臨床試験に必要な統計的考え方

京都大学医学部附属病院
探索医療センター検証部

吉村健一

30 min. + 質疑 10 min.

吉村　健一　　京都大学

本日の内容

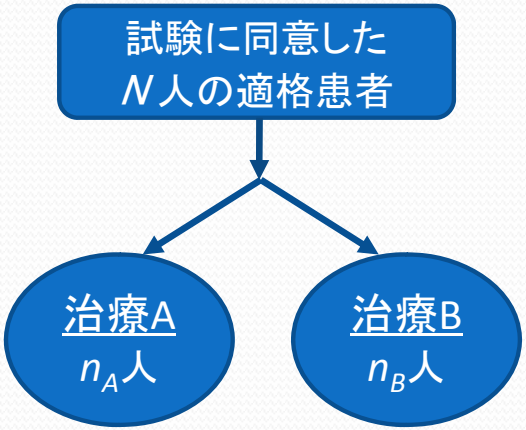
- **ランダム化　randomization**
- 多重性　multiplicity

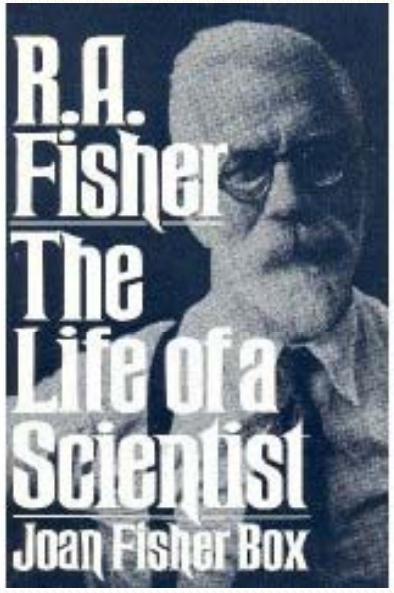
吉村　健一　　京都大学

ランダム化をおこなう意義

① 比較可能性

② P値の算出根拠





Box (1978)

吉村　健一　　京都大学

Association vs. Causation

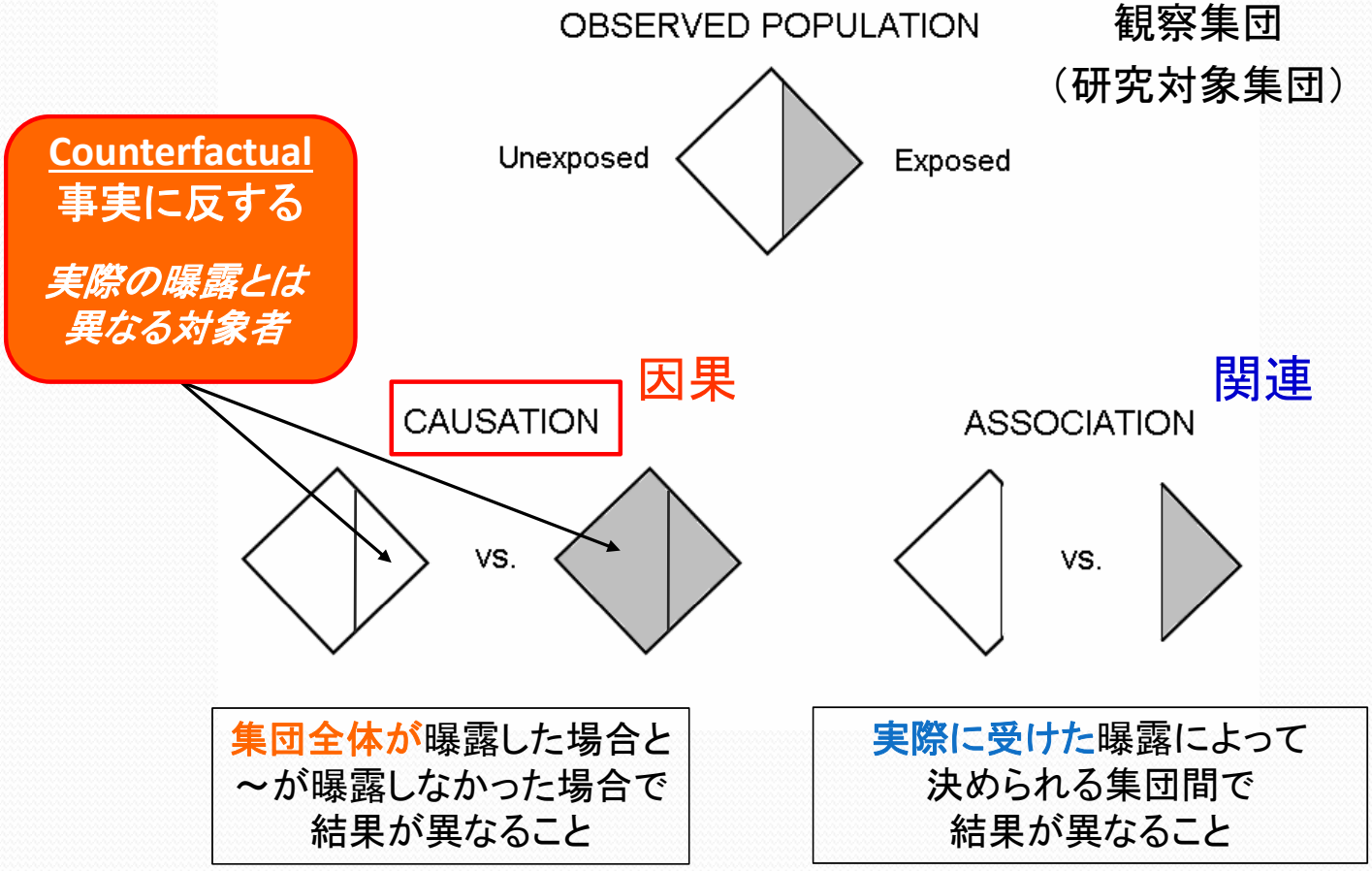
- **関連**　　**Association**
  - **実際に受けた曝露**で決められる集団間、
    つまり、曝露有群と無群で結果が異なること
    - 交絡による“関連”も含む

- **因果（原因と結果）**　　**Causation**
  - 「**集団全体が曝露**した場合」と
    「　　～　　が曝露しなかった場合」
    で結果が異なること

4

吉村　健一　　京都大学

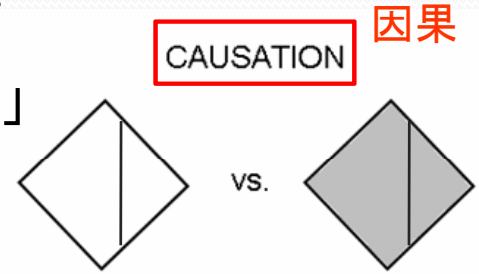
Association vs. Causation, *cont.*



吉村　健一　　京都大学

ランダム化の意義：① 比較可能性

- 理想的なコントロールは自分自身
  ＝同じ予後をもつコントロール・グループ
  - 「このコントロールは比較可能性がある」という



- 比較可能性のあるコントロール・グループ
  - 自分自身をコントロールとすることは事実に反するため実際上は不可能
  - **ランダム化のみによって無条件で実現可能**
    - 既知の交絡因子に限らず、未知の交絡因子に関しても

