

## 製造販売後臨床試験 (2003.11-2006.2)

非小細胞肺癌患者におけるゲフィチニブ投与及び非投与での急性肺障害・間質性肺炎の相対リスク及び危険因子を検討するためのコホート内ケースコントロールスタディ

結果報告書 (2006-09-04)

ゲフィチニブ投与例の治療関連死は1.6%で、主な死因は急性肺障害・間質性肺炎であった

発症リスクは、とくに投与開始後4週間以内で高い

予後不良因子として、①65歳以上の高齢、②喫煙歴、③既存の間質性肺炎、④治療前のCT画像を用いた判定による正常肺占有率が低いことなどが示唆された

## 製造販売後臨床試験

非小細胞肺癌患者におけるゲフィチニブ投与及び非投与での急性肺障害・間質性肺炎の相対リスク及び危険因子を検討するためのコホート内ケースコントロールスタディ

結果報告書 (2006-09-04)

副次目的：投与時の薬物動態学的特性を評価し、発症との関連性を検討する

経口クリアランス影響因子：血清 $\alpha_1$ -酸性糖蛋白、年齢、CYP3A4誘導剤併用

見かけの分布容積影響因子：血清 $\alpha_1$ -酸性糖蛋白、体重

発症例の非発症時と非発症例のゲフィチニブ暴露量に有意差がなかったが、発症例における発症時と非発症時の暴露量は前者で高い傾向

### 創薬：

- ・有機合成や発酵、動植物からの抽出・精製物質の生理活性スクリーニング
- ・動物組織からの抽出物質の遺伝子組み換え技術による大量生産（インスリン、成長ホルモン）
- ・抗体医薬（がん治療薬、リウマチ治療薬）

ゲノム解析による疾患関連遺伝子の探索→未知の受容体、新規整理活性物質の発見→画期的な新薬の創製

### 育薬：

- ・市販後の医薬品情報の収集と適正使用

ヒトゲノム情報、ゲノム解析技術の進歩→個々の患者に応じた薬効・副作用発現の体系的予測（ファーマコゲノミクス）  
個人の遺伝型に応じてくすりの選択、投与量の決定

医薬品の安全性・有効性確保の管理は、医薬品の開発過程から、製品化の過程、市販直後・市販後の使用実態下に至るまでの全過程において必要である

膨大な時間と資金を投入して開発された優れた医薬品が、多くの患者が犠牲になったために国民の信頼を失い市場から姿を消している。人的要因は工夫により解決可能（行政、企業、現場）⇒回避あるいは減少  
医薬品のリスクとベネフィットのバランスを考えた適正使用が不可欠。

物質 + 情報 → 医薬品

評価システム：市販後調査による再調査制度と再評価制度  
市販直後調査：新薬の重篤な副作用の集中的かつ迅速な収集による安全対策（平成13年10月1日）

# 臨床試験をめぐる倫理的問題

～ベルモント・レポートとその問題点～

2013/1/16

臨床試験部

川瀬貴之

## 本日の目次

### 第一章 基本講義の復習

- (1)ベルモント・レポートに至る倫理指針の整備
- (2)ベルモントレポートの内容

### 第二章 発展的な考察

- (1)医療における倫理の役割とは
- (2)ベルモント・レポートの問題点

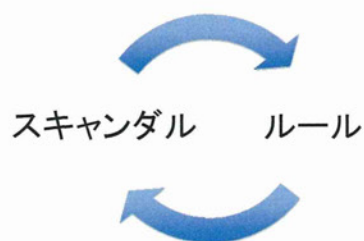
## 第一章 基本講義の復習

- (1)ベルモント・レポートに至る倫理指針の整備

## 歴史的に重要な研究倫理指針

- German Guidelines on Human Experimentation(1931, Prussia)
- Nuremberg Code(1947)
- Declaration of Helsinki(1964, WMA; 2008)
  - DoH: 1975年東京修正
- CIOMS Guidelines(1982; 2002)
  - The Council for International Organizations of Medical Science(国際医科学団体協議会)
- Belmont Report(1979, USA)

## 研究倫理の発展史



## 倫理的統制の歴史の変容

ヒポクラテスの誓い～近世(19世紀): 医師(研究者)個人の職業倫理観への盲目的信頼

20世紀中: 職能集団(ギルド)による自主規制の試み(ex. ヘルシンキ宣言)

20世紀後半: 第三者による評価・統制(IRB/REC)システムの導入

20世紀末・21世紀～:  
法令による統制強化へ?

## (1) ニュルンベルク綱領

- ・「合衆国対カール・ブラント」裁判  
米国一国による軍事裁判

1947年に7人が死刑、9人に終身刑を含む懲役刑が言い渡される

「許される医学実験」が遵守しなければならない10項目の「基本原則」



Karl Brant (1904-1948)



ダッハウでの超高度実験



ダッハウでの低温実験  
(September 1942)

## ニュルンベルク綱領の影響

- 国際人権規約 B規約(1966)
  - 第7条(拷問又は残虐な刑罰の禁止)  
何人も、拷問又は残虐な、非人道的な若しくは品位を傷つける取扱い若しくは刑罰を受けない。特に、何人も、その自由な同意なしに医学的又は科学的実験を受けない。
- 米国の裁判で使われた例
  - ・ 連邦地裁 「連邦の民事・刑事の両方で妥当する」
    - In re Cincinnati Radiation Litigation, 874 F. Supp. 796, 822(S.D. Ohio 1995)
  - ・ メリーランド州最高裁 コモン・ロー上の基準として採用
    - Grims v. Kennedy Krieger Institute, 782 A2d 807 (Md. 2001)
  - ・ 連邦最高裁
    - 連邦法の一部であることは認められたが、5対4で、軍人が、政府への損害賠償請求の根拠として綱領を使用することは認めなかった。  
U.S. v. Stanley, 483 U.S. 669(1987)

## (2) ヘルシンキ宣言の変遷

- ・ 1964年6月 フィンランド、ヘルシンキの第18回WMA総会で採択
- ・ 1975年10月 東京の第29回WMA総会で修正
- ・ 1983年10月 イタリア、ベニスの第35回WMA総会で修正
- ・ 1989年9月 香港、九龍の第41回WMA総会で修正
- ・ 1996年10月 南アフリカ共和国、サマーセットウエストの第48回WMA総会で修正
- ・ 2000年10月 スコットランド、エジンバラの第52回WMA総会で修正
- ・ 2002年10月 第53回WMAワシントン総会で修正  
(第29項目明確化のための注釈が追加)
- ・ 2004年10月 第55回WMA東京総会で修正  
(第30項目明確化のための注釈が追加)
- ・ 2008年10月 第59回WMAソウル総会で修正

### (3) ベルモント・レポートとは

- 1974年に米国で設置された、National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research(「生物医科学と行動研究における被験者の保護のための国家委員会」)によって公表された短い報告書
- 1978年9月30日に公表された

### 国家委員会の構成員

- 11人の委員のうち医学専門家は5人まで、と法律で決められていた

Kenneth John Ryan, M. D., Chairman, Chief of Staff, Boston Hospital for Women.  
Joseph V. Brady, Ph. D., Professor of Behavioral Biology, Johns Hopkins University.  
Robert E. Cooke, M. D., President, Medical College of Pennsylvania.  
Dorothy I. Height, President, National Council of Negro Women, Inc.  
Albert R. Jonsen, Ph. D., Associate Professor of Christian Ethics, Pacific School of Religion.  
David W. Louisell, J. D. \*, Professor of Law, University of California at Berkeley.  
Donald W. Seldin, M. D., Professor and Chairman, Department of Internal Medicine, University of Texas at Dallas.  
Eliot Stellar, Ph. D. \*, Provost of the University and Professor of Physiological Psychology, University of Pennsylvania.  
Robert H. Turtle \*, LL. B., Attorney, VomBaur, Coburn, Simmons & Turtle, Washington, D. C.

