

3歳児(95cm, 14kg)におけるテオフィリン投与量
成人(170cm, 60kg)投与量600mg/日とした場合

換算式		1日投与量 (mg/日)	体重あたりの1日投与量 (mg/kg/日)
年齢基準	Young	120	8.6
	Augaberger (I)	192	13.7
体重基準	Augaberger (II)	186	13.3
体表面積基準	Crawford	208	14.9
	von Harnack	200	14.3
Pediatric Dosage Handbook® #1		280-336	20-24
日本の添付文書		224	16

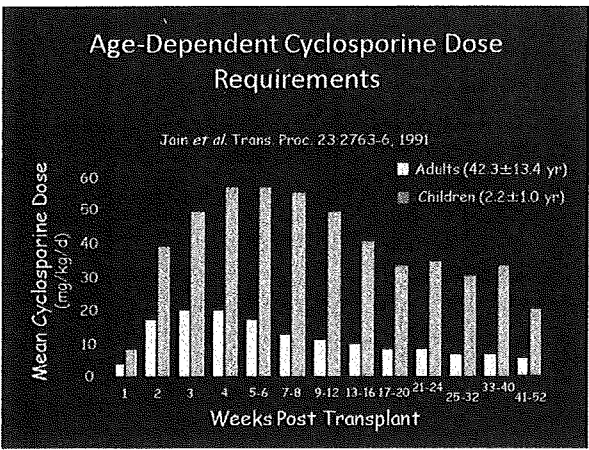
*1. 米国の添付文書に 2mg/kg/日(体重別)が成人状態の血中濃度セーブ10-25μg/mlとの記載あり。

開原園の外産小児科学 改訂5版

Inadequacy of traditional dosing schema

Comparison of total daily gabapentin (Neurontin®) maintenance doses calculated via "traditional" and current dosing guidelines for a 4-year-old, 17 kg child

	Young - Cowling	Clark	BSA	2003 Peds Dosage Handbook
Fraction of adult dose	25%	21%	25%	N/A
Total daily dose (mg)	225 - 450		391-783	680
Total daily dose (mg/kg)	13 - 26		23 - 46	40



体表面積や体重から
小児の投与量を換算できる魔法の
ような式はない

薬ごとによって、投与量の設定は
変わる可能性がある

臨床試験を行って、
投与量や有効性、安全性を
評価せねばならない

- テオフィリンの血中濃度を上げる因子
- CYP 1A2 (70%) > 2E1 > 3A4によって代謝
 - CYP 3A4を抑制する薬: 上げても2-3割
 - マクロライド系抗生物質
 - エリスロマイシン: 上げても2-3割
 - クラリスロマイシン: 上げても2割
 - azole系抗真菌剤、グレープフルーツジュース等
 - CYP 1A2を抑制する薬:
 - フルボキサミン、キノロン系抗生物質(例: シプロフロキサシン)
 - うっ血性心不全、肺炎、肝硬変など

- テオフィリンの血中濃度を下げる因子
- フェノバルビタール、フェニトインなど
 - CYP 3A4を誘導
 - さほど大きくは下げない
 - CYP1A2を誘導する薬のほうが影響は大きい
 - 喫煙者: 非喫煙者に比べると3-4割血中濃度は低い

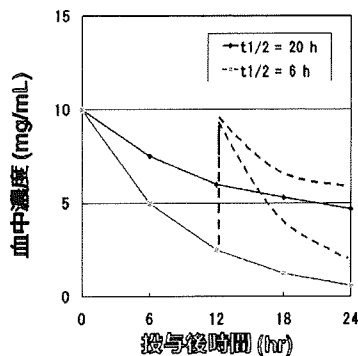
新生児におけるテオフィリン

- 成人、年長児とは適応も薬物動態も異なる
- 未熟児無呼吸発作の治療に用いる
- 50%以上未変化体のまま腎排泄
 - 成人では~90%肝代謝
- かなり、カフェインの血中濃度が上がる
 - カフェインの代謝が未発達
 - テオフィリン+カフェインの効果