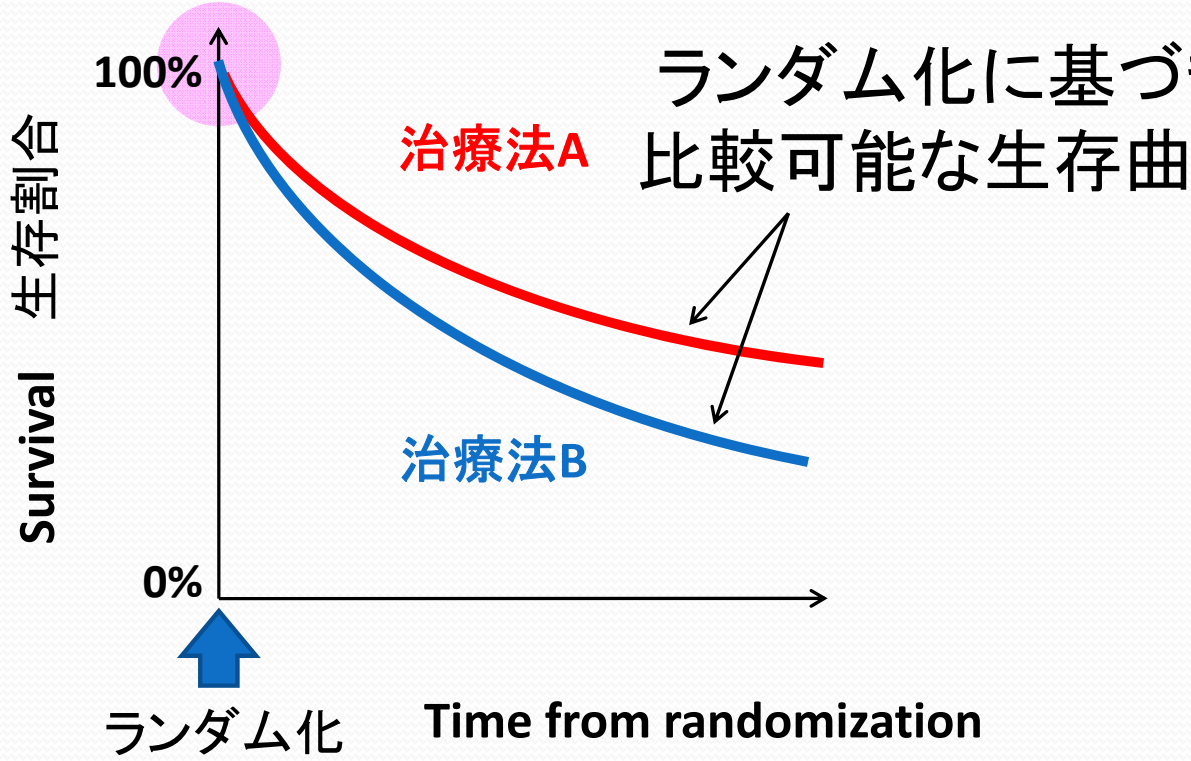
ランダム化によってのみ
“無条件”で
群間が比較可能になる

7

吉村　健一　　京都大学

ランダム化試験の結果の一例：生存曲線

- 生存割合を表す曲線：一般にKaplan-Meier法によって推定

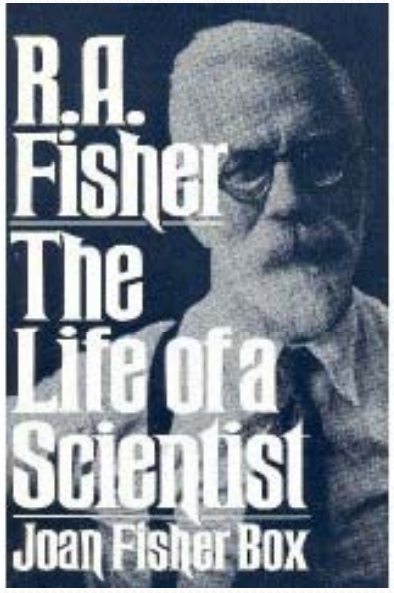


吉村　健一　　京都大学

ランダム化の意義

① 比較可能性

② **P値の算出根拠**



Box (1978)