### (2) ベルモント・レポートの内容

### ベルモント三原則

#### (1) 人格の尊重

- インフォームド・コンセント

#### (2) 善行

- リスク・ベネフィット評価

#### (3) 正義

- 研究参加者の選択

### (1) 「人格の尊重」とは

個人は自律的な主体として扱われるべきである

- 個人の自律を認めよ

自律性が減じている人は保護を受ける権利がある

- 自律性が減じている人を保護せよ

### 「自律的な人」とは

- 自分自身の目標について深く考えることができ、その考えに従って行動できる人

### 「自律的な人を尊重する」とはどういうことか

- 自律的な人の熟慮した上での意見や選択を尊重し、明らかに他者を害することになるのではない限りは、その人の行動を妨げることを差し控える、ということ
- 「対象となる人は自発的に、十分な情報が与えられて研究に参加すべきである」

### 「自律能力に欠ける人を尊重する」とは

- 未成熟あるいは能力を欠く人々を尊重するとは、その人たちの成熟の過程において、あるいは能力を欠いている間においては、その人たちに保護を与えることを必要とするかもしれない

### (2)「善行」とは

#### 個々人の決定を尊重し害から保護

- 害を為してはならない

#### 福利を確保するよう努力する

- ベネフィットの可能性を最大にし、リスクの可能性を最小にする

### (3)「正義」とは

#### 分配の公正さ

- 研究における負担と利益を公平に分担すること
- 被験者の選択

### 社会的弱者

- 繰り返し研究対象になっているグループにだけ負担を負わせない
- 黒人→タスキギー梅毒研究
- 兵士→プルトニウム注入実験
- 子供→ウィローブルック肝炎実験
- 学生・研究員→黄禹錫によるヒト胚性幹細胞捏造事件(研究員が卵子提供)  
ジョンズ・ホプキンス大での死亡事故  
(テクニシャンのEllen Rocheが試験に参加して死亡)

他の被験者では研究の意味をなさないような場合でないかぎり、社会的弱者を研究に組み入れるべきではない。

## 第二章 発展的な考察

• (1) 医療における倫理の役割とは

医学研究、特に「臨床研究」における葛藤と倫理

医学研究者の役割

=社会的要求への応答  
=未だ見ぬ「将来の患者」の必要と最善の利益のために行為

1. 治療法の改善、病因等の理解
2. 普遍化可能な知識の獲得
3. 社会全体の健康増進

医療従事者の役割

=個別具体的な「目の前の患者」の必要と最善の利益のために行為



責務相反

「研究」と「治療」の本質的差異

研究の本質

- 普遍化可能な科学的知識の獲得と構築
- 社会や「将来の患者」への貢献
- 介入法:  
効能・AEは未確立(未知)
- 目の前の被験者は目的達成のための「手段」
- 倫理的非対称性の構造:  
判断者=被験者本人=リスク負担者 受益者

治療の本質

- 「目の前の患者」の最高の利益(=治療、高QOL、高ADL)の確保・追求
- 介入法:  
効能・AEは確立(既知)
- 目の前の患者は「目的それ自体」
- 倫理的対称性は保持:  
判断者=患者本人=リスク負担者 受益者

目の前の個別の被験者を、  
医学・医療の進歩と未来の患者のために「使う」ことについて、  
どこまでならば正当化できるのか???

ちなみに・・・(倫理学の性質について)

- よくある問いかけ(1)  
「～の問題について、倫理的に言うとうどうなるのか？」

「倫理」という単一の思想があるわけではない。  
「〇〇主義、××主義だと、この問いにどう答えるのか」という質問には、答えがあるが、「倫理は、どうなのか」という質問は、答えのない、空虚な問いである。  
倫理とは、全く異なる多様な主張をする諸々の主義の総称にすぎない。異なる主義は、同一の問いに全く異なる答えを持つ。

- よくある問いかけ(2)

「これさえ守っておけば大丈夫という、最低限のルールは、どういうものか」

「法制度を紹介してほしい」という要請は、このような問題意識に由来する。確かに、法的責任を問われないために、法制度の理解を深めることは重要。ただし、これは倫理の問題というより、事務・行政・教育という問題。  
倫理的な問題とは、制度や常識がどのようなものに関係なく、「本当に道徳的に正しいとはどういうことか」を徹底的に批判考察すること。とりえず、ルールを暗記して墨守するのは、思考停止である。

- (2) ベルモント・レポートの問題点

## (1) 人格の尊重原理の問題点

ベルモント・レポートの原則

「自律的な者は自己決定、自律的でない者はパターナリズム」

そもそも・・・**パターナリズム(父権的干渉主義)**とは何か

被干渉者の自由その他の「利益」を理由として、その「自由」を制限すること。シートベルトやヘルメットの着用など。

本人に判断能力が乏しいと判断されるとき、あるいは高度な専門知識が必要とされるケースで、問題となる。

パターナリズムが許容される条件とは？

1) 本人の同意があるとき

→果たして、同意によって自由を放棄することはできるのか？

2) 本人の同意が取れない場合、「もし彼に合理的判断能力があれば、このように判断していたであろう」という道理にかなった反実仮想ができるとき

また、実際に本人のICを取るのか、あるいは近親者や医師などによるパターナリスティックな代理同意にするのかを判断する場合、本人の判断能力が、どの程度であれば、「自律的」とされるべきなのか？

十分なコミュニケーション能力など・・・

治療・研究(の意義や方法)に関する専門的知識の理解能力までは必要ないだろう。

もし、本来ならば自律能力ありと判断されるべき人が、誤って自律能力なしと判断されれば・・・

→重大な「**自律・自由・自己決定権**」の侵害になる

もし、本来ならば自律能力なしと判断されるべき人が、誤って自律能力ありと判断されれば・・・

→向う見ずな判断で、本人の「**福利**」が重大に損なわれるかもしれない

## 小括

被験者・患者の自律を尊重するためには、その自律的判断能力の有無を、慎重に見極めなくてはならない。

自律能力の有無の判定の手続きの厳格化という「**技術的**」な問題

そもそも、どれだけの知的能力があれば自律的と判定されるべきか、という「**規範的**」な問題

自律的なICを得ていれば、どこまで高レベルのリスクも引き受けることが可能であるべきかという規範的問題

→自律的な人は自分の利益を合理的に判断できるはずだから、不合理に高レベルのリスク引き受けには同意しないだろうという前提は成り立つだろうか？

同意撤回する権利を、あらかじめ放棄する(離脱不可能にする)ことを最初の同意文書に盛り込むことは許されるべきかという規範的な問題

「自律の尊重」というだけでは、その意味や内容が、極めて漠然としているので、その明確化が必要。

## (2) 善行原理の問題点

さて、自律的判断能力の判定が完了したとする。とりわけ、パターンリスティックな介入が求められる場合、**善行原理**、つまり被験者にとっての**ベネフィットを増加し、リスクを最小化**することが要求される。

臨床研究に伴うリスクやベネフィットには、様々な種類が存在する(金銭的・健康的・精神的・・・)