新生児における アミノフィリン/テオフィリンの投与

- アミノフィリン
 - Loading dose: 4 - 6 mg/kg IV 30分かけて
 - Maintenance dose: 1.5 - 3.0 mg/kg/dose IV 8 - 12時間おき
- テオフィリン経口への変更は同量で
 - アミノフィリンは~78%テオフィリン
 - テオフィリンの生体内利用率は~80%
- 半減期は新生児では20-30時間

半減期と血中濃度推移

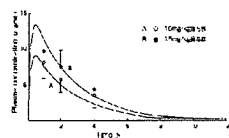


アセトアミノフェン

経口・鎮痛薬

- 消炎効果は極めて弱い
- 投与量は、
 - 体重1kgあたり1回10~15mg
 - 投与間隔は4~6時間以上
 - (成人量300-500mgを超えない)
 - 1日総量として600mg/kgを限度
- 症状を抑える効果しかない
- 熱が高くても元氣なら必要ない
- 15 mg/kgを4時間おきで2-3回投与しても構わない
- 寝ているのを起こしてまで投与する必要はない

カロナール2%シロップ投与後の血中濃度推移



ちなみに2007年9月までは...

- 「10-15mg/kg 原則として1日2回まで」、「小児での安全性は確立していない」など
- アンヒバ®坐薬の投与回数は、「通常1日1回投与」

1日2回や、1回5mg/kg等というunderdosingをしている医師が多い！！
適切な量を必要な時期にしっかりと投与することが大切！！

臨床薬理学とは

- 薬物治療におけるEvidence-based Medicineそのものである
 - 臨床薬理学を理解していない臨床医には薬物治療の個別化は難しい
- Evidenceを作るための学問である
 - 質の高い臨床試験を出来ない医療施設は一流と認めてもらえない

小児臨床薬理学

- 臨床薬理学にさらに発達の要素がからむ
- 小児の薬物治療の質の向上に必須
- クスリの臨床研究に必須
- 解明すべきことは多く残されている
 - 研究テーマの宝庫！

臨床研究論文の読み方

滋賀医科大学医学部附属病院治験
管理センター
中川雅生

考えてみよう

医師になってはじめて受け持つ疾患の患者さんの治療方針を決めないといけなくなりました。

さて、治療法の情報を得る方法は？

治療法の情報を得る方法

1. 先輩のアドバイス
2. ガイドラインや教科書
3. 学会での発表や討論
4. 研究論文

研究論文を読む目的

1. 診療に必要な医療情報収集
最新の情報
個別の臨床問題解決
2. EBM (Evidence-based Medicine)の
実践

EBM (Evidence-based Medicine) (Wikipedia)

エビデンスに基づく医療

医療行為における治療法の選択などにあたっては、理論や経験や権威者の判断ではなく、確固とした疫学的証拠に基づき、科学的に最良の判断をすべきであるという考え方。

中国語：循証医学、実証医学、証拠医学

エビデンス

- ・ 証拠、根拠の意味
- ・ ある治療法がある病気・怪我・症状に効果があることを示す証拠、検証結果。
- ・ 医療行為において、治療法を選択する際の根拠。
ある医学的事実に対する臨床的、学問的な証拠、裏付け

エビデンスレベルの分類

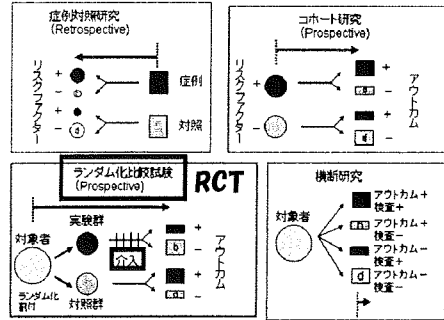
(診療ガイドライン作成におけるエビデンスレベルの分類)

- I. システマティックレビュー／メタアナリシス
- II. 1つ以上のランダム化比較試験による
- III. 非ランダム化比較試験による
- IV. 分析疫学的研究（コホート研究、症例対照研究）
- V. 記述研究（症例報告やケースシリーズ）による
- VI. 患者データに基づかない、専門委員会や専門家個人の意見