9

吉村　健一　　京都大学

例．二値の結果変数を群間比較する

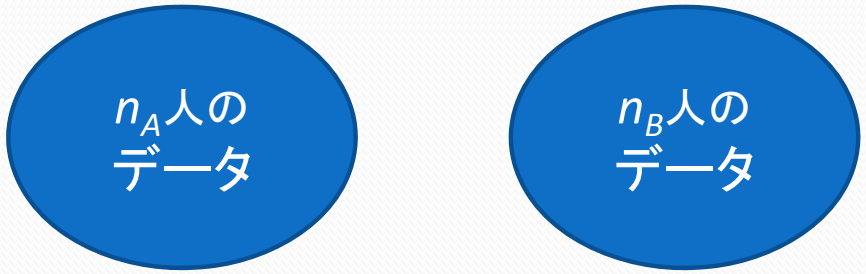
- p値を算出するために

  2つの検定法が一般によく用いられている

  - **（Pearson）χ^2 検定**
  - **Fisherの直接確率検定**

吉村　健一　　京都大学

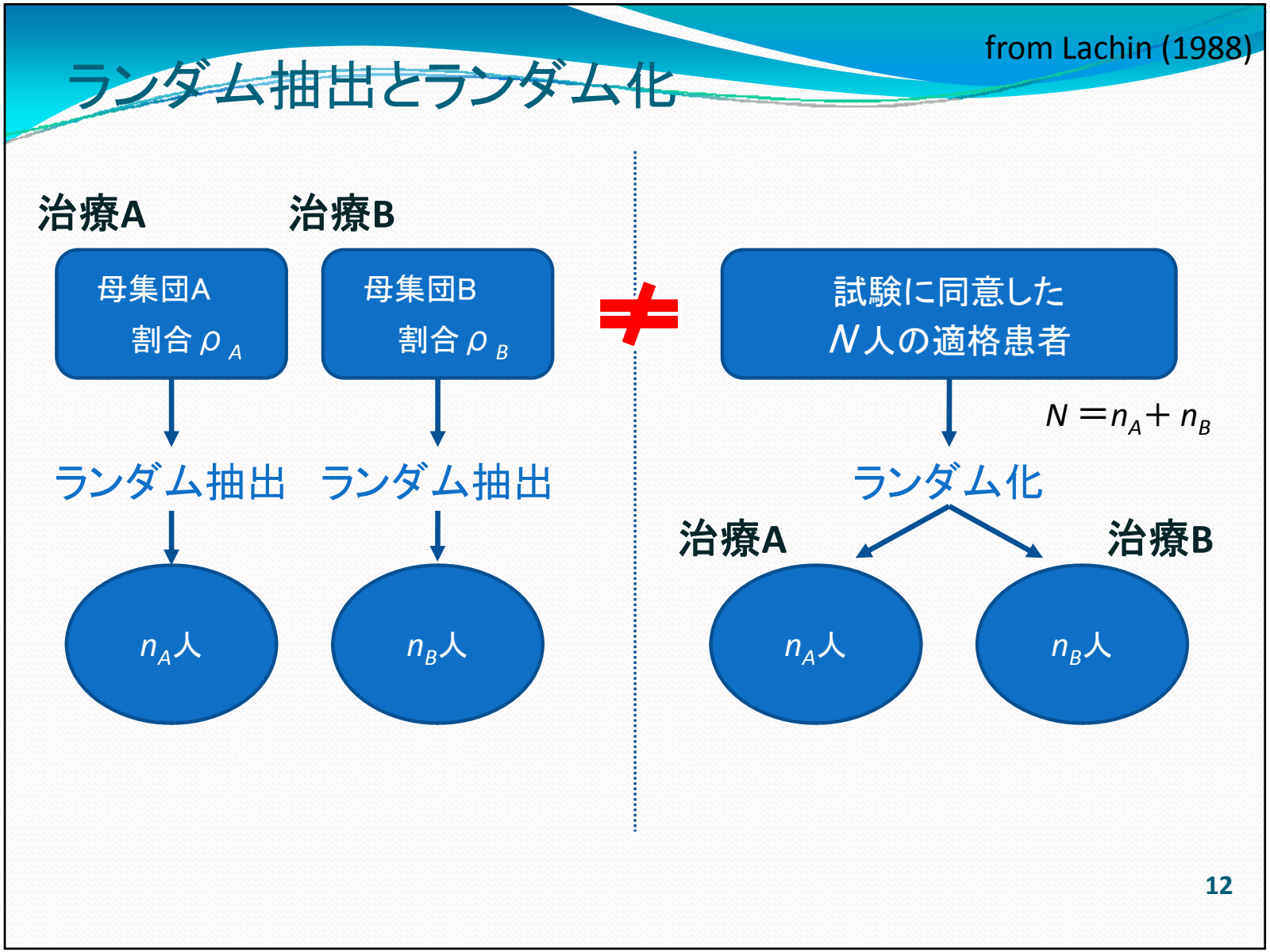
データをどのように得たか？



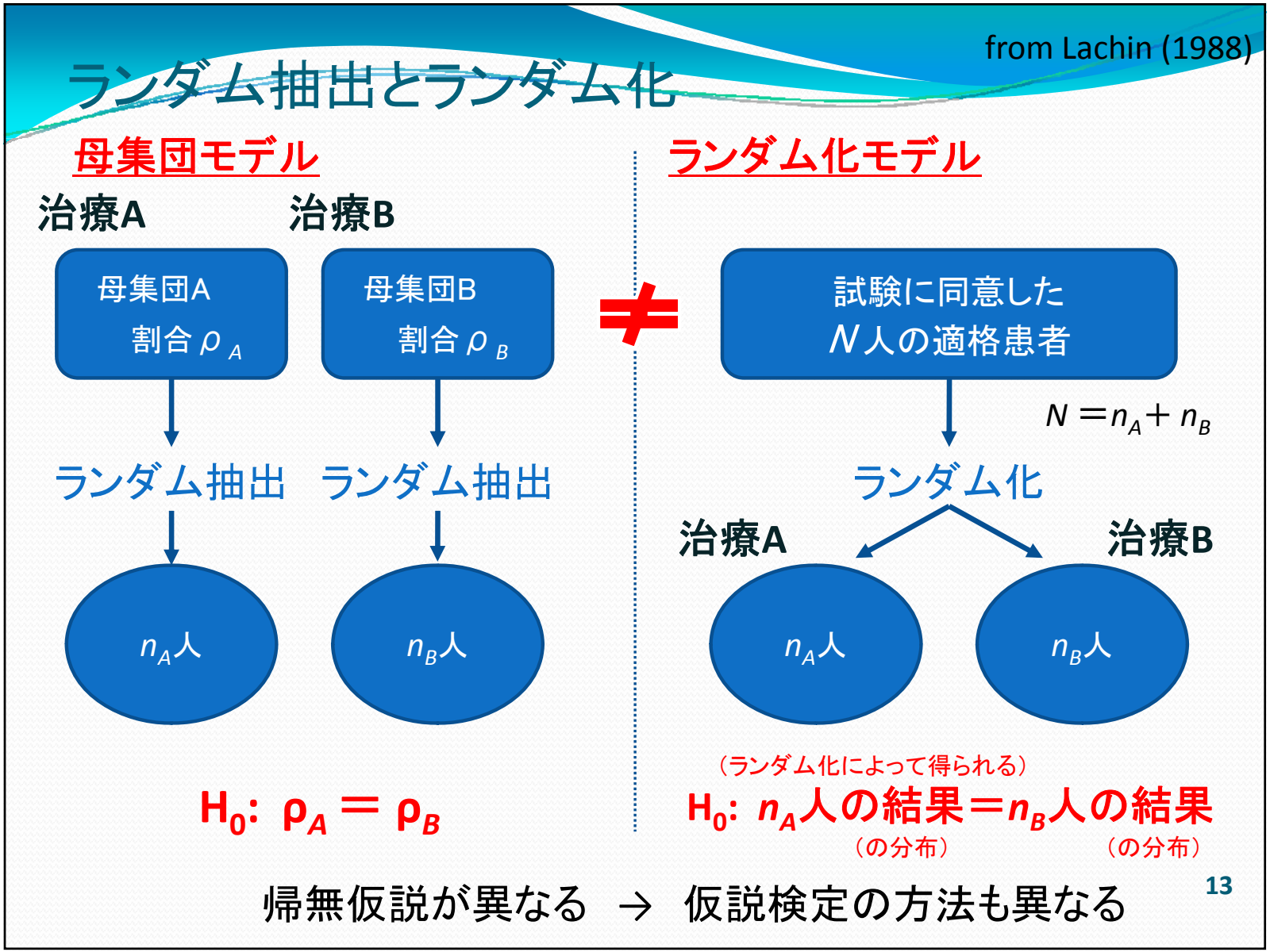
- データをどのように得たかによって“p値の求め方”が異なる
  - ランダム抽出？
  - ランダム化？

11

吉村　健一　　京都大学


ランダム抽出とランダム化
from Lachin (1988)
治療A
治療B
母集団A
割合 ρ_A
母集団B
割合 ρ_B
≠
試験に同意した
N人の適格患者
$N = n_A + n_B$
ランダム抽出
ランダム抽出
ランダム化
治療A
治療B
n_A人
n_B人
n_A人
n_B人
12

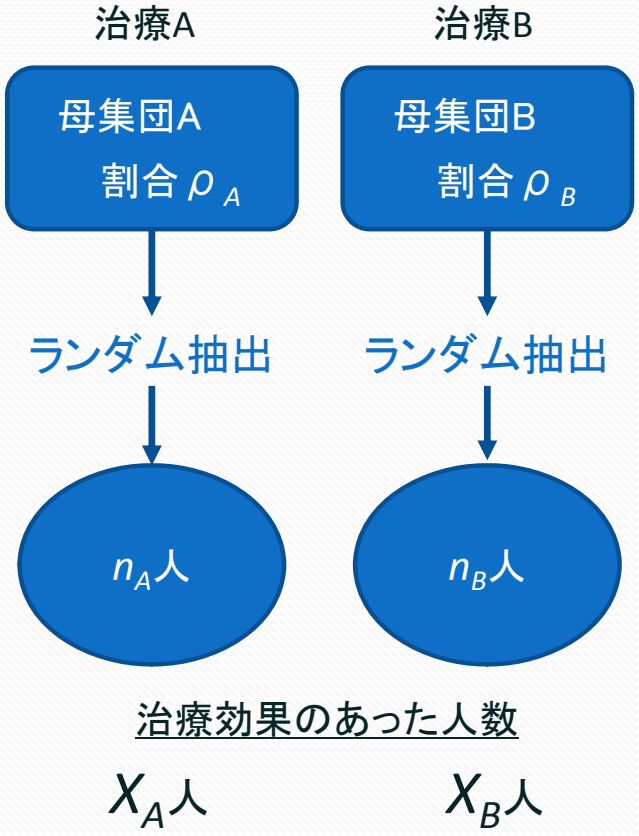
吉村　健一　　京都大学


ランダム抽出とランダム化
from Lachin (1988)
母集団モデル
治療A
治療B
母集団A
割合 ρ_A
母集団B
割合 ρ_B
ランダム抽出
ランダム抽出
n_A人
n_B人
H_0: $\rho_A = \rho_B$
ランダム化モデル
試験に同意した
N人の適格患者
$N = n_A + n_B$
ランダム化
治療A
治療B
n_A人
n_B人
（ランダム化によって得られる）
H_0: n_A人の結果＝n_B人の結果
（の分布）
（の分布）
帰無仮説が異なる → 仮説検定の方法も異なる
13

吉村　健一　　京都大学

Model-based test

- 各母集団の分布を仮定（モデル）
  - **母集団での割合**をρ_A、ρ_B
  - 各母集団から**ランダム抽出**すると
    - X_A　～　二項分布（n_A, ρ_A）
    - X_B　～　二項分布（n_B, ρ_B）
- 検定統計量＝割合の差／その標準誤差を正規近似してZ検定（χ^2検定でも等価）によりP値が求められる
  - ロジスティック回帰やCox回帰などの統計モデルを用いる場合も同様に“モデル”が仮定される