そこで問題は、リスク・ベネフィット評価において、どのような種類のリスクやベネフィットを考慮すべきか、あるいは考慮すべきでないか？

CIOMS(国際医学団体協議会)の「人を対象とする生物医学研究についての国際倫理指針」によれば

ICができない個人は、研究参加の利益が**直接的なものである**ときのみ、参加可能とすべきである。

この「直接的」とは、どういう意味だろうか。

### ①「近い将来」という意味

遠い将来の利益を与えることは、現在のリスクを賦課することの正当化にならない

### ②リスクと「同種のもの」とであるという意味

研究参加に伴うリスクの種類は、「健康上」のものである。これに対して、金銭・物質的・精神的利益で補償することは、同種のベネフィットの提供ではないので許されない。

もちろん、交通費・日当などは、経済的コストを経済的ベネフィットで補償するものだから、許されるだろう。

③当該研究に「由来する」ものであるという意味  
同じ種類である「健康上」のベネフィットで補償するとしても、当該研究とは、何の関係もない検査・診察・治療を提供することで足りるとすれば、それ自体に何のベネフィットもなく、単に高リスクの研究への参加を、全く関係ない治療をいくつでも付属させることで正当化することになってしまうだろう。



### (3) 正義原理の問題点

- 正義には、2つの種類がある
  - i) **分配的正義**・・・利益と負担の公正な分配を要求する。公正さとは、具体的に何かについて、多様な考え方が存在する。
  - ii) **矯正的正義**・・・公正な分配状態が、不正に侵害された場合、公正な原状の回復を要求する。損害賠償・補償など。

ベルモント・レポートの正義原理とは、このうち、分配的正義に関するものである。

第一章でも見たとおり、民族的少数派・囚人・小児など社会的弱者を不均等に参加させることが不公正であることは言うまでもない。

問題は、何を考慮して、被験者を選択することが公正なのか。(被験者の属性のうち、考慮すべきものと、すべきでないものがある。それを分類するのが、ここでの課題。←(2)のリスク・ベネフィット評価における問題意識と似た構造。)

- ①研究にとって**レレバンス(重要性・関連性)**がある属性は考慮されるべき。

タスキギー事件では、単に恣意的に黒人が対象となったが、グローバル治験の場面では、人種などの属性が重要な科学的意味を持ちうる。黒人のことを調べる科学的必要性があるなら、黒人だけを選んでよいだろう。性別・年齢など、他の属性についても...

もちろん、科学研究の背後に不正な意図(たとえば人種主義)がある場合は、問題だが...

- ②レレバントな属性において違いのない被験者候補の間では、**公正な手続き**が求められる。

ただし、手続的正義の具体的な内容は、多様なものが考えられる。

- i) くじ引き(ただしコストがかかりすぎる)
- ii) 早い者勝ち(待ちリスト方式。これが現状であり、現実的)・・・などなど

ともかく、考慮すべきではない属性の違いをランダム化する手続が必要。

### おわりに

- 倫理は、「ここまで考えれば十分」ということはない
- 倫理は、とにかく常識(法制度を含む)を疑う態度が必要
- 倫理は、その思考方法に革命的進歩は少なく、2000年前の古典を読むことが重要
- 倫理は、医療を規制するためだけのものではない(これは大きな誤解)。多様な主義の間で神学論争しているが、その中には、医療の進歩を後押しし、正当化する主義もたくさんある。
- 倫理は、...

「平成24年度臨床研究応用」

臨床統計学

柴田大朗  
国立がん研究センター 多施設臨床試験支援センター  
薬事安全管理室長

2013/1/31 18:00-19:30 千葉大学医学部附属病院

Outline

- 新しい治療法の評価に関する米国での苦い経験
- 交絡とランダム化
- 統計にまつわる身近なpitfall
- 生存時間解析の基礎

2

治療法評価の失敗例

CLINICAL TRIALS PERSPECTIVE Clinical Trials 2006; 3: 149-153

Listen to *all* the voices: an advocate's perspective on early access to investigational therapies

Musa Mayer<sup>1</sup>

*In this issue of the journal, The Society for Clinical Trials, whose members are clinical investigators and regulators in academia, government and industry, has published the rationale for their opposition to US Senate bill S. 1936, the so-called "ACCESS" act, which would allow the marketing of therapies that have undergone only Phase I studies (dose finding and toxicity testing) in humans, in patients with "serious and life-threatening" disease who have exhausted other treatment options. While this bill reflects the views and efforts of some cancer advocates and their supporters, it does not represent the beliefs of the cancer advocacy community overall. The essay that follows is inspired by the need for other voices from the patient and advocacy communities to be heard in this debate.*

<http://ctj.sagepub.com/cgi/content/abstract/3/2/149>

1990年代の米国における、乳がんへの大量化学療法の受け止められ方

- Mayerによると、1990年代、乳がんに対する移植を伴う大量化学療法は、
  - 用量と奏効に関連性があった
  - Phase IIの結果、致死的なリスクはあるものの高い奏効率が確認された
  - ヒストリカルコントロールとの比較でも良好な成績であった
  - 1995年に小規模ではあるがランダム化比較試験の結果で良さが証明された結果があった

ことから、1990年代末までに4万2680人の米国人が臨床試験外でこの治療を受けたと言われている

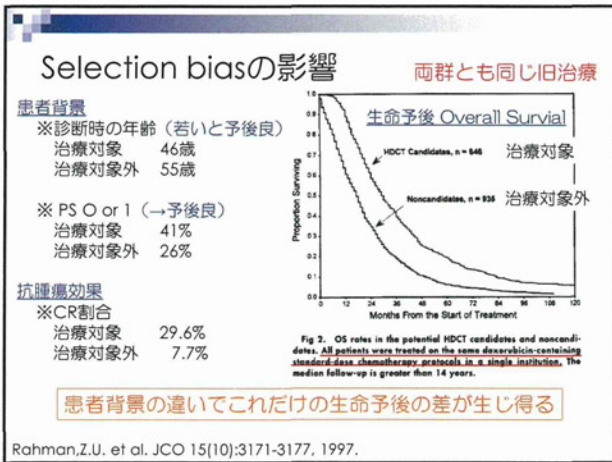
Mayer, M., Clinical Trials 3(2): 149-153, 2006.

2000年に公表された大規模なランダム化比較試験結果

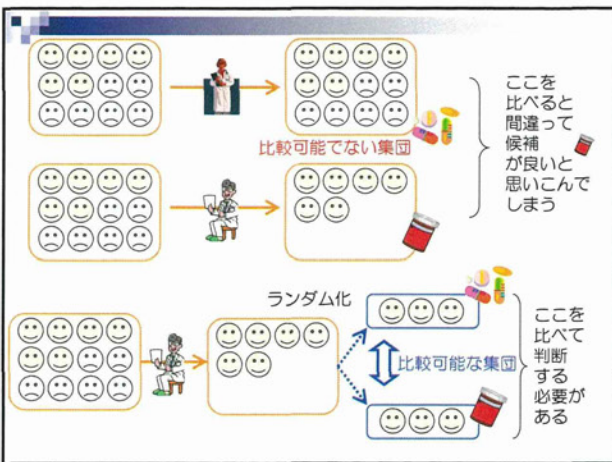
NEJM 342(15):1069, 2000.

- 初期には5人に1人の患者が、移植技術が進歩した後も2~5%の患者が治療関連死したといわれている
- ヒストリカルコントロールとの比較で非常に良好な成績であったことは、selection biasによるものと考えられた
  - Rahman et al. (1997)
- 2000年に大規模な比較試験の結果、効果がないことが判明した
  - Stadtmauer et al. (2000)
- 2006年1月時点で、Cochrane systematic reviewで生存への寄与はないとされている

— Mayer, M., Listen to all the voices: an advocate's perspective on early access to investigational therapies, Clinical Trials 3(2): 149-153, 2006.

