障の視野狭窄

CAST

- Cardiac Arrhythmia Suppression Trial -

- 目的
 - 抗不整脈薬の心筋梗塞後の突然死予防効果の検証
- デザイン: 二重盲検無作為化
 - encainide: 実薬432例 vs. プラセボ425例
 - flecainide: 実薬323例 vs. プラセボ318例
- 結果: 早期中止
 - 不整脈死または非致死的心停止
 - 実薬 4.5% vs. プラセボ 1.2%
 - 死亡率
 - 実薬 7.7% vs. プラセボ 3.0%

統計学的考察

- 目標症例数の設定根拠
- 解析対象集団
- 解析項目・方法
- 中間解析
 - 試験治療の有効性・安全性の観点から試験の早期中止の必要性を判断する際に、その根拠を得るために実施する統計解析。

症例数 (標本の大きさ、sample size)

- 科学性に重きを置くと、有意であるという結果を得やすくするために症例数を多くしたくなる。
- 一方、倫理性に重きを置くと、試験の対象となる患者数を可能な限り減らしたくなる。
 - しかしながら、症例数不足の試験を行って統計的に有意でないという結果が得られた場合に、真に効果がなかったのか、それとも真に効果があったにもかかわらず検出力不足のため有意でなかったのかを区別できないことは、無駄な試験を行ったという意味で非倫理的と言うこともできる。
- 多すぎても少なすぎてもいけない。

一般論

- 一時的な仮定に基づく概算である
- ある仮定の下で、
 - 何人の対象者が必要か？（症例数算出）
 - n人の対象者しか参加できない場合、実施する価値があるか？（検出力解析）
- 仮説検定という方法を用いると、定式化が容易である

仮説検定

- 仮説の設定
 - 帰無仮説 (null hypothesis) : H_0
 - 対立仮説 (alternative hypothesis): H_1
- 検定統計量の選択
 - 通常は、S/N比(シグナルとノイズの比)
- 統計的有意性の評価
 - データと帰無仮説の間の整合性
 - p値と有意水準を比較(帰無仮説を棄却/採択)

仮説検定における2種類の誤り

検定結果	母集団 (真)	
	差なし	差あり
採択 (有意差なし)	正しい	β エラー (type II error)
棄却 (有意差あり)	α エラー (type I error)	正しい

↓
1- β : 検出力 (power)

臨床試験のインパクト

- 消費者リスク
 - 効果のない、あるいは有害な医療技術が承認あるいは使用されるリスク
 - 第1種の過誤 (α)
- 生産者リスク
 - 効果がある医療技術が承認あるいは使用されないリスク
 - 第2種の過誤 (β)

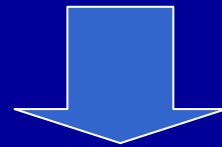
発生割合の比較

群	イベント発生		計
	有	無	
治療群	a	b	n_1
対照群	c	d	n_0
計	m_1	m_0	n

$$\text{治療間差} = \frac{a}{n_1} (\text{治療群の発生割合}) - \frac{c}{n_0} (\text{対照群の発生割合})$$