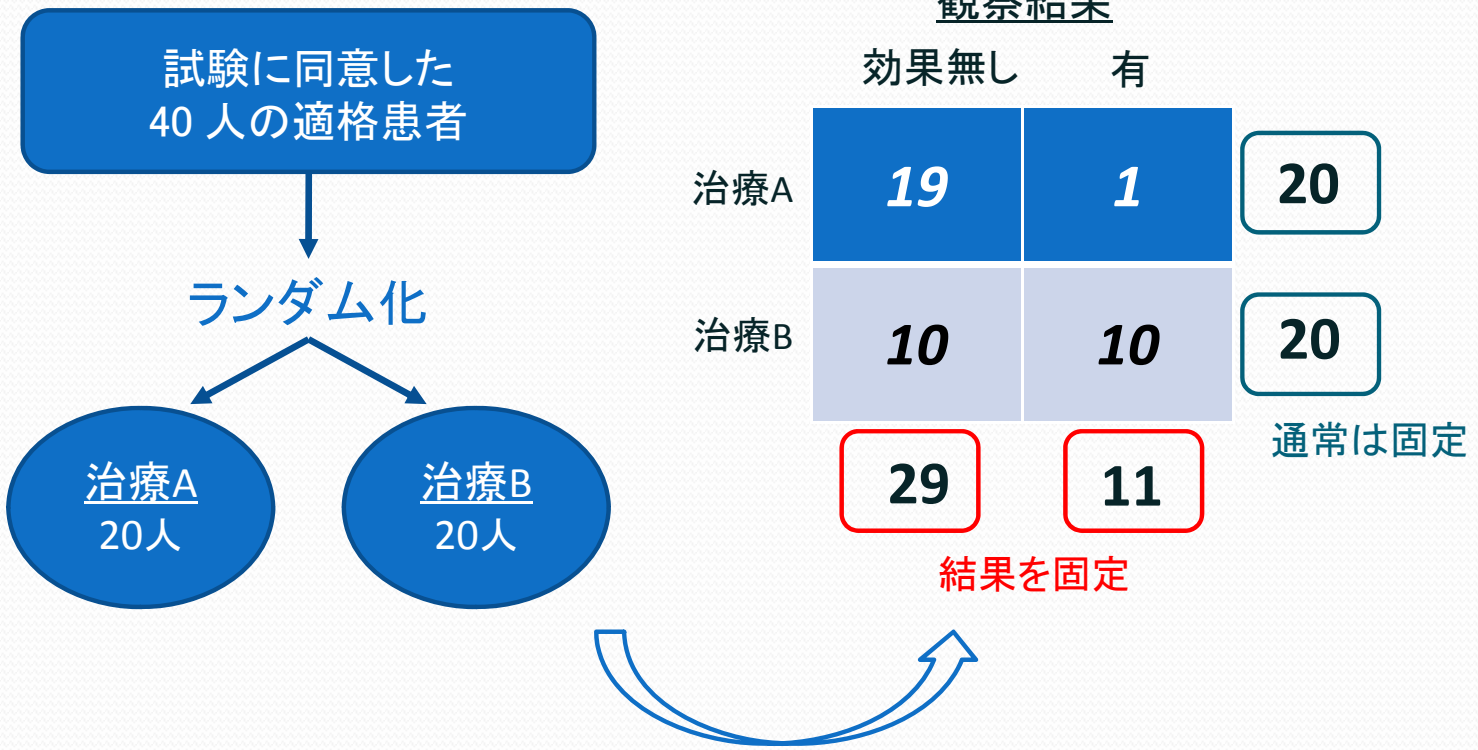


14

吉村　健一　　京都大学

Design-based test

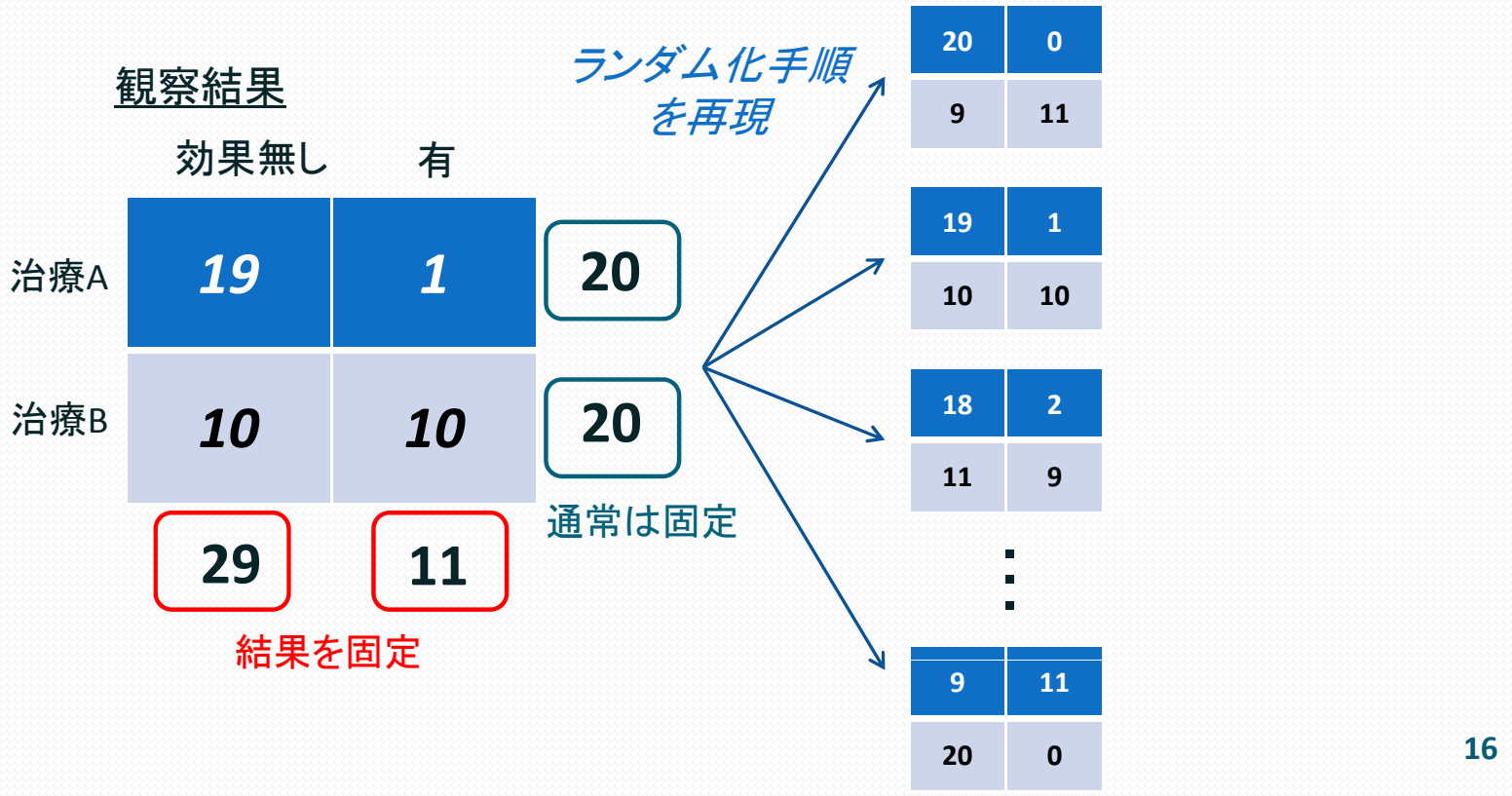
- ランダム化した“事実”のみが前提、その他には仮定不要
- 各対象者の「結果」は固定、「割付結果」がバラつく（確率変数）