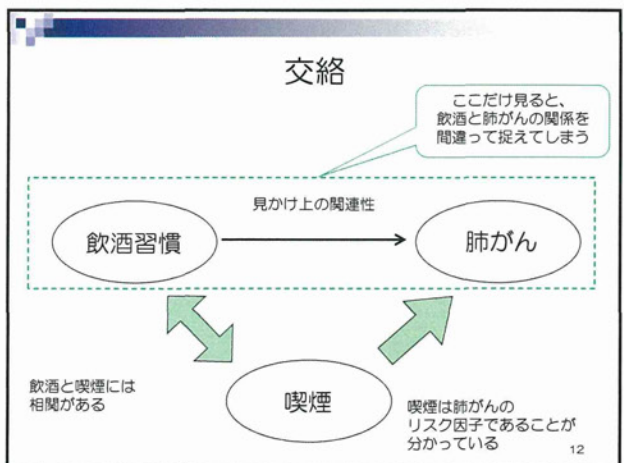


- 「治療を受けた患者のデータ」と「治療を受けていない患者の(ヒストリカル) データ」
  - これらの比較によってある治療法の良さを主張する際、
    - 「治療あり」
    - VS
    - 「治療なし」
 の比較をしているつもりで、実際には、
    - 「治療を受けることの出来る患者の状態」
    - VS
    - 「治療を受けることの出来ない患者の状態」
 の比較になっていることが多々ある
- 8

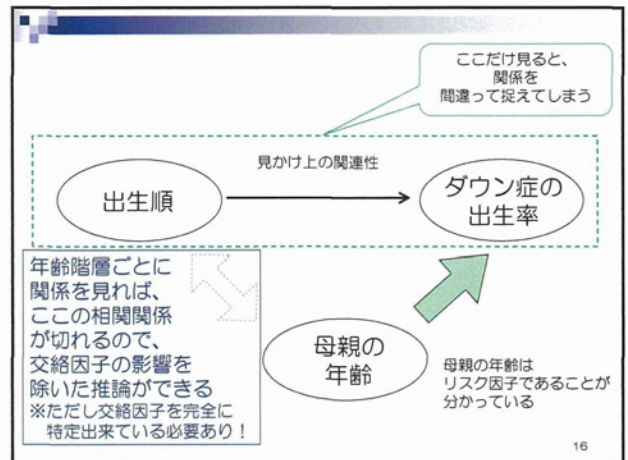
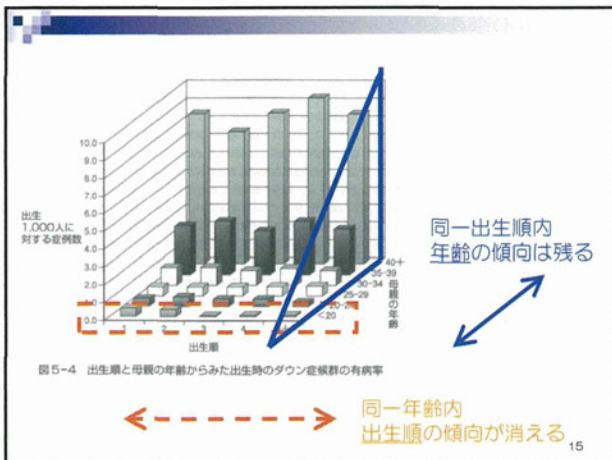
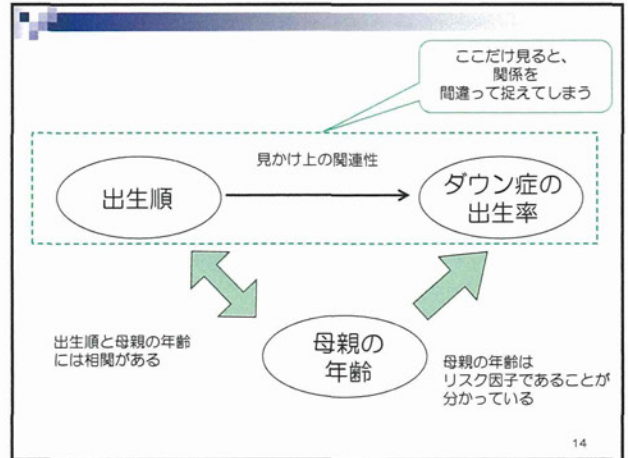
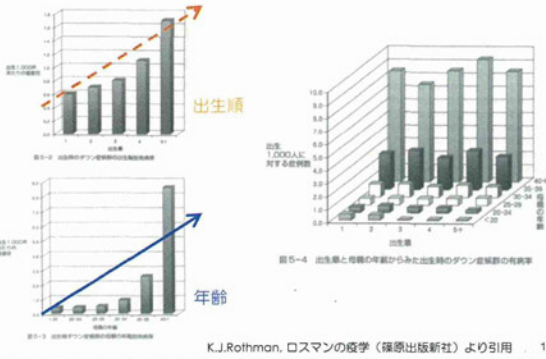


- ### Outline
- 新しい治療法の評価に関する米国での苦い経験
  - 交絡とランダム化
  - 統計にまつわる身近なpitfall
  - 生存時間解析の基礎
- 10

- ### 臨床研究に伴う誤差・バイアス
- 偶然誤差 (random error)
    - nを増やすと小さくなる
  - 系統誤差 (systematic error)
    - nを増やしても小さくならない
    - バイアス (bias、偏り)、交絡など
- 「nが大きいから正しい結果を導くことができる」  
 ・・・・は早計
- nをどれだけ増やしても正しい結果が得られないことがある!
- 11



## 交絡の例、分析方法の例



## いろいろな交絡の調整方法

- 交絡因子が同一の値をとるサブグループ内で比較する
- 層別解析
  - 層毎に差を要約し、全体で併合する
- 標準化する
  - 比較するグループ間で交絡因子の分布が同じになるように調整する
  - 年齢調整死亡率など
- 多変量解析で調整する
- マッチングをとる
  - コホート研究の場合は問題なし
  - ケース・コントロール研究の場合はオーバーマッチングなど、難しい問題あり
- 傾向スコア (Propensity Score) 等、因果推論の方法を応用する