(その他、AHCRR (Agency for Health Care Policy and Research) 現在は AHRQ (Agency for Healthcare Research and Quality) の分類もある)

より客観性の高いもの！

主な研究デザインの分類



RCT: randomized controlled trial

CAST Study

心筋梗塞後の抗不整脈薬の使用の有効性評価

心筋梗塞は急性期が過ぎてから合併する不整脈が時として致命的となるため、抗不整脈薬が有効であるという理論、予測が従来からあり、抗不整脈薬が予防的に投与されていた。どのグループの薬剤がもっとも効果的かを調べるため、無作為二重盲検法による臨床実験が行われた。しかし中間報告で最も死亡率の低いのは薬剤非投与群だった。

EBM (Evidence-based Medicine)

大規模臨床試験において結果の良かったほうの選択肢を選ぶ (△)

+

エビデンスに対して批判的吟味を加え、目の前の患者にどう適応するか熟考する

||

EBM

EBMの効果

1. テータの提示
医療関係者 の責任として
国民 に説明する
2. 医療の標準化
3. 医療の透明化
医師のパターナリズムから脱却

研究（あるいは論文）のタイプ

- ・ RCT
- ・ コホート研究
- ・ ケースコントロール研究
- ・ 横断的研究
- ・ メタ・アナリシス
- ・ 総説
- ・ ケースシリーズ研究
- ・ 症例報告
- ・ ガイドラインや指針

批判的吟味をしてみよう

論文に書いてあることを

そのまま信用していいか？

そのまま診療に利用できるか？

批判的吟味

RCTに対するチェックポイント

1. 患者背景
2. 方法論 (RCT、フラインド、フローチャート)
3. 主要な結果
4. 層別解析
5. 副作用

患者背景

1. 患者背景と症例の類似性
2. 自分の患者にあてはめられるか・・・

方法論

方法論が明確か・・・

エビデンスレベル (客観性)
研究のタイプ (RCT、DBT)
フローチャートの適切性
CONSORT声明

(観察研究においてはSTROBE声明)

CONSORT声明

ランダム化並行群間比較試験報告の質向上のための
改訂版勧告 (JAMA2001;285:1987-91)

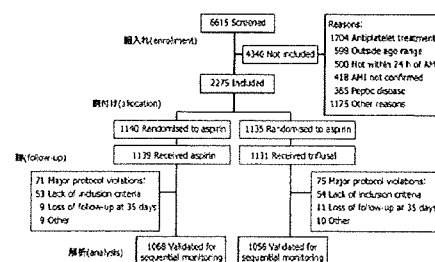
①チェックリスト

知見の信頼性や関連性の判断根拠確保
22項目 (タイトル、抄録、はじめに、方法、結果、考察) 読むときもこの視点で!

②フローチャート

臨床試験の4つの段階の流れを表示
組入れ (enrollment)
割付け (intervention allocation)
追跡 (follow-up)
解析 (analysis)

RCTのフローチャート例



脱落例や不備な点をチェック!

(折笠秀樹「EBM実践のための臨床試験論文の読み方」より引用、一部改変)

CONSORT声明

RCT報告を改善

企図の理由

実施方法

解析方法

臨床試験実施のプロセスの理解

結果の妥当性の評価

主要な結果

明らか〔有意〕な結果が得られているか・・・

治療効果は十分か（大きい）か・・・

対照群と比較してどうか・・・

相対的な治療効果

相対リスク減少率 (RRR: Relative Risk Reduction)

必要治療例数 (NNT: Number Needed to Treat)

層別解析

1. 自分の患者が適合する（属する）層の結果はどうか・・・

2. どういった背景の患者により有効か・・・

副作用

記載された（発現した）副作用はどんなものか・・・

目の前の患者に何を注意すべきでどんな検査をすべきか

偏り（バイアス）の評価

妥当な比較のためにRCTを行うことにしても偏りは生じるので配慮が必要

1. 選択バイアス (selection bias)