吉村　健一　　京都大学

Design-based test

- ランダム化した“事実”のみが前提、その他には仮定不要
- 各対象者の「結果」は固定、「割付結果」がバラつく（確率変数）

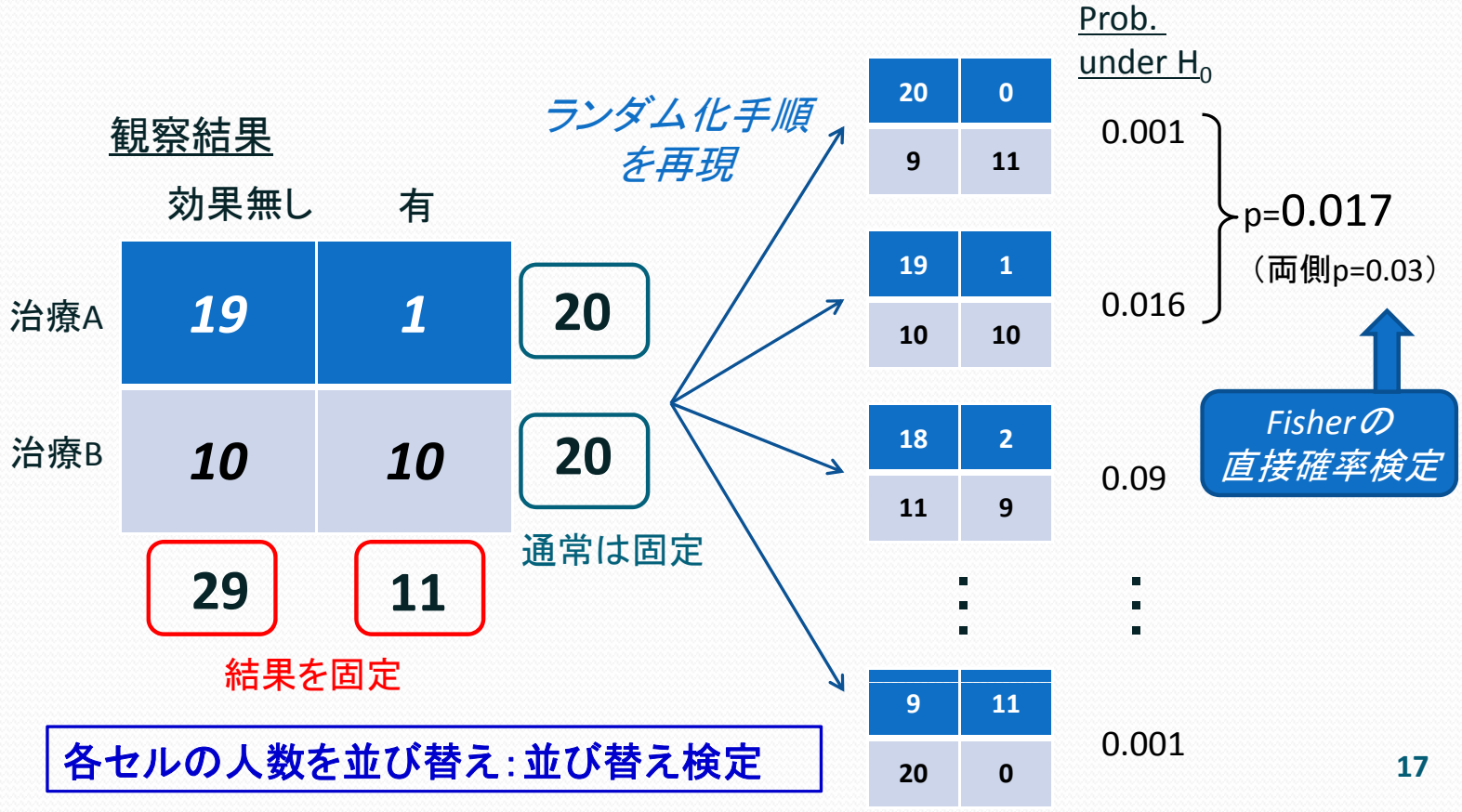


16

吉村　健一　　京都大学

Design-based test

- ランダム化した“事実”のみが前提、その他には仮定不要
- 各対象者の「結果」は固定、「割付結果」がバラつく（確率変数）



吉村　健一　　京都大学

例：実際の解析場面

- 二値結果変数
  - 母集団モデルに基づくmodel-based test
    → Z 検定、χ^2検定など
  - ランダム化に基づくdesign-based test
    → Fisher の直接確率検定など
- 生存時間解析
  - 母集団モデルに基づくmodel-based test
    → Cox回帰など
  - ランダム化に基づくdesign-based test
    → ログランク検定、並び替え検定？
- ランダム化により後者の根拠が与えられる
  - （多くの場合に同等の結果となる）前者による代替もなされる

吉村　健一　　京都大学

ランダム化により
p値の算出根拠がえられる
(*design-based tests)*

19

吉村　健一　　京都大学

本日の内容

- ランダム化　randomization
- **多重性　multiplicity**

20

吉村　健一　　京都大学

多重性の問題

- α=5%の検定を繰り返して行うと

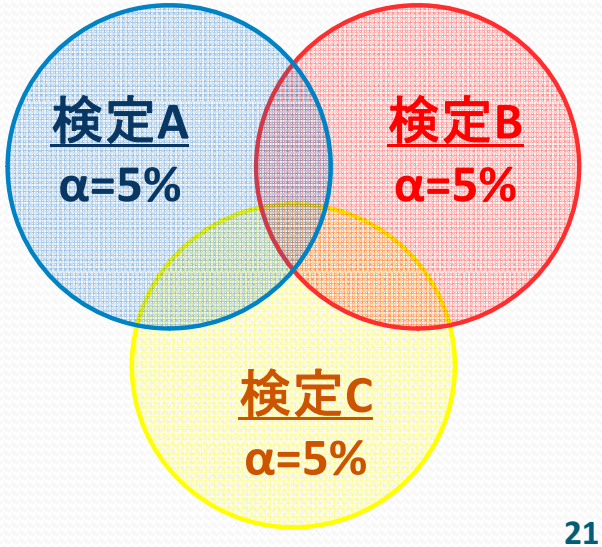
“インフレ”

試験全体でのαエラー（Family-wise error: **FWE**）　＞5%

全て差がないのに1つ以上で誤って有意差有りとなる確率

検定回数	FWE
1	0.05 =α
2	0.10
5	0.23
10	0.40
20	0.64
40	0.87
100	0.99

（独立な検定を繰り返す場合のFWEを算出）



吉村　健一　　京都大学

臨床試験における「多重性」

- 試験中に複数回評価　　中間解析
- 複数のエンドポイントを評価　　多重エンドポイント
- サブグループごとに複数回評価　　サブグループ解析
- 複数の群を比較　　多群試験

22

吉村　健一　　京都大学

中間解析　Interim analysis

- **試験途中**に行われる有効性（又は安全性）に関する**試験治療群間の比較**を意図した全ての解析
  - 医師は潜在的な利益よりもリスクが高いと判断される場合、または有効かつ利益のある結果の決定的証拠が得られた場合は、直ちに研究を中止しなければならない。（ヘルシンキ宣言第20条）
    2008年改訂、日本医師会訳
  - 研究者の利益よりも、**被験者保護**の点から重要
  - 研究者とは独立な委員会（**D**ata **M**onitoring **C**ommittee）によって行われることが一般的
    - FDAのガイドライン：DMCs should be established for controlled trials with mortality or major morbidity as a primary or secondary endpoint.

吉村　健一　　京都大学

試験途中で何度も比較を行うと：FWEの“インフレ”

中間解析回数	FWE
0 （最終解析のみ）	0.05 =α
1	0.08
2	0.10
3	0.12
4	0.13
5	0.14
6	0.15
7	0.16
8	0.17

有意水準0.05の比較を試験期間中に等間隔で実施

- “6年間”の試験で“毎年”比較 → **FWE=約14%**

吉村　健一　　京都大学

αエラー（FWE）の制御：中間解析法

- 事前に決めた少数回の比較
- 試験全体でのαエラー（FWE）を名義水準以下に制御
  - 例．厳しめの “有意水準” を用いて、**FWE=0.05**となるよう調整

総解析回数（中間+最終）		中間解析法	
		Pocock法	O’Brien-Fleming法
2	第1回中間解析	.0294	.0052
	最終解析	.0294	.0480
3	第1回中間	.0221	.0005
	第2回中間	.0221	.0141
	最終	.0221	.0451
4	第1回中間	.0182	.00005
	第2回中間	.0182	.0042
	第3回中間	.0182	.0194
	最終	.0182	.0430

25

※　有意水準αの比較を試験期間中に等間隔で実施した場合

吉村　健一　　京都大学

臨床試験における「多重性」

● 試験中に複数回評価	中間解析
● 複数のエンドポイントを評価	多重エンドポイント
● サブグループごとに複数回評価	サブグループ解析
● 複数の群を比較	多群試験

26

吉村　健一　　京都大学

ある臨床試験の目的：多重エンドポイント??

- 本試験の“主”たる目的は、「セレン剤投与」により
  - **前立腺がん**を予防し
  - **肺がん**を予防し
  - **大腸がん**を予防し
  - **狭心症**を予防し
  - **心筋梗塞**を予防し
  - **白内障**を予防し
  - **高血圧**を予防し
  - …を予防するか

を検証することにある??

