

# 臨床試験に関する 生物統計学の基礎

東京大学大学院医学系研究科  
公共健康医学専攻 生物統計学  
松山 裕

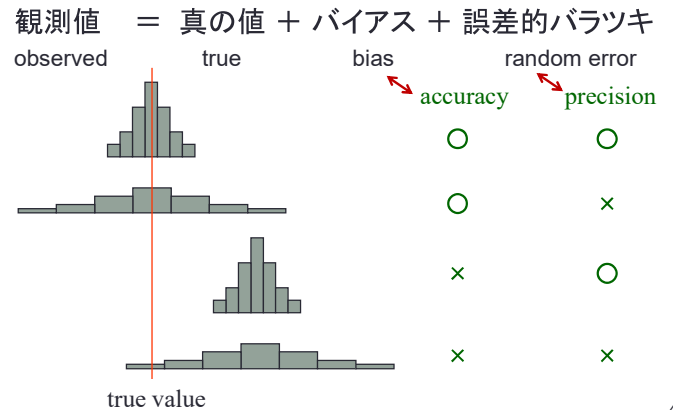
## 本日の話

- 統計学とバラツキ
  - ランダムなバラツキ
  - バイアスによるバラツキ
- 統計解析(生存時間解析)
  - Kaplan-Meier法による生存曲線の推定
  - Log-rank検定による生存曲線の比較
  - ハザード比の推定・信頼区間

## “統計学”とは何か?

- 統計学で問題にするのはバラツキ (Variation)
  - バラツキとは、「同じようなものを測っているのに値がいろいろと違ってしまふこと、違っていること」
- そのバラツキを
  - 誤差的な部分(ランダムエラー)と意味(physical meaning)のある部分(対応可能部分)とに分解し、誤差を考慮に入れて意味を理解しようという視点・態度
  - さらに、バイアスとなりうるバラツキを(積極的に)確率論で評価可能な誤差に転化し、真実に迫ろうとする視点・態度
    - 方法: ランダムサンプリングとランダム化

## 精密度と正確度 Precision and Accuracy



## なぜバラツキ?

進行・再発大腸癌患者に対する5FUの成績

Authors	# of patients	response (%)	奏効率(CR+PR)
Sharp and Benefell	13		85
Hall and Good	19		63
Rochlin et al.	47		55
Allaire et al.	17		47
Cornell et al.	13		46
Every	12		41
Field	37		41
Bell	22		36
Weiss and Jackson	37		35
Ferguson and Humphrey	12		33
Hurley	150		31
ECOG	48		26
Talley	271		21
Hyman et al.	30		20
Moore et al.	80		19
Ansfield	141		17
Mayo	358		17
Ellison	87		12
Kennedy	22		9
Knoepp et al.	11		9
Olson and Green	12		8

## 誤差的バラツキとバイアス

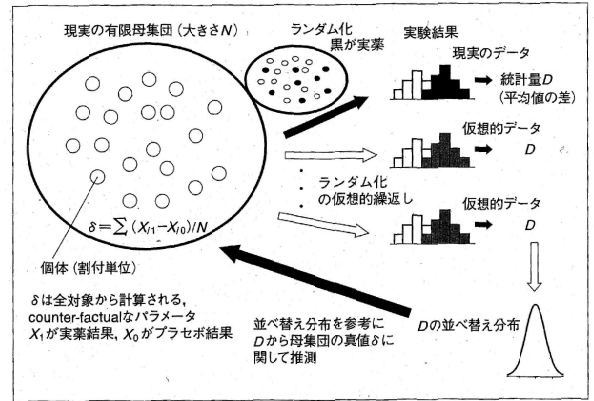
- さまざまな理由による「バイアス」
  - 例えば、進行癌に対するある化学療法の奏効率
    - 患者全身状態(PS)とADMEの違い
    - 腫瘍の性質(病理・分子生物学的)
    - dose-intensityとコース、評価部位と評価方法、出版バイアス、...
    - 施設差
- モデルとしての「誤差」
  - バラツキの理由は同定できない、あるいはあえてしない
  - 確率論の応用が可能
- 誤差的バラツキとバイアスの相対性
  - 知識が深まれば / 情報が得られれば、誤差的バラツキは制御可能なバイアス要因に転化

## 医学研究デザインで必要なこと

- 誤差的バラツキを小さくすること(精度を高くすること)
  - サンプルサイズを増やす
  - 測定の精度を上げる
- 偏り(バイアス)を小さくすること
  - デザイン上の工夫が必要(サンプルサイズを増やしても減少しない)
  - その中で最強の方法が、**ランダム化**(randomization)
  - 統計解析でも制御可能
- そして、得られた結論の一般化可能性を高めること