対象者を選択する際に生じる偏り

2. 割付バイアス (allocation bias)

割付の結果として評価項目に影響を与える重要な要因が偏る

3. 評価バイアス (assessment bias)

どちらに割付られたか知ることによって評価が偏る

4. 出版バイアス (publication bias)

良い結果のみが公表され、うまく行かない結果が公表されない

メタアナリシス

解析対象となった論文の信頼性は確保できているか・・・

解析対象とした論文収集は適格か・・・

QUOROM声明

Quality Of Reporting Of Meta-analysis

MOOSE チェックリスト

Meta-analysis of observational studies in epidemiology. A proposal reporting

QUOROM声明

(Moher D et al, The Lancet 354:1896-1900, 1999)

メタアナリシス結果の見方

関連文献のサーチ法の記述
選択された研究のデザイン要約の記載
各研究についての質の評価
特徴的困における結果の描き方
異質性の検定p値
出版バイアスの評価 (漏斗プロット)
効果サイズとその信頼区間
併合結果のp値はみない
層別解析結果の検討 (担当患者に近い点)
相対リスク以外にNNT (有効性の絶対数値) とNNTH
(危険性の絶対数値) の参照

MOOSEチェックリスト

(Stroup D et al, JAMA 283:2008-2012, 2000)

観察研究のメタ・アナリシスにおける著者、 編集者および査読者用のチェックリスト

背景の報告には以下の事項が含まれるべき
検索ストラテジーの報告
方法の報告
結果の報告
考察の報告
結論の報告

総説

エビデンスに基づいた記載になっているか・・・
筆者の学説を中心に記載が展開されていないか・・・
バランスよく文献引用されているか・・・

等々

症例報告

診療に利用できる内容か・・・
特異な症例ではないか・・・
科学的・倫理的な検討がされているか・・・
報告の内容は妥当か・・・

結論
観察期間





等々

おわりに・・・

研究論文を正しく読んで、診療に活用しましょう！

また、診療に活かせる臨床研究をし、論文を書いてください！

臨床研究と生物統計学・医療統計学

(財)パブリックヘルスリサーチセンターがん臨床研究支援事業(CSPOR)運営委員長
 NPO日本臨床研究支援ユニット理事長
 スタットコム(株)取締役会長
 NPO日本メディカルライタ協会理事長
 東京大学医学系研究科 公共健康医学専攻 生物統計学

大橋靖雄

医療のあゆみ

エビデンス(EBM) 疫学・統計

<p>1895 X線の発見 (レントゲン)</p> <p>1922 CTの発表 (ハウンスフィールド)</p> <p>1972 ハーセプチン発売 (ロシユ・ジェネンティク)</p>	<p>治療とくに薬物</p> <p>1897 アスピリン合成 (バイエル社ホフマン)</p> <p>1973 スタチンの発見 (三共 遠藤章)</p> <p>1991 EBM</p>	<p>1849 コレラの伝播様式(スノー)</p> <p>1858 ナイチンゲール覚書き</p> <p>1947 初めてのランダム化試験</p> <p>1960 フラミンガム研究 がん臨床試験のシステム</p> <p>1980 Milano研究開始</p> <p>1980-ARPA-NET</p> <p>1981 IBM/PC発表</p> <p>1996 PUBMED</p>
--	---	---

情報
コミュニケーション?

なぜ生物統計?

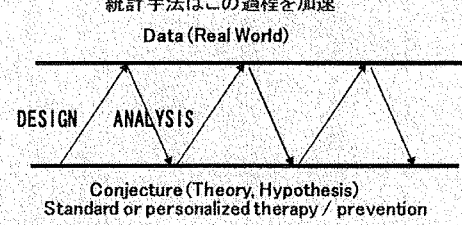
Design Design Design Design Design Design
 Design Design Design Design Design Design
 Design Design Design Design Design Design
 Design Design Analysis Design Design
 Design Design Design Design Design Design
 Design Design Design Design Design Design
 Design Design Design Design Design Design
 Design Design Design Design Design Design
 Design Design Design Design Design Design

+ Interpretation and Communication

Why Statistical Methods?