17

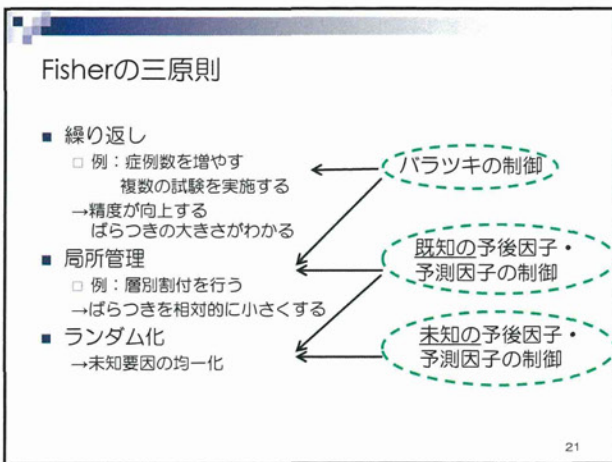
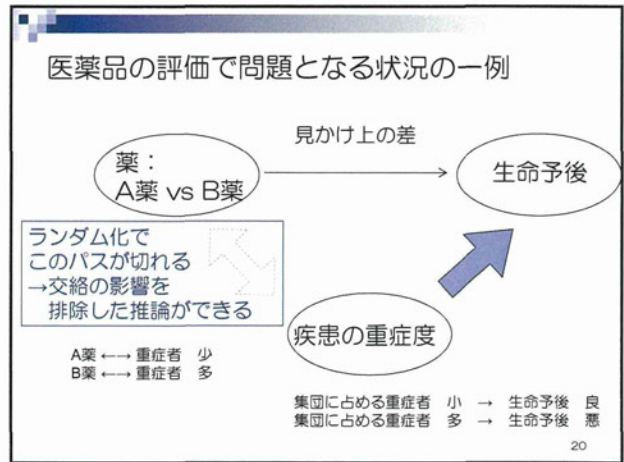
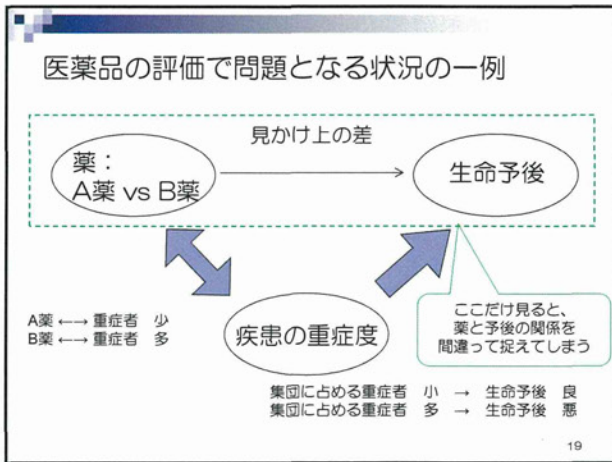
## 交絡を解析時に調整することの限界

- 交絡因子が特定できなければ、調整のしようがない・・・

### → ランダム化比較試験

- 既知の交絡因子の影響だけでなく未知の交絡因子の影響も排除できる
- 系統誤差を偶然誤差に転化する

18



- ### Outline
- 新しい治療法の評価に関する米国での苦い経験
  - 交絡とランダム化
  - 統計にまつわる身近なpitfall
    - データの集め方 (サンプリング方法)
    - データの取り扱い
    - グラフ・数値の読み方
    - 比較対照の取り方
  - 生存時間解析の基礎
- 22

### データの取り方と解釈

- 下記の2×2分割表から、ある薬を服用した場合の問題発生割合は98%(90/92) 非服用の場合の問題発生割合は11%(22/208) と解釈して良いか?

	問題あり	問題なし
服用	90	2
非服用	22	186

23

### データの取り方と解釈

	奏効	非奏効
因子 (+)	a	b
因子 (-)	c	d

- 同じ様にみえる2×2分割表でもデータの取り方によって解釈できる内容が変わる
  - ① 奏効例、非奏効例を集め、因子±の構成割合を調べる
    - 陽性的中位・陰性的中位の算出は出来ない
    - 因子 (+) 例のうち何割が奏効するかは計算出来ない
  - ② 因子 (+) 例、(-) 例を集め、奏効例、非奏効例の構成割合を調べる
  - ③ 一定数の患者を集め奏効、非奏効、因子±の構成割合を調べる

※臨床試験・臨床研究結果の解釈を行う際には、データがどのように取られているかを 確認・理解することが必須

24

## 歴史をひもといてみると・・・

- サリドマイド問題で有名なLenz報告には統計学的に問題があり、このデータからは結論を出せない、と主張した統計学者がいた
  - 日本医事新報1969.5.17
  - 「いわゆるサリドマイド問題に関する統計的考察」

25

	問題あり	問題なし
服用	90	2
非服用	22	186

(前出の表)



	奇形	非奇形	
服用	90	2	92
非服用	22	186	208
	112	188	300

(下記論文中の表)

「この表の非服用の行を横に見ると計二〇八例中奇形児が二例出生したことになるが、この一〇%以上の高率はどうみても到底納得し難いデータであろう。(略) このような四分表に $\chi^2$ 検定を行い(略) 相関ありといつても、もとのデータが根本的に怪しいのであるから、この検定自体およそ意味のないものになってしまうのである。」

前述の日本医事新報1969.5.17号から引用

26

	奇形	非奇形	
服用	90	2	92
非服用	22	186	208
	112	188	300

—Lenzの表—

ケース・コントロール研究であるので、  
22/208 (非服用者における割合)  
を計算することは不適切

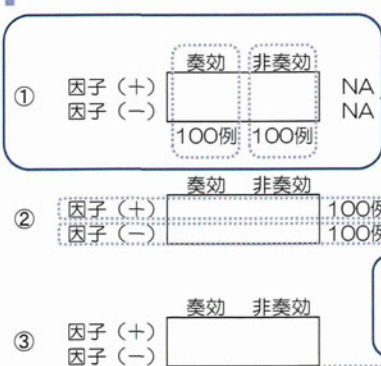
※データのサンプリング(抽出)方法を  
無視した解析を行ってはいけない

27

## ちなみに・・・

- この問題の論文を医学部の4年生に読ませ、当否について論じるように課題を出したところ、課題を提出した79人中、
    - 60人 → 統計学者の間違いに全面的賛意
    - 8人 → 疑問を感じつつも、同意見
    - 11人 → 適確に批判
- という結果であったとのこと。結果が分かっているにもかかわらず問題を指摘するのは困難なようです。

日本計量生物学会ニュースレター第103号(2010年7月31日発行)  
巻頭言「サリドマイド」 柴田義貞(長崎大学)  
<http://www.soc.nii.ac.jp/jbs/newsletter/all/kaiho103.pdf>



このデータの取り方では、因子(+)  
のうち、何割が奏効  
するかを算出する  
ことは出来ない

①のデザインで  
陽性適中度、陰性的  
中度を計算している  
ケースは多い

29

## Memo: 抗菌薬の皮内反応テストの意義をめぐり臨床家との対話

- ある抗菌薬の皮内反応テストの感度・特異度は比較的良い(80~90%程)
- 「診断精度がこのような高いならば、副作用を回避するために、検査を必須とすべき」(臨床家の意見)
- 実際に使ってみると、殆ど正しい判定が下されないことが臨床現場で問題となった
- 「統計はしょせん統計であり、そのようなものに振り回される方がナンセンスである」(別の臨床家の意見)

→1万人に数人という副作用を感度・特異度9割程の検査によって高い確率で見いだせないことは、陽性的中度を計算すれば判る

→統計が役に立たないのではなく、臨床的意思決定に必要な情報が何であるのか(必要なのは感度ではなく陽性的中度)が明確にされないまま、それに対応しない統計量の見た目の数字の善し悪しで検査の要否を議論したこと、つまり、統計の使い方が間違っていたことが問題の本質

30

## 感度、特異度と陽性的中度

※感度90%、特異度95%の検査の陽性的中度は・・・

	要因+	要因-	
検査+	90	5	94.7% (←要因+の人が集団に50%いる時)
検査-	10	95	
	100	100	

	要因+	要因-	
検査+	90	45	66.7% (←要因+の人が集団に10%いる時)
検査-	10	855	
	100	900	

	要因+	要因-	
検査+	90	495	15.4% (←要因+の人が集団に1%いる時)
検査-	10	9405	
	100	9900	

・同じ検査でも、対象集団が異なれば判定の確からしさは異なる  
 ・「奏効例100名、非奏効例100名を集めた研究」の結果を見て、検査+の人の奏効確率が94.7%になると判断するのは早計

31

## Outline

- 新しい治療法の評価に関する米国での苦い経験
- 交絡とランダム化
- 統計にまつわる身近なpitfall
  - データの集め方 (サンプリング方法)
  - データの取り扱い
  - グラフ・数値の読み方
  - 比較対照の取り方
- 生存時間解析の基礎