7

## ランダム化による統計的推測



大橋靖雄著. 医師のための臨床統計学. 医歯薬出版株式会社. P108

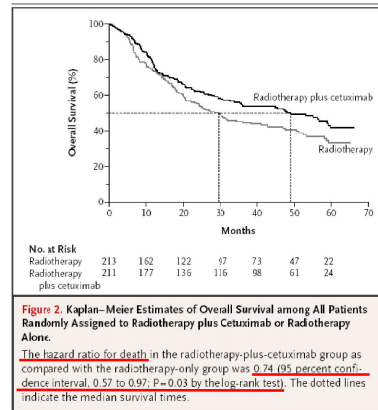
8

## 結果変数の型による統計手法の分類

解析目的	変数の型		
	二値	連続	生存時間
分布の記述	頻度集計 分割表	ヒストグラム 平均・SD、相関	生存曲線 Kaplan-Meier法
単純な 群比較	$\chi^2$ 検定、 リスク推定	$t$ 検定 平均値の推定	Log-rank検定 ハザードの推定
層別解析	MH法 標準化	分散分析	層別 log-rank
回帰モデル	Logistic回帰	分散分析 重回帰分析	Cox回帰

9

## よく見かける結果

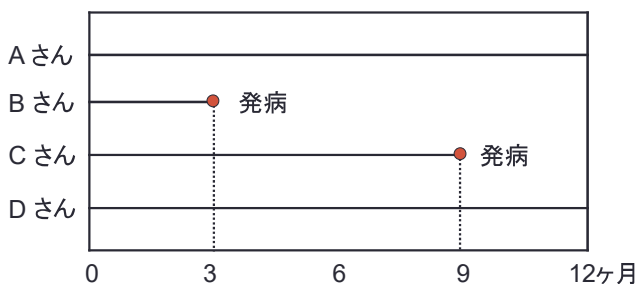


Kaplan-Meier曲線  
Log-rank検定の結果  
P値 = 0.03  
↓  
統計的に有意であった  
ハザード比 = 0.74  
95%信頼区間: 0.57 - 0.97

10

## 疾病発生確率(リスク)

- 次の4人の集団について、疾病発生**確率**を計算してください



11

## 発生確率(リスク、割合)

- 2名発病、最初に4人いたから  
 $2/4 = 0.5 (= 50\%)$
- 分子が分母に含まれる分数
  - 割合は次元(単位)をもたない
  - 必ず、0から1(0%から100%)の間の値をとる

12

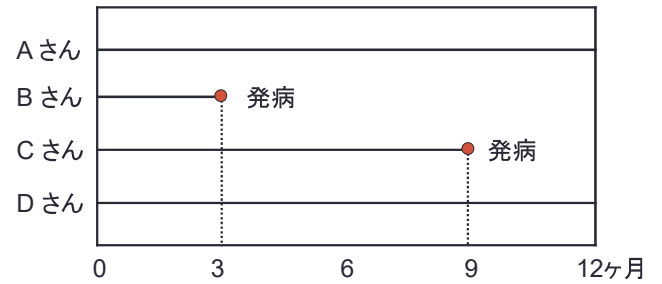
## 発生リスク

- ・当該のイベントが何人中何人に発生したか?
- ・問題点
  - ・イベント発現までの時間を考慮していない

13

## 疾病発生率(ハザード)

- ・次の4人の集団について、ハザードを計算してください



14

## 人時間 (Person-Time) 法

- ・最も単純な「ハザード」の計算法
- ・人時間
  - ・ At Risk にある時間の合計 (総観察期間)
  - ・ 測定単位によって、人年、人月、...
- ・人時間法による発生率の計算
  - ・  $\frac{\text{発生数}}{\text{人時間}}$  (/時間)

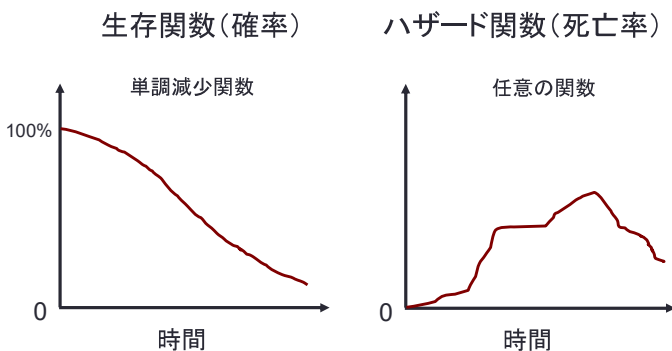
15

## 発生率 Incidence Rate, Hazard

- ・2名発病
- ・打ち切り症例を含む観察期間の合計
  - ・  $1年 + 1/4年 + 3/4年 + 1年 = 3人年$
- ・発生率(ハザード)
  - ・  $2/3人年 = 0.67 (/year)$
  - ・ = 1人年あたり 0.67
  - ・ = 100人年あたり 67