Iterative process of scientific investigation
 (Box and Tiao, 1973)

統計手法はこの過程を加速



Conjecture (Theory, Hypothesis)
 Standard or personalized therapy / prevention

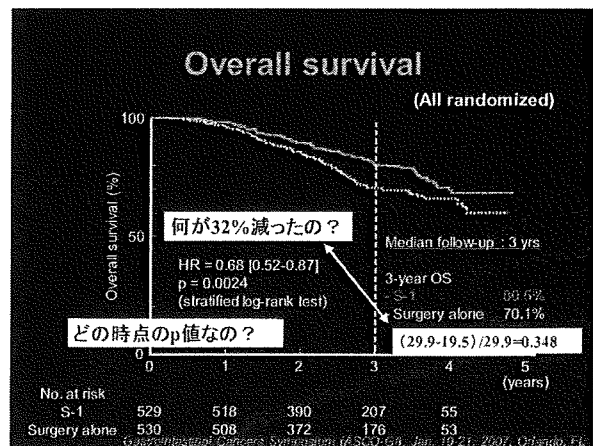
**Sakuramoto et al.,
 Adjuvant chemotherapy for
 Gastric cancer with S-1,
 an oral fluoropyrimidine,
 NEJM 2007; 357:
 1810-20.**

Abstract
 Adjuvant chemotherapy improved survival in patients with curable gastric cancer.

Background
 Patients who undergo resection of gastric cancer with curative intent are at high risk of relapse or death. We conducted a randomized, controlled trial to compare the efficacy of adjuvant chemotherapy with S-1 or no adjuvant therapy in the adjuvant setting.

Methods
 We randomly assigned 527 patients to the S-1 group and 526 patients to the surgery-only group between October 2002 and December 2004. The trial was stopped on the recommendation of the independent data and safety monitoring committee because the benefit was statistically significant (P < 0.001) in favor of the S-1 group. The median time to relapse or death in the S-1 group was 29.9 months, compared with 29.9 months in the surgery-only group. The hazard ratio for death in the S-1 group, as compared with the surgery-only group, was 0.68 (95% confidence interval, 0.52 to 0.87) (stratified log-rank test). The overall survival rate at 3 years was 59.5% in the S-1 group and 53.1% in the surgery-only group. The median overall survival time was 29.9 months in the S-1 group and 29.9 months in the surgery-only group. The primary end point was overall survival.

Conclusions
 S-1 as adjuvant chemotherapy improved survival in patients who had undergone a R0 resection for localized gastric cancer. (ClinicalTrials.gov number, NCT00152227.)



Q:統計手法とくに生存時間(time-to-event)解析に関する質問

- ◆ 図表などでハザード比とp値の両方が記載されていることが多いが、このときのp値はハザード比とは全く別に算出されたものなのか？
- ◆ PFS、DFS、TTP、OSなどの図で、特に断り書きの無いp値はどの時点でのp値なのか？

内容

- ◆ なぜ統計学が必要か？ バラツキの存在
- ◆ 確率を用いた判断とは？ P値と検定(test)
- ◆ 「良い」推測とは？ 正確性と精密性
- ◆ データの適切な解釈とは？

なぜ生物統計？

推論・判断の際にバラツキが無視できないから

バラツキの存在下で正しく推論・判断するためにデータをどうとるか？
バラツキのもとでの客観的な推論・判断をどう行うか？

バラツキを如何に小さくするか、如何に効率的にデータを得るか？
(品質管理、データマネージメントの範疇でもある。)