FWEの“インフレ”

27

吉村　健一　　京都大学

多重エンドポイントへの対処
臨床試験におけるエンドポイントの決め方

- 主・副を事前に明確に区別して規定
  - 主：　primary endpoint　　（主要評価項目ともいう）

    → 検証的、試験の結論を導くもの
    - 標準的な検証試験の多くは1つだけ規定
    - 当該試験の仮説を評価する上で“最良”のもの
  - 副：　secondary endpoint　　（副次評価項目ともいう）

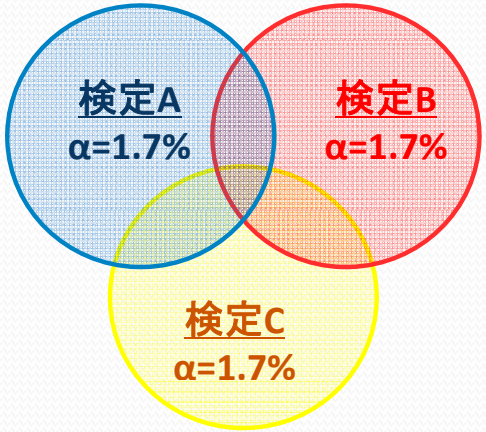
    → 探索的、検証を将来の研究で行うべきもの
    - 結果を過大に解釈しない方が賢明

28

吉村　健一　　京都大学

どうしても複数のprimary endpointを評価したい

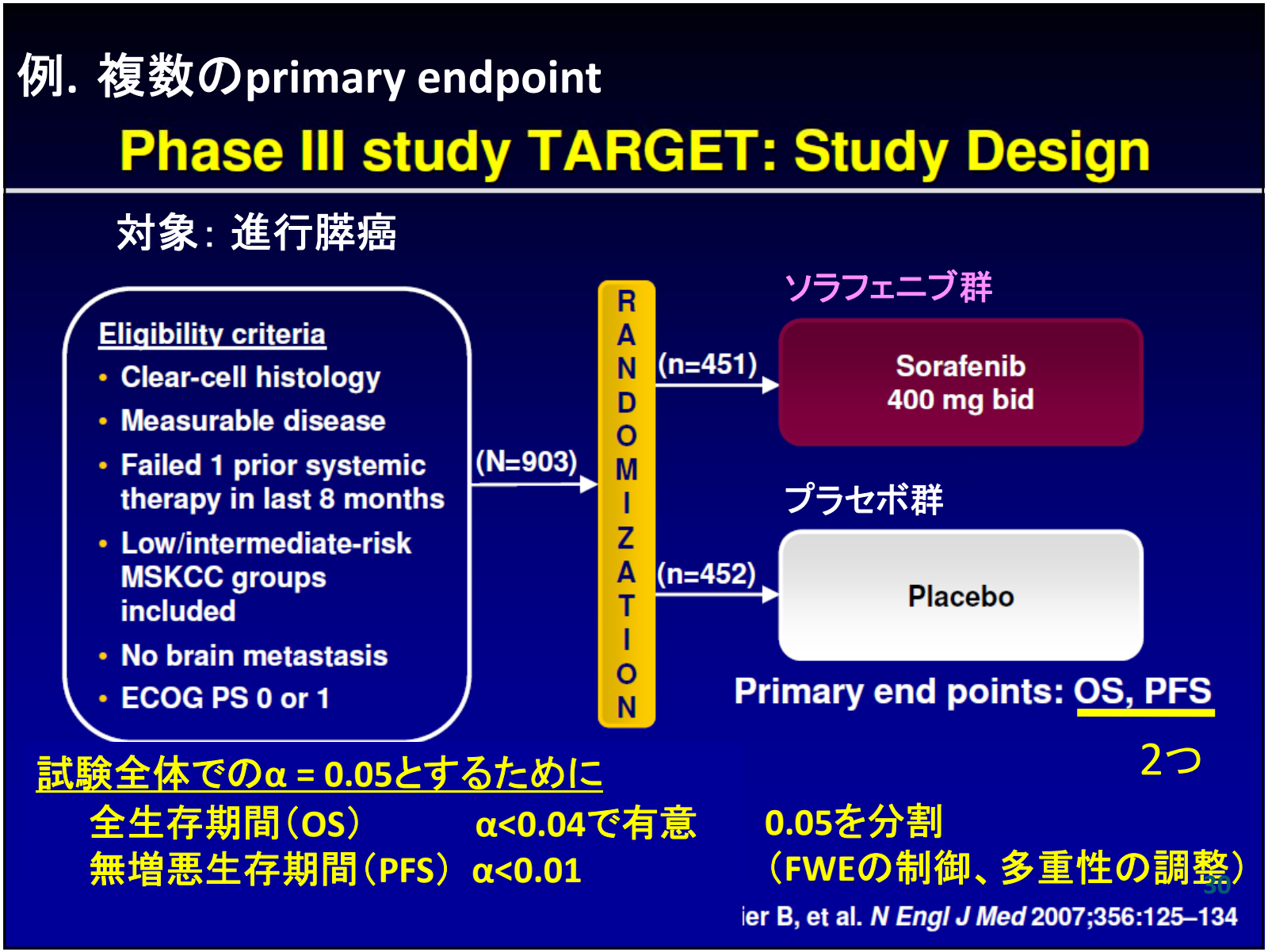
- 何をどのように評価したいかに依る
- 簡便的にはαを配分（分割）：ボンフェローニ法
  - 2つ
    - primary endpoint　X：　<0.025 を有意
    - primary endpoint　Y：　<0.025
  - 3つ
    - primary endpoint　A：　<0.0166 を有意
    - primary endpoint　B：　<0.0166
    - primary endpoint　C：　<0.0166



29

吉村　健一　　京都大学

例. 複数のprimary endpoint



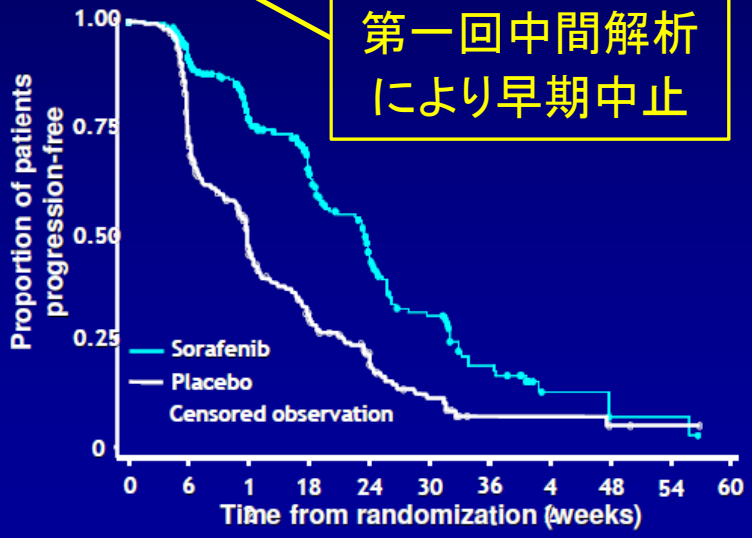
吉村　健一　　京都大学

例．複数のprimary endpoint

無増悪生存期間（PFS）

Median PFS:
Sorafenib=24 weeks
Placebo=12 weeks
Hazard ratio (S/P)=0.44 (95% CI: 0.35, 0.55)
P<0.000001

第一回中間解析により早期中止

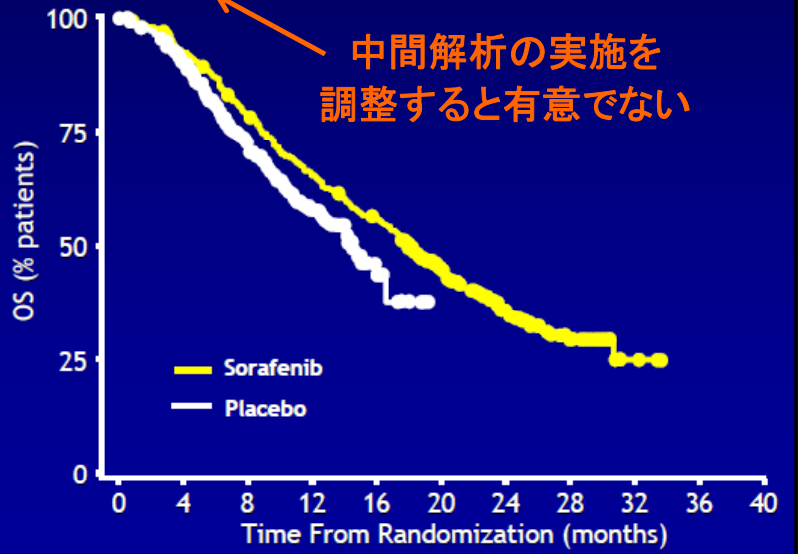


Escudier B, et al. N Engl J Med 2007;356:125–34

全生存期間（OS）

Median OS:
Sorafenib (n=451) = 17.8 months
Placebo (n=452) = 14.3 months
HR (sorafenib/placebo) = 0.78; 95% CI: 0.62–0.97
P=0.0287*