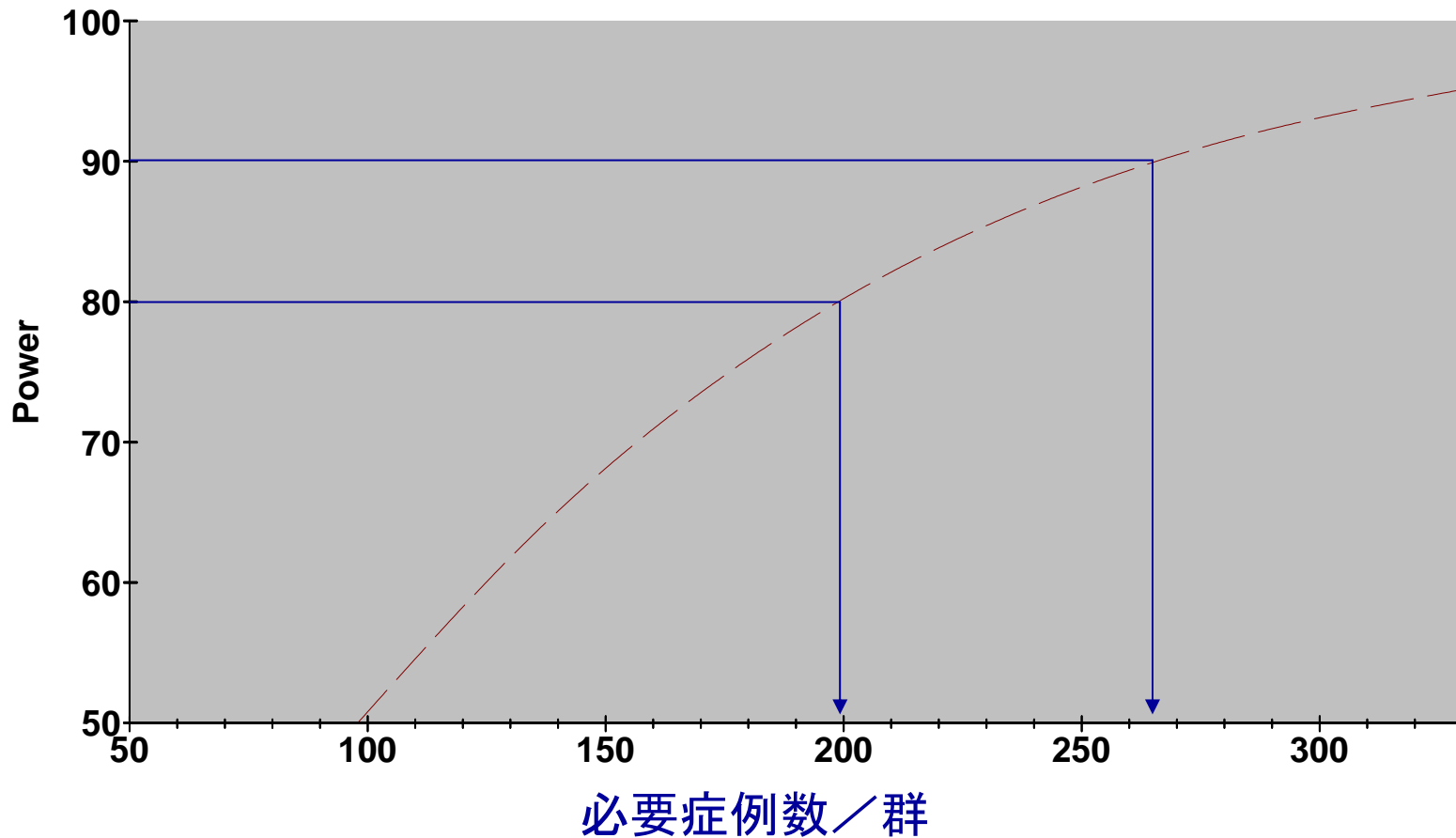
例：急性心筋梗塞後の β 遮断薬

- エンドポイント：発作後90日以内の死亡
- 対照群の死亡割合を0.20と予測
- 治療群の死亡割合が0.10になれば、臨床的に意味があると仮定
- $\alpha = 0.05$ (両側)



n(必要症例数) と $1 - \beta$ (検出力) との関係？

必要症例数と検出力の関係



解析対象集団

- Intention-to-treat (ITT) の原則
 - プロトコルに従って治療した(できた)群ではなく、治療しようとした群を対象
 - 全登録(割付)例を解析対象とする。
 - 除外が許されるのは、不適格、全く治療せず、など。
 - 製品を研究的に評価するのではなく、コンプライアンスなども含めた実践的な技術の評価
 - プロトコル治療の不遵守などが多く発生すると効果が薄まり、差が検出困難。

統計手法の例

目的	連続変数	分類変数	時間-イベント変数
分布の記述	ヒストグラム、 箱ヒゲ図、散布図	ヒストグラム、 分割表	生存曲線 (Kaplan-Meier法)
要約統計量	平均、分散、中央 値、パーセント点、 相関係数	頻度、一致度、 相関係数	x年生存割合、 中央生存期間
検定(単純)	t検定、分散分析、 Wilcoxon検定	χ^2 検定、 Fisher正確検定	logrank検定
検定(層別)	分散分析	Mantel-Haenszel 検定	層別logrank検定
回帰モデル	分散分析、 重回帰分析	logistic回帰分析	Cox回帰分析

同じだけの時間と労力をかけたとしてもデータ収集の過程、または実験計画を厳密に検討しているか否かによって、得られる収穫は10倍から12倍にもなる。

実験終了後に統計学者に相談を持ちかけるのは、統計学者に、単に死後診療を行って下さいと頼むようなものである。統計学者はおそらく何が原因で実験が失敗したかという実験の死因について意見を述べてくれるだけであろう。

R. A. Fisher (1938)

参考文献

(臨床試験デザイン)

- これからの臨床試験. 椿広計、藤田利治、佐藤俊哉編. 朝倉書店、1999.
- 無作為化臨床試験. 丹後俊郎. 朝倉書店、2003.
- 米国SWOGに学ぶがん臨床試験の実践. Green S, Benedetti J, Crowley J. 福田治彦、新美三由紀、石塚直樹訳. 医学書院、2004.
- 臨床試験ハンドブック. 丹後俊郎、上坂浩之編. 朝倉書店、2006.
- Fundamentals of Clinical Trials. Friedman LM, Furberg CD, DeMets DL. Springer, 1998.
- Design and Analysis of Clinical Trials. Chow SC, Liu P. Wiley, 2004.

参考文献

(統計学に関する読み物)

- 統計学とは何か. Rao CR. 藤越康祝、柳井晴夫、田栗正章訳. 丸善、1993.
- R.A.フィッシャーの統計理論. 芝村良. 九州大学出版会、2004.
- 確率と統計のパラドックス. Senn S. 松浦俊輔訳. 青土社、2005.

(医療統計学の基礎)

- 医学統計学ハンドブック. 宮原英夫、丹後俊郎編. 朝倉書店、1995.
- はじめて学ぶ医療統計学. Swinscow TDV, Campbell MJ. 折笠秀樹監訳. 総合医学社、2003.
- 実践医学統計学. Matthews DE, Farewell VT. 宮原英夫、折笠秀樹監訳. 小田英世、手良向聡、森田智視訳. 朝倉書店、2005.