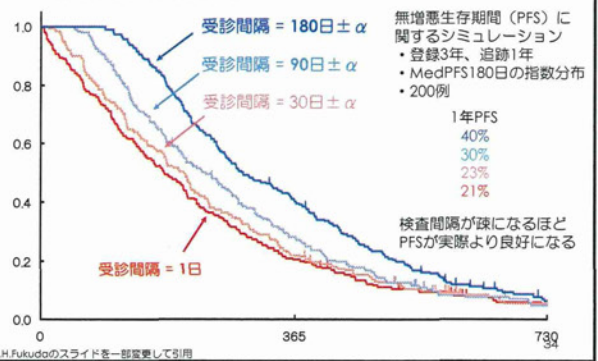
32

## 無増悪生存期間

- 増悪までの期間を測る
  - 増悪した人は増悪と判定された日でイベントとして扱う
  - 観察終了までに増悪しなかった人は観察終了時点で「打ち切り」としてあつかう
    - ただし、検査しなければ、本当は増悪していても増悪なしと評価されることになる

33

## 検査の間隔によって生じるバイアス — 無増悪生存期間 —



## 検査間隔が密な場合



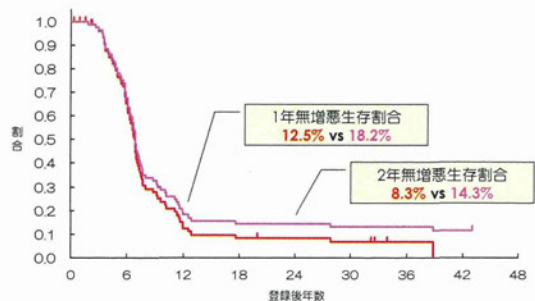
## 検査間隔が疎な場合



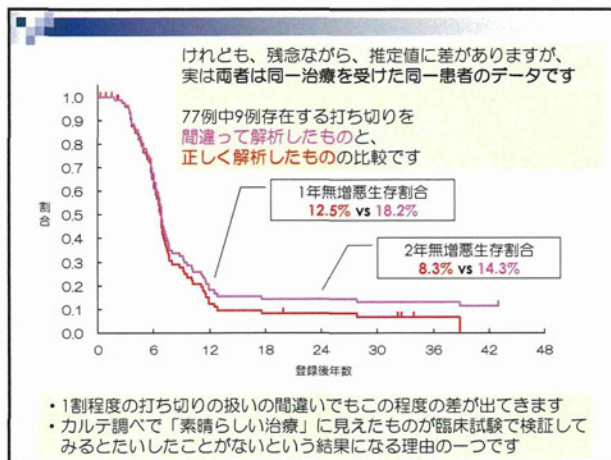
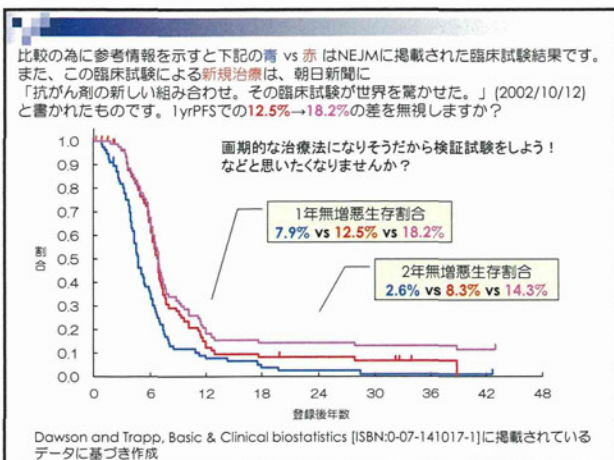
- ・ 検査が疎であるほど治療成績が良く見える
- ・ 検査間隔が群間で異なっている場合、PFSの評価にバイアスが生じる

35

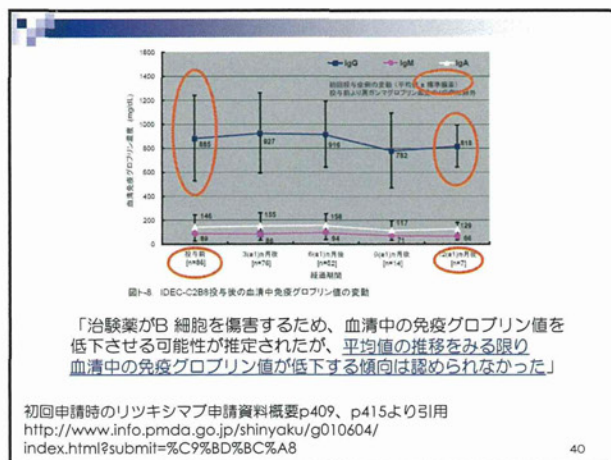
## ある予後不良ながんの無増悪生存期間



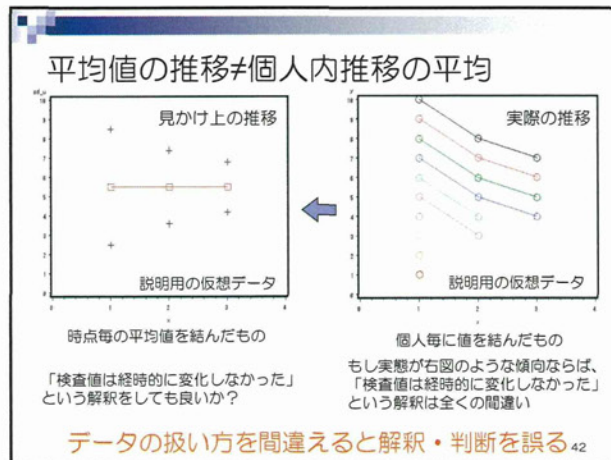
統計学的に有意差はありませんが、この差を臨床的に大きいと思いますか？



- ### Outline
- 新しい治療法の評価に関する米国での苦い経験
  - 交絡とランダム化
  - 統計にまつわる身近なpitfall
    - データの集め方 (サンプリング方法)
    - データの取り扱い
    - グラフ・数値の読み方
    - 比較対照の取り方
  - 生存時間解析の基礎
- 39



- リツキシマブ (遺伝子組換え) 審査報告書H15.8.28
  - 「免疫グロブリンの推移については、indolent B細胞性非ホジキンリンパ腫対象国内臨床第II相試験と本試験間に相違はなく、いずれの試験においても減少傾向は認められなかった (審査センター注: 申請者の主張は不適切な集計方法に基づくものであり、免疫グロブリンの推移については、平均値の経時推移からは減少傾向が見られないものの、個別症例ごとの推移を確認すると減少傾向を示している。この点について申請者に指摘したところ、申請者から、低下傾向の見られる症例があり、IgMの減少を示した症例が多いことが説明された。)」
- 効能追加のための申請時の規制当局 (PMDA) の判断  
審査報告書p26から引用  
<http://www.info.pmda.go.jp/shinyaku/r03.html>
- 41