16

## 生存関数とハザード関数



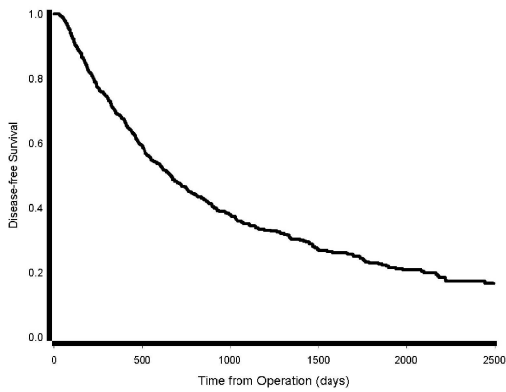
17

## 発生率(ハザード)

- ・当該のイベントが何人中何人に発生したかだけでなく、イベント発現までの時間を考慮
- ・速度の概念
  - ・ 単位時間当たりのイベント発生数
- ・割合(確率、リスク)と違って、単位がある
  - ・ 多くは、時間の逆数

18

## 肝細胞がん術後の無再発生存曲線



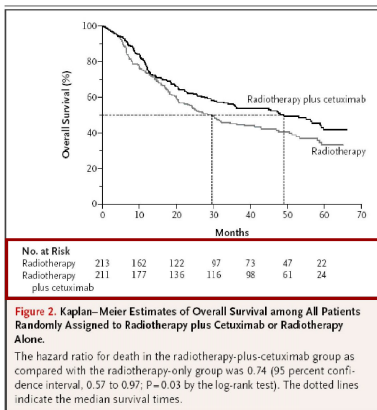
19

## 肝細胞がん術後の再発ハザード



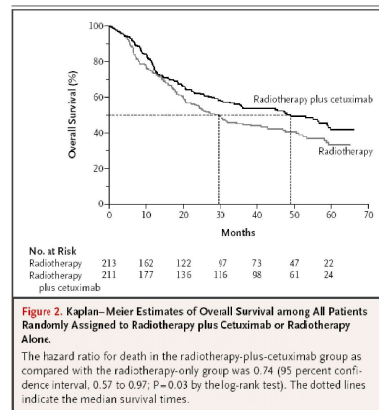
Imamura et al, Journal of Hepatology 2003 20

## At Riskの人数(リスクセット)の記載



各群における  
number of at risk

## 生存曲線の比較



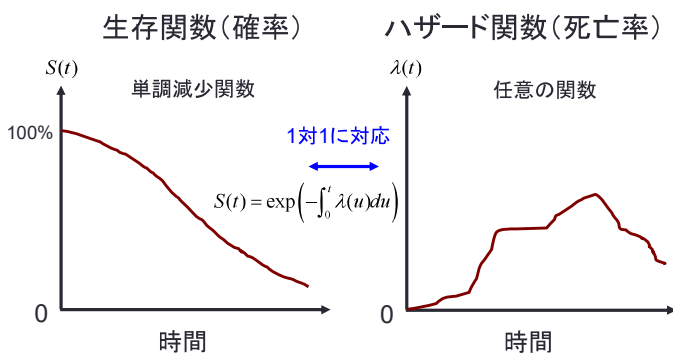
2群間の生存曲線に  
違いがあるように  
見えるが、もう一度、  
同じ研究をすれば、  
2つの曲線が重なる  
ことはないのか？

実は1本の曲線？

21

22

## 生存関数とハザード関数(再)



23

## 2本の生存曲線が同一とは？

- 各グループの「死亡ハザード」が等しい
- ハザード比 = 1

$$\frac{\text{グループAのハザード}}{\text{グループBのハザード}} = 1$$

24

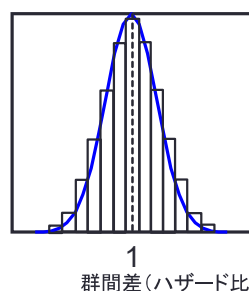
## 「ハザード比 = 1」を否定したい

- 帰無仮説 Null hypothesis
  - 否定するための仮説
  - 治療効果なし
  - 今の例では、ハザード比 = 1
- 統計的仮説検定とは、帰無仮説が正しいという仮定のもとで、
  - 仮想的な試行実験を考える
  - 同じ人数の研究を何度も繰り返したとする