中間解析の実施を調整すると有意でない



*Statistically significant: O'Brien–Fleming threshold for statistical significance α=0.037.
Adapted from Bukowski RM et al. Presented at ASCO 2007; Chicago

吉村　健一　　京都大学

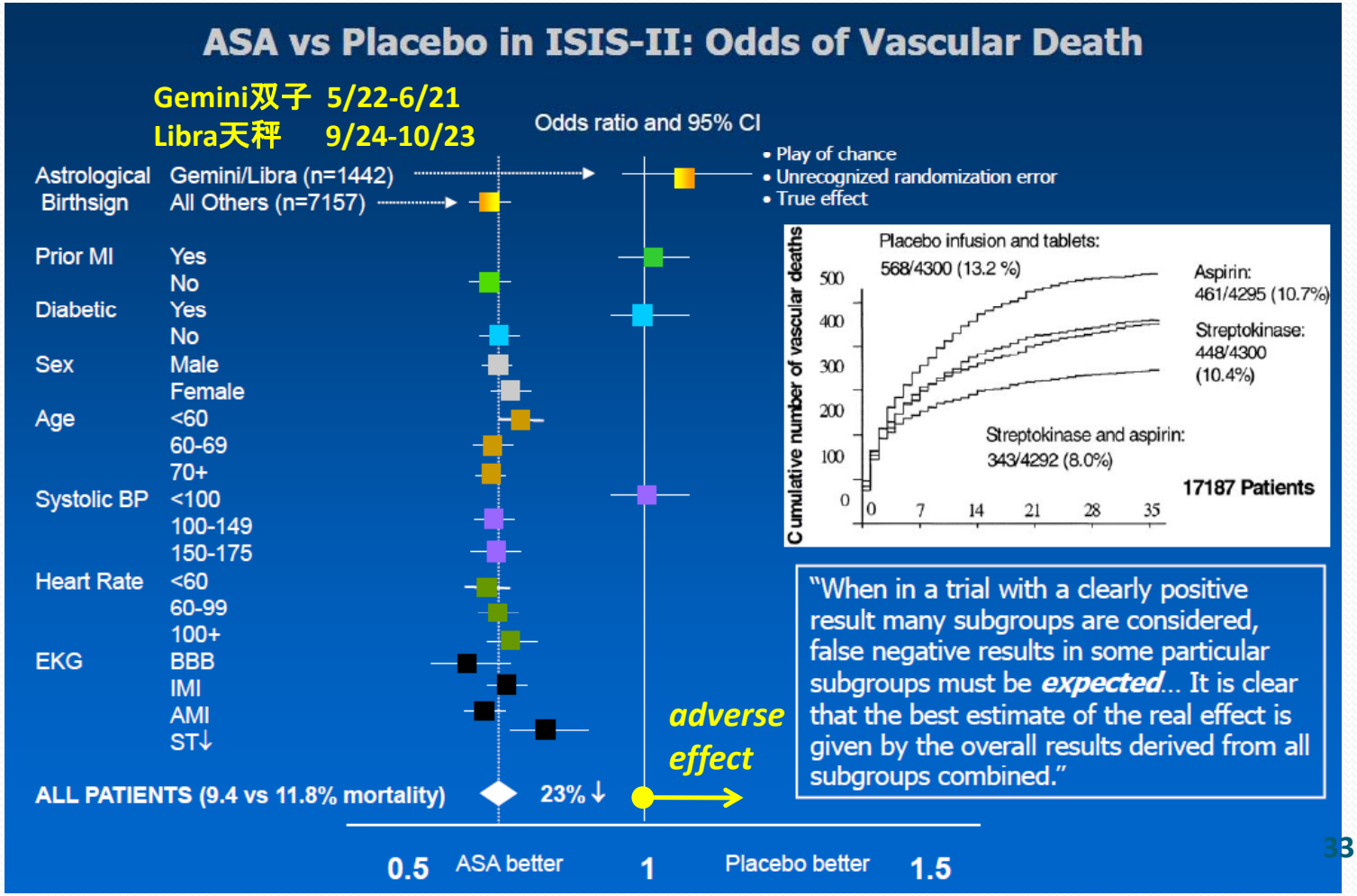
臨床試験における「多重性」

- 試験中に複数回評価　　中間解析
- 複数のエンドポイントを評価　　多重エンドポイント
- サブグループごとに複数回評価　　サブグループ解析
- 複数の群を比較　　多群試験

32

吉村　健一　　京都大学

Subgroup analysis in ISIS-2 study



Subgroup analysis in ISIS-2 study

- The ISIS-2 investigators reported that “subdivision of the patients in ISIS-2 with respect to their astrological birth sign appears to indicate that for persons born under Gemini or Libra, there was a slightly adverse effect of aspirin on mortality (9% increase, SD 13; NS), while for patients born under all other astrological signs, there was a striking beneficial effect (28% reduction, SD 5; 2p <0.00001.)”
  - 双子座、天秤座生まれのヒトに限ると
    アスピリンは死亡リスクを若干増加させる可能性??

FWEの“インフレ”

34

吉村　健一　　京都大学

NEJM's Guidelines for Reporting Subgroup Analysis (2007)

SPECIAL REPORT

Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials

Rui Wang, M.S., Stephen W. Lagakos, Ph.D., James H. Ware, Ph.D., David J. Hunter, M.B., B.S., and Jeffrey M. Drazen, M.D.

Such analyses, which **assess the heterogeneity** of treatment effects in subgroups of patients, may provide **useful information** for the care of patients and for future research.

However, **subgroup analyses** … can lead to **overstated and misleading results.**

For example, Jackson et al.[9] reported the outcomes of a study in which 36,282 postmenopausal

analyses, and it sets forth guidelines for their use

35

吉村　健一　　京都大学

NEJM's Guidelines for Reporting Subgroup Analysis (2007)

- 事前に規定したPrimary endpointに対するサブグループ解析のみを
  それら解析全体として解釈できる場合に限って要旨に記載
- 方法にサブグループ解析の数（報告数、実施数）を明記
  - 事前に規定して実施と事後的に実施を区別
- エンドポイントと解析法の明記
- 特に興味のあるものをhomogeneityの評価と区別して記載
- αエラーへの影響の程度、それに対する対応法を明記
- 治療効果の不均一性の評価には交互作用に対する検定
- discussionでは、過度の解釈を避ける、信憑性の評価、限界を記載
  - 結果を支持/矛盾する先行研究結果を提示

36

吉村　健一　　京都大学

多重性は問題となりうるが適切な方法により対処する

- 中間解析法
- primary endpoint と secondary endpoint
- サブグループ解析に関するガイドライン（NEJM）など

吉村　健一　　京都大学

The most important medical developments of the past millennium, by Editors (2000) *N Eng J Med*

1. lucidation of human anatomy and physiology
2. Discovery of cells and their substructures
3. Elucidation of the chemistry of life
4. **Application of statistics to medicine**
5. Development of anesthesia
6. Discovery of the relation of microbes to disease
7. Elucidation of inheritance and genetics
8. Knowledge of the immune system
9. Development of body imaging
10. Discovery of antimicrobial agents
11. Development of molecular pharmacotherapy

疫学に限らず
臨床試験でも

38

Take Home Messages

- ランダム化により群間の**比較可能性**
- ランダム化は**p値の算出根拠**
- 多重性により試験全体でのαエラー（**FWE**）**が上昇**
- 適切な**中間解析法**により、αエラーを制御
- **primary endpoint**と**secondary** ～の役割は異なり
  両者を事前に明確に区別
- **サブグループ解析**に関する新しい**ガイドライン**

39

Further readings

- Piantadosi S. Clinical Trials -A Methodologic Perspective. Wiley. 2005.
- Wang R et al. Statistics in medicine -reporting of subgroup analyses in clinical trials. *New Eng J Med*, vol.357:pp2189-94. 2007.
- 椿広計, 佐藤俊哉, 藤田利治（編）. これからの臨床試験―医薬品の科学的評価-原理と方法. 朝倉書店. 1999.
- 福田治彦, 新美三由紀, 石塚直樹（訳）. 米国SWOGに学ぶがん臨床試験の実践 –臨床医と統計家の協調をめざして. 医学書院. 2004.（Green S *et al.* Clinical Trials in Oncology, 2nd ed. CRC press. 2003.）
- 丹後俊郎. 無作為化臨床試験. 朝倉書店. 2003.
- 丹後俊郎, 上坂浩之（編）. 臨床試験ハンドブック. 朝倉書店. 2006.