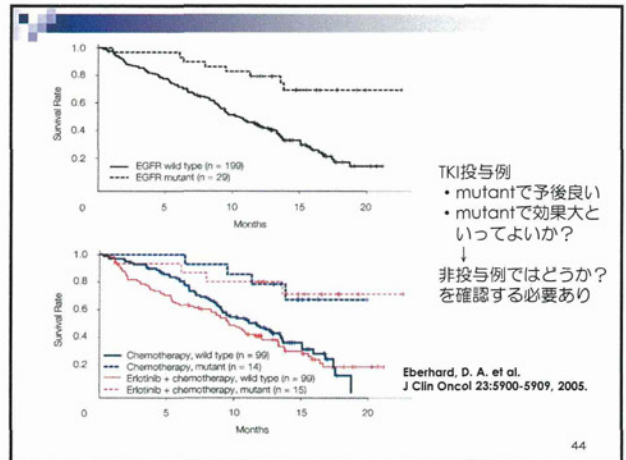




## Outline

- 新しい治療法の評価に関する米国での苦い経験
- 交絡とランダム化
- 統計にまつわる身近なpitfall
  - データの集め方 (サンプリング方法)
  - データの取り扱い
  - グラフ・数値の読み方
  - 比較対照の取り方
- 生存時間解析の基礎

43



44

## 「比較の対照は何か？」を明らかにする

- もし調べたいことが「TKIはEGFR mutantで効果が大きい」(\*)であれば、その比較対照はTKIが投与された wild typeの集団ではない
- 「TKI投与例でのEGFR mutantとwild typeの差」と「TKI非投与例でのEGFR mutantとwild typeの差」とを比較することによって(\*)を調べることができる

※nがどれほど大きくとも、比較対照が適切に選ばれていなければ、正しく結論を導くことはできない  
 →適切に問題を定式化する必要がある

45

## 医学研究には常に比較対照がある

- ランダム化比較試験は当然試験内部に比較対照がある
- シングルアームの試験の場合はヒストリカルコントロールとの比較をしている
- ケースレポートですら比較対照がある(!)
  - 副作用報告に対して因果関係を判断するときには、当該疾患での一般的な臨床経過を対照として推論している
  - しかし、このことは当たり前すぎて意外と見落とされている  
 →非医療者に誤解を招く原因になるので注意が必要
- ケースレポートを医師が個別に評価することが有用だからといって、ケースレポートの情報をまとめた要約値が有用だとは限らない(データを読む側が要約値の比較対照を持っていない場合、役に立たない)

46

## 比較対照を明示しないことによるコミュニケーションギャップ

- がんの専門家が「この分子標的薬は非常に安全である」という判断を下す際に、implicitに用いている比較対照は重篤(時には致命的)な副作用が必発の既存の細胞障害性の薬
- 一般の方の多くは抗がん剤で治療を受けたことはないで、そのような人がimplicitに用いる比較対照は重篤な副作用が非常にまれにしか出ない薬  
 →安全だと言われたのに副作用が出た、だまされた、という印象を持たれる

※データに基づきどのような推論をして結論を導いたのかを常に意識することで(≠統計を正しく使うことで)、結論を導く上でcriticalな情報・伝えなければならない情報が何か明らかになる

47

「間質性肺炎が四十一例あり、うち十六人が死亡。」  
 「発症は女性が多く、死者のうち間質性肺炎の既往歴がある患者が十二名を占めた。」

2005/2/25 日本経済新聞

リウマチ治療薬 副作用で患者死亡16人

女性であることはリスク因子か？  
 間質性肺炎の既往歴があることはリスク因子か？

副作用例に占めるある要因の構成割合が高いか否かは、当該因子がリスク因子か否かを判断する根拠にはなっていない

↓

データに基づいた判断をしているようで、実は、何か他の情報を加味している

何を加味して判断していますか？

## 個人と集団の情報の違い

- ケースレポートは経験豊富な臨床医が個のデータとして読む限り、解釈可能
  - しかしケースレポートの蓄積を統計的に取り扱う場合に、解釈不可能になる状況がある（例：タミフルを服用し、異常行動を起こしたケースのみが蓄積したデータなど）
    - 個のデータは臨床医の経験の中に比較対照がある
    - 集団のデータは、適切な比較対照が選択できない場合、全く解釈不可能
- 統計的にデータを扱うのであれば、適切なデータの取り方（比較対照の置き方を含む）を計画時に決める必要がある

49

## Outline

- 新しい治療法の評価に関する米国での苦い経験
- 交絡とランダム化
- 統計にまつわる身近なpitfall
- **生存時間解析の基礎**

50

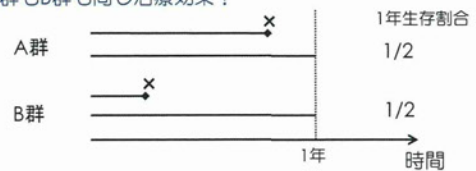
## 生存時間データを解析する際に必要となること

(仮説をたて試験デザインを決定)  
(解析方針を決定) → 解析 → (解釈)  
(調査を実施) (伝達)

- **記述と要約：生存時間の記述**  
→生存関数のKaplan-Meier法による記述
- **検証的解析：生存関数の比較**  
→群間で生存曲線が異なるかどうかの検定 (Logrank検定など)
- **探索的解析：**
  - 観察された差（されなかった差）の説明
  - 生存関数に影響を与える要因の探索  
→Cox回帰などのモデルを用いた解析