25

## 帰無仮説が正しいければ、

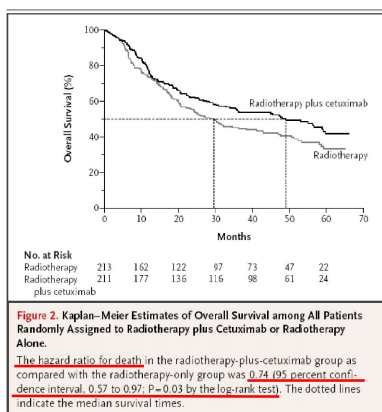
- 観察されるハザード比は1を中心に左右対称にバラツクはず



仮想的に同じ規模の研究を繰り返したときのハザード比の分布

26

## 自分のデータでは、HR=0.74



Log-rank検定の結果  
P値 = 0.03

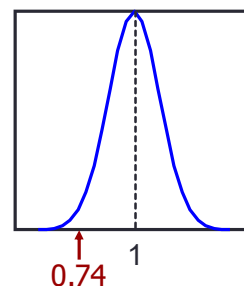
↓  
統計的に有意であった

ハザード比 = 0.74  
95%信頼区間: 0.57 - 0.97

27

## 帰無仮説が正しいとき

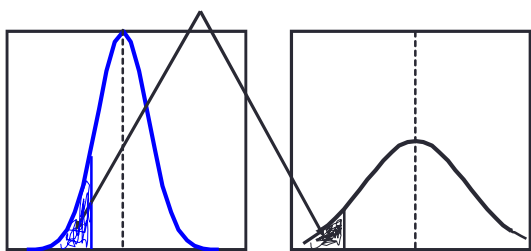
- 観察されたハザード比が、1から十分離れていたら、帰無仮説が間違っているのでは?



28

## 帰無仮説から十分離れた?

- この面積が小さければ、どちらの場合でも帰無仮説から十分離れたと考えよう



精度の高い研究

精度の低い研究

29

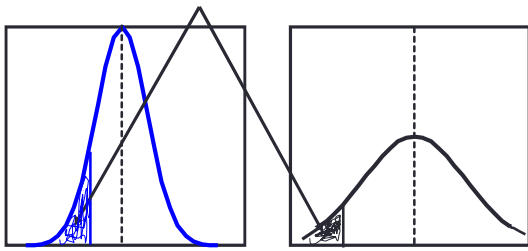
## P値 P-value (さっきの面積)

- 帰無仮説「2つのグループの生存曲線が同一」が正しい場合
- P値:
  - 群間差が観察された値と同じか、それよりもっと仮説から離れた値をとる確率

30

## 帰無仮説から十分離れた?

- この面積が小さければ、どちらの場合でも帰無仮説から十分離れたと考えよう



精度の高い研究

精度の低い研究

31

## 帰無仮説から十分離れている

- 絶対値と研究の精度の両方を考慮
- 絶対値
  - 群間差(ハザード比)の大きさ
- 研究の精度
  - サンプルサイズの大きさ
- 研究の精度が低くても、差が大きければ
- 差が小さくても、研究の精度が高ければ

帰無仮説から十分離れていると考えよう

32

## 検定の結果 1

- P値が事前に定めた値(有意水準:通常は5%)よりも小さければ、「帰無仮説を棄却する」
  - ハザード比は1ではないと判断する
  - 2本の生存曲線の間には統計的に違いが認められる

33

## 検定の結果 2

- P値が事前に定めた値よりも大きければ、「帰無仮説を採択する」
  - ハザード比は1だと判断する
  - 2本の生存曲線は同一のものだと考える

34

## 検定の結果 2

- P値が事前に定めた値よりも大きければ、「帰無仮説を棄却しない」
  - ハザード比は1であることを否定できなかったと判断する
  - 「2本の生存曲線が同一」かどうかは判断保留と考える

35

## 検定結果が有意

- 統計的に「差がある」といえる  
Statistically Significant

36

## 有意でなかった場合は...

- 帰無仮説「ハザード比 = 1」は否定するためのもの
  - 否定できなければ、積極的に何も統計的にはいえない
- さもなければ、
  - いい加減な(人数の少ない)研究を行えば、常に同等(群間差なし)の結果が得られる