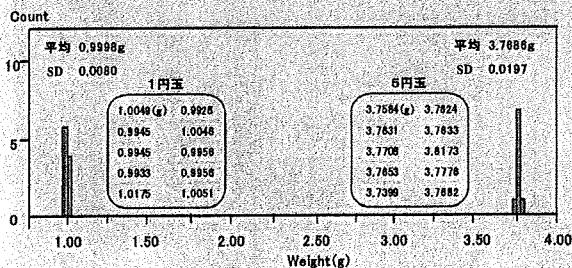
なぜ統計学が必要か？

バラツキの存在下での判断

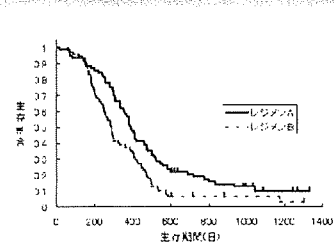
例題：5円玉と1円玉はどちらが重いか？
SCLCで(CPT11+CDDP)と(VP-16+CDDP)の
どちらの予後が良いか？

SCLC: 小細胞肺癌
CPT11: 塩酸イリノテカン
CDDP: シスプラチン、VP16: エトポシド

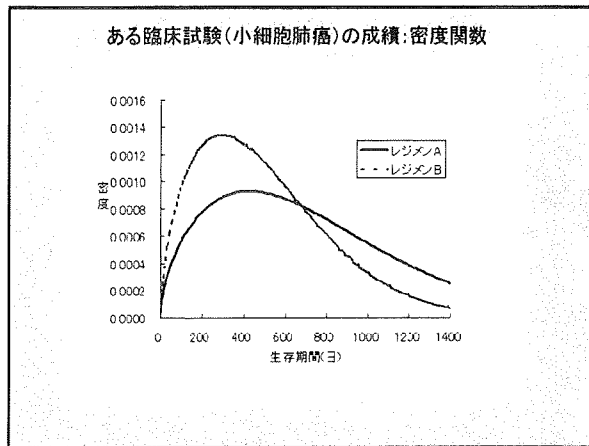
5円玉と1円玉:ヒストグラム(密度関数)



JCOG9511(小細胞肺癌に対する臨床試験)の成績:生存曲線



グラフの見方:
(念のため)グラフが上にある方が生存成績が良い
端の方は信頼性が低い
MST Median-Survival-Timeは50%の線と交わる時点



バラツキの存在下での判断

なぜ「5円玉の方が1円玉より重い」と判断するか？

- ◆ ともにバラツキている: 「分布」の概念の導入
バラツキを客観的・定量的に扱うために「確率論」を応用
- ◆ 実際の測定値の分布の比較 internalな証拠
比較してよいか、不公平な比較ではないか

5円玉	SCLC
「日本中5円玉、1円玉はどれも同じ」 無作為(ランダム)抽出したとみなして良い 同じ測定方法	選択に偏り?しかし ランダム化 同じ測定・追跡方法

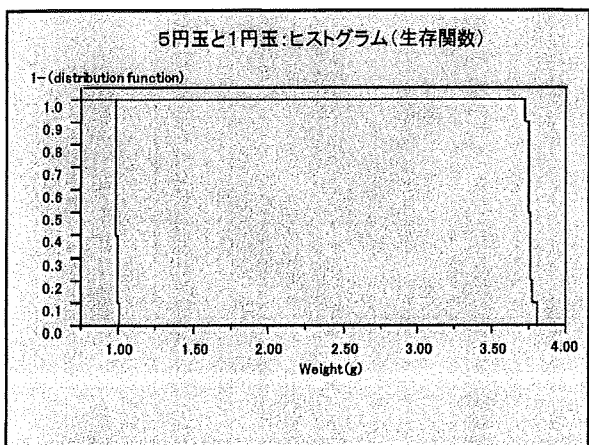
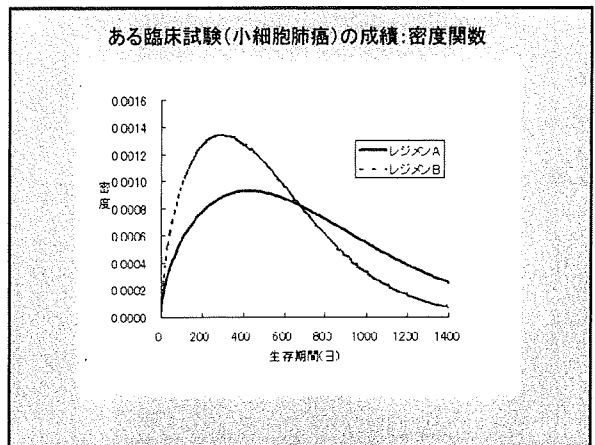
群内のバラツキと群間の違いを比較: シグナルとノイズの比較