吉村　健一　　京都大学

附録

吉村　健一　　京都大学

比較可能性と交絡因子の三条件

- 比較可能性がない＝交絡がある と定義すると
  - 交絡因子とは比較可能性を崩す因子
  - 交絡因子の3条件は比較可能性を崩すための必要条件として導出可能
    - 両群で分布が異なる
    - 予後に影響をする
    - 因果推論の一般論から中間変数は調整すべきでない

42

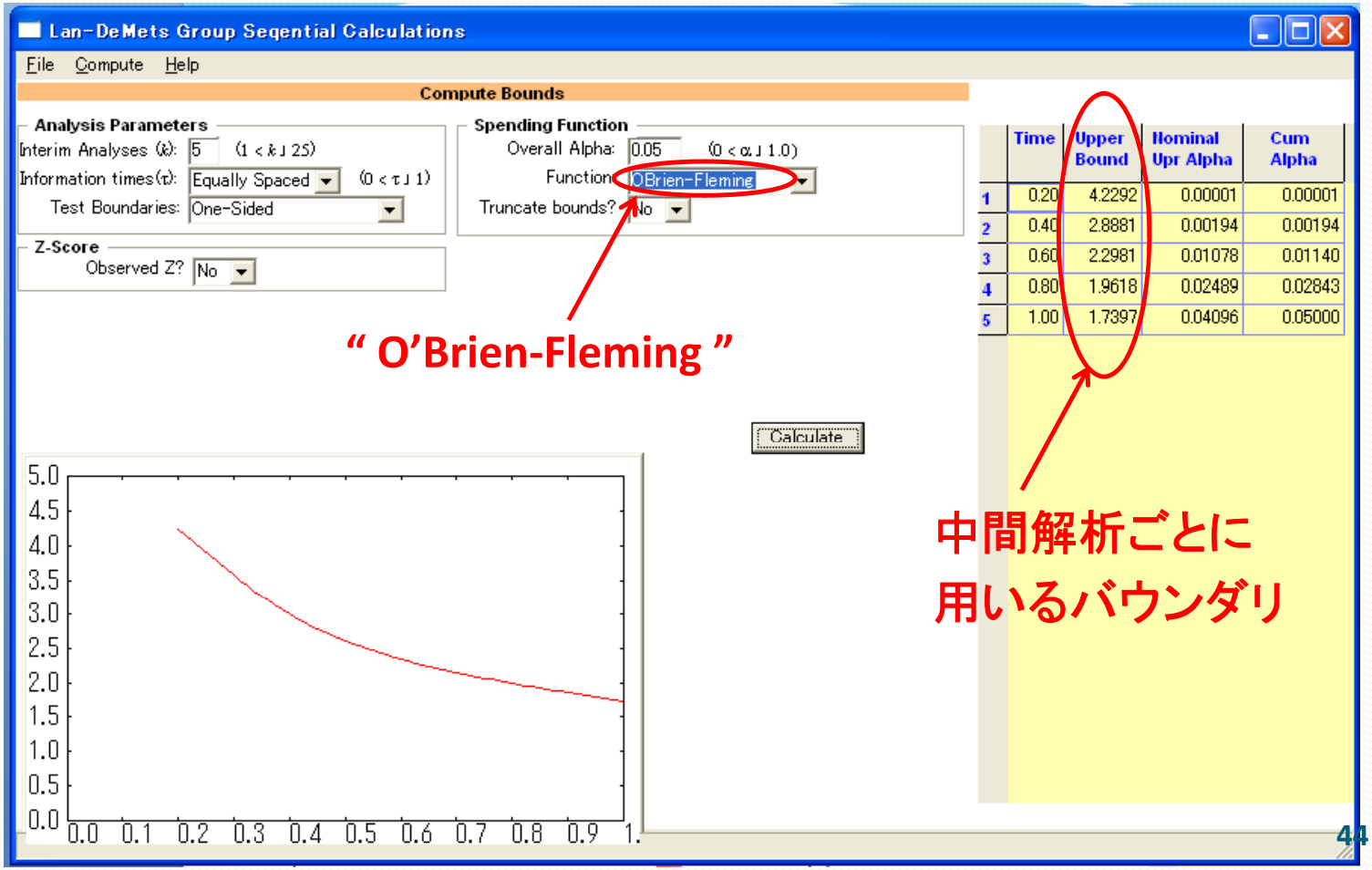
吉村　健一　　京都大学

中間解析法：ソフトウェア

- **S+SeqTrial** (Insightful Corporation)
- **EaST** (Cytel)
- **PEST 4** (University of Reading)
- **Lan-DeMets Group Sequential Calculator** (University of Wisconsin)
  - http://www.biostat.wisc.edu/landemets/

吉村　健一　　京都大学

出力の例（中間解析ソフトウェア）



吉村　健一　　京都大学

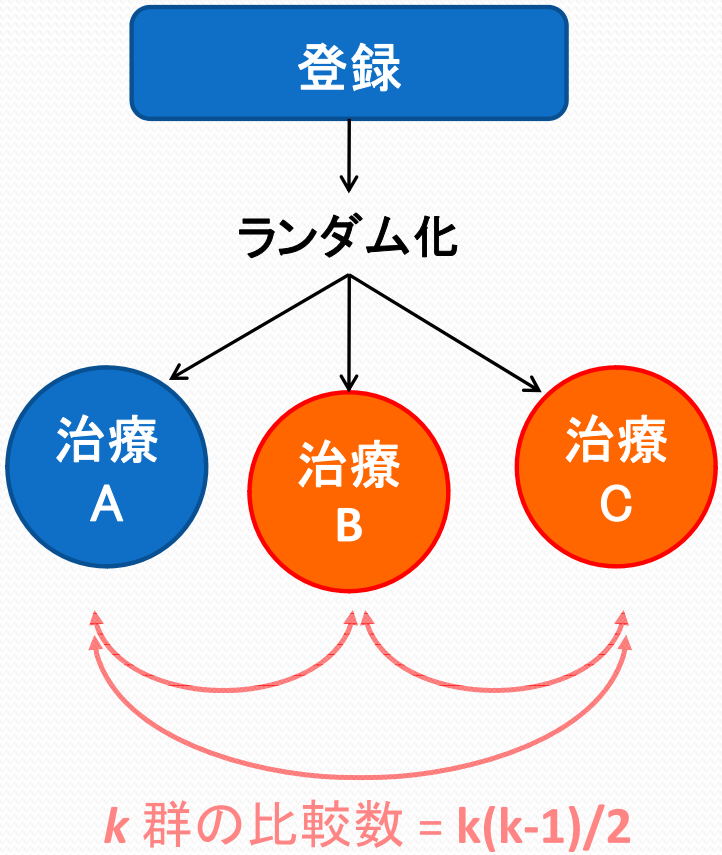
臨床試験における「多重性」

- 試験中に複数回評価　　中間解析
- 複数のエンドポイントを評価　　多重エンドポイント
- サブグループごとに複数回評価　　サブグループ解析
- 複数の群を比較　　多群試験

45

吉村　健一　　京都大学

多重性：多群試験



- **検証したい仮説（比較）の数**に応じて調整できる
  - 標準治療 *vs.* 2つの新治療
    - A vs. B　　α=0.05/**2**
    - A vs. C　　α=0.05/**2**

      by Bonferroni 法
  - 全ての対比較に興味
    - A vs. B　　α=0.05/**3**
    - A vs. C　　α=0.05/**3**
    - B vs. C　　α=0.05/**3**

      by Bonferroni 法