37

## 検定の結果の解釈

- 有意差あり = 帰無仮説は間違い
- 有意差なし ≠ 帰無仮説は正しい
- ただし、
  - 十分な検出力が確保された質の高い試験であれば、例えば、 $P$ 値 = 0.90は「群間差なし」と解釈しても大きな間違いではない
  - 本来は、2群が同等であることを示したければ、**非劣性試験**

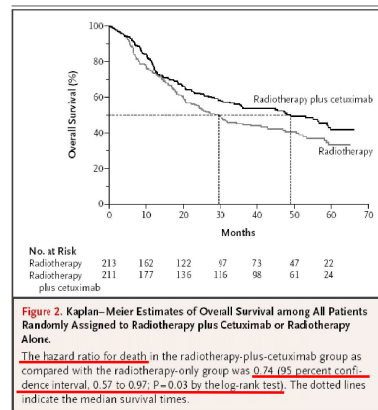
38

## 検定だけでなく、推定も重要

- 試験結果をP値で議論
  - 臨床的には意味のない差を検出することもある
  - 重要な群間差を見逃すこともある
- 治療効果の大きさの推定
  - **ハザード比**
  - リスク比
  - NNT (Number Needed to Treat)
  - ...

39

## 自分のデータでは、HR=0.74



Log-rank検定の結果  
P値 = 0.03

↓  
統計的に有意であった

ハザード比 = 0.74  
95%信頼区間: 0.57 - 0.97

40

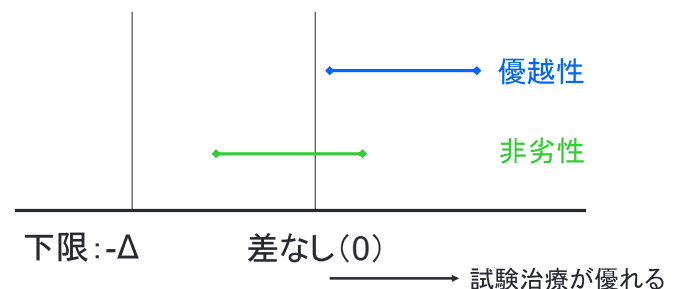
## 95%信頼区間

- 治療効果の大きさの推定精度
  - 症例数(あるいは、イベント数)が多いと狭くなり、少ないと広がる
- データと矛盾しない効果の大きさの範囲
- 例えば、
  - ハザード比 = 0.74、95%CI = 0.57 - 0.97
  - 帰無仮説(ハザード比=1)が95%CIに含まれていないので、有意水準5%で統計的に有意

41

## 優越性試験と非劣性試験

- 治療効果の差(test - control)と95%信頼区間



42

## バイアスの少ない研究を目指すためには、

- 比較の妥当性 Comparison Validity → ランダム化
  - ・そもそも比較群はよく似ている集団か？
- 追跡の妥当性 Follow-Up Validity
  - ・対象者の選択的な脱落はないか？
- 測定の妥当性 Measurement Validity
  - ・様々なデータの測定誤差(観測の不完全性)は大きくないか？
- 解析の妥当性 Specification Validity
  - ・統計モデルの現実からの乖離は大きくないか？

43

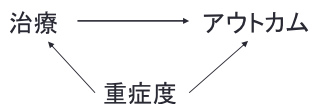
## ランダム化研究(RCT) v.s. 観察研究

妥当性	RCT	観察研究
比較	○	×
追跡	△	△
測定	△	△
解析	○	△

44

## 比較の妥当性に欠ける

- 交絡(Confounding)バイアスがある
  - ・比較群間で予後因子の状態がことなる
  - ・観察研究では常に存在する
- 例えば、「手術 v.s. 内科的治療」では...
  - ・重症な患者ほど内科的治療を受けていれば...
  - ・何を比較しているのか？ 治療法か、重症度か？



45

## 交絡バイアスの補正・調整

- 層別解析
  - ・交絡因子のレベルごとに解析して、それらの結果を併合
  - ・例えば、ステージ別に解析
- 回帰モデル
  - ・アウトカムと交絡因子の間に数学モデルを仮定
  - ・例えば、比例ハザードモデル(Cox回帰)
    - ・ある変数Xを有している対象者の時点tでのイベント発生ハザード:  $h(t,x)$

$$h(t,x) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)$$

治療変数

重症度

$\beta$ : 回帰係数

交絡調整済みのハザード比:  $\exp(\beta_1)$

46

## まとめ

- 医学研究における統計学の利用
  - ・データのバラツキとバイアス
  - ・バイアスの除去とバラツキの最小化
- 様々な統計解析手法
  - ・アウトカム変数の型による統計解析手法の分類
- 統計的仮説検定
- 治療効果指標の推定と95%信頼区間

47