バラツキの存在下での判断(続き)

なぜ「5円玉の方が1円玉より重い」と判断するか？

- ◆ 先験的な知識・常識(externalな証拠)の活用と一般化

	5円玉	SCLC
先験情報	これまでの「持った感じ」	第II相の結果
対象選択	「日本中5円玉、1円玉はどれも同じ」 無作為(ランダム)抽出したとみなして良い	選択に偏り? ×
一般化	確率的な不確実さがあるもの可能	?



Random sampling vs randomization

ランダム化(無作為化) randomization 内的妥当性の確保
無作為抽出 random sampling 外的妥当性

無作為抽出: 多くの統計理論において理論上重要な概念であるが、臨床試験においてはほとんど現実味をもたない。私見であるが、不用で誤解を招く概念
Senn, S(1997), *Statistical Issues in Drug Development*, Wiley: glossary

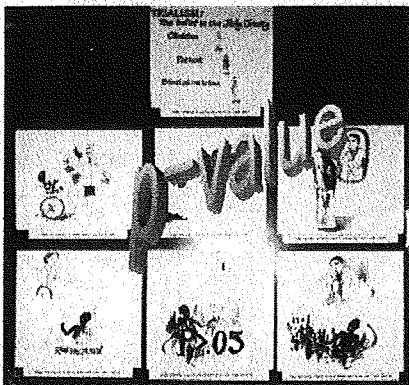
一般化可能性のために

- ◆ 患者背景の解析
結果の差異の説明
- ◆ 部分集団の解析
部分集団における効果の差、交互作用の検討
- ◆ 再試験
FDAの方針
- ◆ メタアナリシス
複数の独立な研究結果の統計的併合
- ◆ 緩い選択条件
large-scale-randomized-evidence
- ◆ ITT (Intention-to-treat)解析

内容

- ◆ なぜ統計学が必要か？ バラツキの存在
- ◆ 確率を用いた判断とは？ P値と検定(test)
- ◆ 「良い」推測とは？ 正確性と精密性
- ◆ データの適切な解釈とは？

医師・患者さん・統計家の天国と地獄(EORTCの漫画):



ある臨床試験の結果: 薬は効いたと考えてよいか

リユーマチに対する多施設共同プラセボ対照2重盲検無作為化比較試験

エンドポイント(反応):改善あり、無し response: yes, none

	yes	none	total
test	n ₁₁ =28	n ₁₂ =13	n _{1.} =41
placebo	n ₂₁ =14	n ₂₂ =29	n _{2.} =43
	n _{.1} =42	n _{.2} =42	n=84

割合の差 68.3-32.6=35.7% p値 0.00108!
Koch et al. :From Pease ed. (1988),
Biopharmaceutical statistics for drug development, Dekker, 翻訳有り(サイエンティスト社)

どちらのレストランをデートに選びますか？

おいしいレストランにいきたい。A、BのうちAはかなり高い。
まだ行ったことが無いので、人の評判をあてにすることにした。
Aが良いというはっきりした証拠があればA、無ければBにしたい。