51

## A群もB群も同じ治療効果？

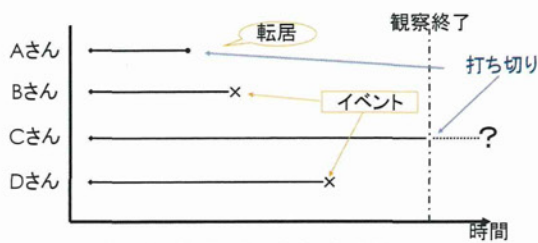


## A群もB群も同じ治療効果？ B群の方が劣る治療？



52

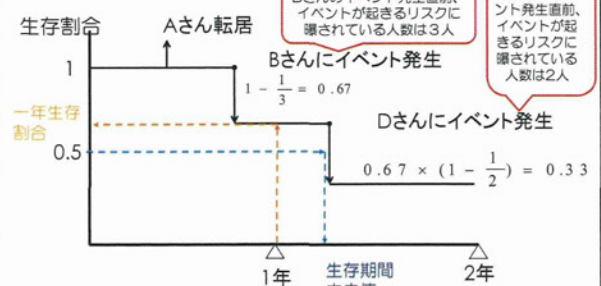
## 簡単な例



「1年後の時点での生存率は？」  
「median survivalは何年？」

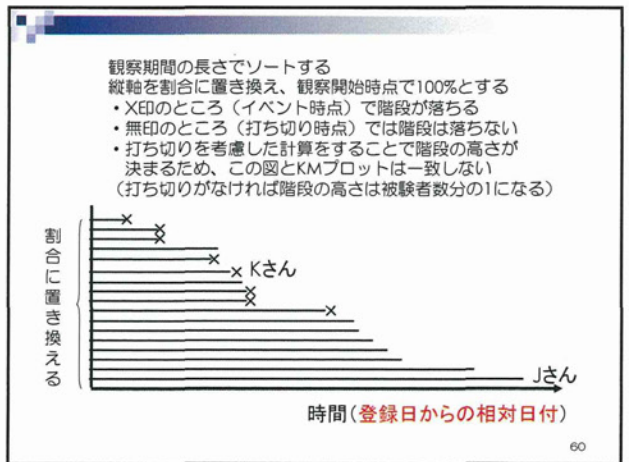
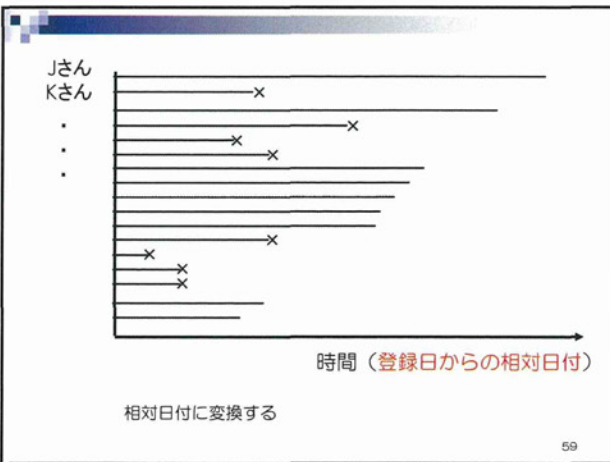
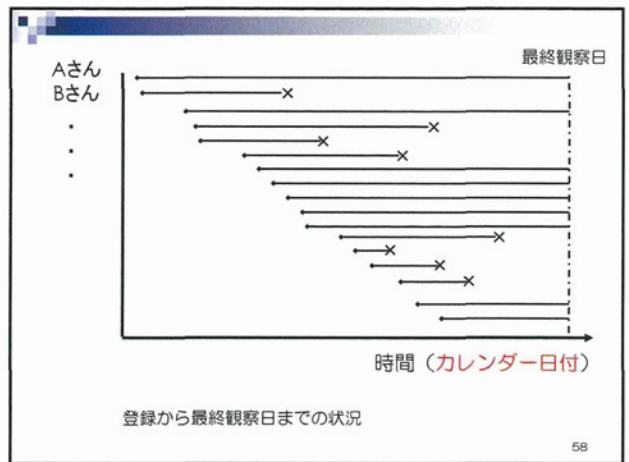
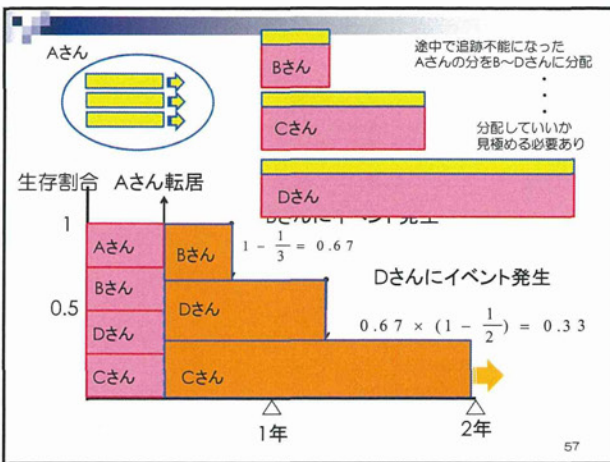
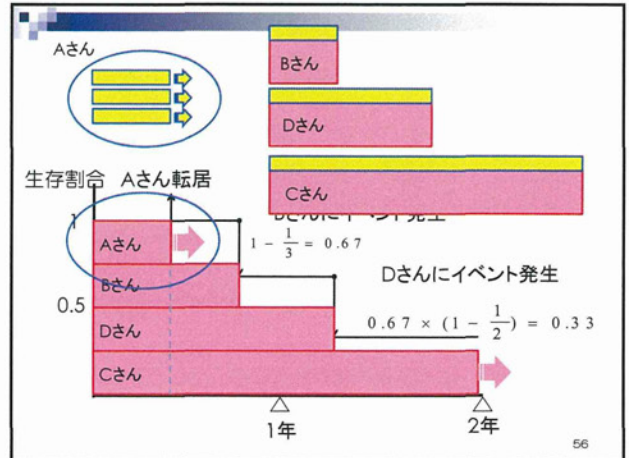
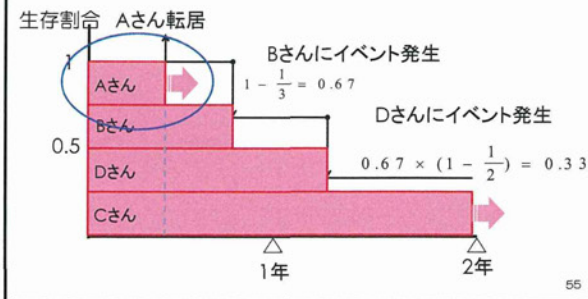
53

## Kalpan-Meierプロット

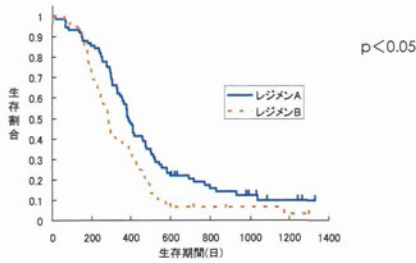


54

# Kalpan-Meierプロットのイメージ図



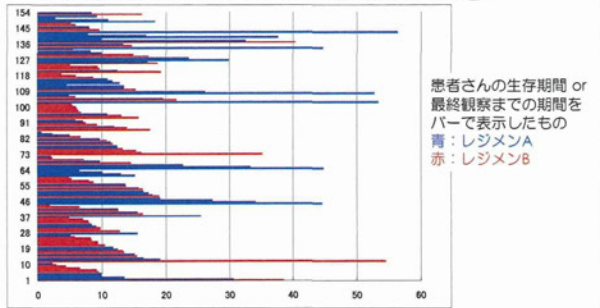
### 小細胞肺癌に対する試験結果①



Noda et al. NEJM, 2002.

Dawson and Trapp, Basic & Clinical biostatistics [ISBN:0-07-141017-1]に掲載されているデータに基づき作成

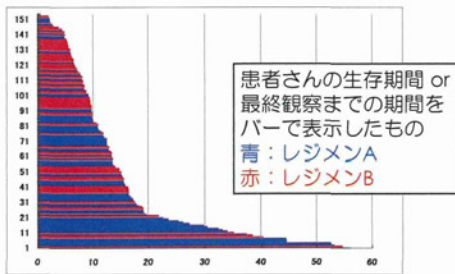
### 小細胞肺癌に対する試験結果②



Noda et al. NEJM, 2002.

Dawson and Trapp, Basic & Clinical biostatistics [ISBN:0-07-141017-1]に掲載されているデータに基づき作成

### 小細胞肺癌に対する試験結果③



Noda et al. NEJM, 2002.

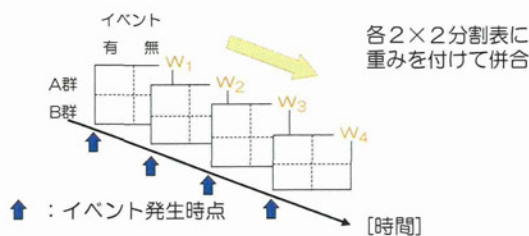
Dawson and Trapp, Basic & Clinical biostatistics [ISBN:0-07-141017-1]に掲載されているデータに基づき作成

### Logrank検定・一般化Wilcoxon検定

- イベントが発生した時点の情報をもとに群間の比較を行う（二群間で生存関数に違いが有るかを検定）
- 二つの手法の違い
  - 各時点の情報の重みの違い
  - Logrank検定は群間のリスク比の推定が可能

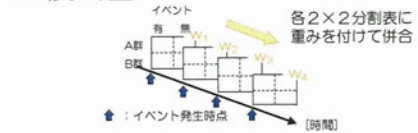
64

### Logrank検定・一般化Wilcoxon検定の模式図



65

### Logrank検定・一般化Wilcoxon検定の模式図



○各時点毎に実際に観察されたイベント数と期待イベント数との差を取り、重みをつけて併合

$$A \text{ 群の「期待と観察とのズレ」} : \sum w_i (d_i - e_i)$$

○この「ズレ」を帰無仮説の元での「ばらつき」で割る

$$\frac{\sum w_i (d_i - e_i)}{(H_0 \text{ の下でのばらつき})}$$

66