さて評判データは？

A	B	A	B
10	0	100	0
9	1	90	10
8	2	80	20
7	3	70	30
6	4	60	40

どう考えるか？ 背理法と検定の考え方

背理法 (2重否定の論理)
証明したい命題を否定する

このもとで論理を進めると矛盾が起きる → 命題を認める

例:21人の招待者がいる。
4つのテーブルに席を作るとすると1テーブル何人の席まで用意しなければならぬか？
命題「6人以上のテーブルが最低1つは必要になる。」

すべてのテーブル5人以下で足りる、と仮定すると・・・
招待者の総数は20人以下となる。
矛盾！

どう考えるか？ 背理法と検定の考え方

(統計的)検定 (2重否定の論理)
 差はないと仮定する(帰無仮説) レストランに差はない

このもとで観察されたデータが発生しやすいか確率を計算する。稀ならば → 仮説を棄却するすなわち差があると考えあるいは他の説明を求める

確率を計算する

差無しのもとではレストランを選ぶ確率はともに0.5

10人全部Aを選ぶ確率は？
 1人目がA $p_1=0.5$ 2人目がA $p_2=0.5$...
 それぞれの判断が独立なら「AND」の確率は 掛け算
 答えは $0.5^{10}=0.00098$

9人がAを選ぶ確率は？
 1人目がB 残りがA
 2人目がB 残りがA
 ... 全部で10パターン ${}_{10}C_1$
 それぞれの確率は $0.5^{10}=0.00098$ 、「OR」の確率は 足し算
 答えは $10 \times 0.5^{10}=0.0098$

p値とは

帰無仮説のもとで、実際の観察結果以上に仮説からずれたデータが得られる確率

A	B	p-value
10	0	0.00098
9	1	$0.00098 + 10 \times 0.00098 = 0.0107$
8	2	0.0547
7	3	0.1719
6	4	0.3769

上記は片側p値、両側p値は2倍して得られる

情報量(症例数)とp値

A	B	p-value
100	0	0.00000
90	10	0.00000
80	20	0.00000
70	30	0.00004
60	40	0.0284
59	41	0.0443
58	42	0.0666
57	43	0.0967

症例数が増えれば、(正確なデータである限りは)それだけ情報量が増え、正しい判断を行える可能性が増える

データ = 真実 + バイアス + 確率的ゆらぎ(誤差)

どの程度の確率なら無視できるのか？ 100万人に1人が死ぬリスクの比較

Pochin EE: The acceptance of risk. *Brit. Med. Bull.* 1975;31: 184-90

- 650kmの航空機飛行
- 100kmの自動車旅行
- 紙巻タバコ3/4本の喫煙
- 岩登りを1.5分間続ける
- 60歳の人間が20分過す
- 経口避妊薬を2.5週間飲む
- ワインを半分飲む
- 10ミリレムの放射線被曝

p値の目安は？ 他の説明とは？

5%を下回れば 統計的に有意 statistically significant
 5%を用いるのは伝統・習慣 有意水準5%
 ICHG9 統計的原則では 両側5%に統一 (研究では片側5%も)

5% 阪神タイガースが優勝するくらい稀

他の説明とは？
 レストランの外形にだまされているのでは？ 情報バイアス
 回答者に偏りあるのでは？ 選択バイアス
 私にはあてはまらないのでは？ 治療の個別化
 賄賂をもらっているのでは？ 品質保証

ある臨床試験の結果

リウマチに対する多施設共同プラセボ対照2重盲検無作為化比較試験

エンドポイント(反応):改善あり、無し response: yes, none

test	yes	none	total
placebo	$n_{11}=28$	$n_{12}=13$	$n_{1\cdot}=41$
	$n_{21}=14$	$n_{22}=29$	$n_{2\cdot}=43$
	$n_{\cdot 1}=42$	$n_{\cdot 2}=42$	$n=84$

割合の差 35.7% p値 0.00108!

Koch et al. : From Peace ed. (1988),
Biopharmaceutical statistics for drug development, Dekker, 翻訳有り(サイエンティスト社)

ランダム化に基づく検定

前提:ランダム化のみ

反応に個人差を考慮してもよい

仮説 H_0 : 介入(薬剤)による反応に差はない

i番目の患者の反応 test= X_{iT} placebo X_{iP}

$$X_{iT} = X_{iP}$$

仮説の下で、ランダム化は n_{11} 個の黒玉と n_{12} 個の白玉をランダムに n_{11} と n_{12} に分ける作業に対応

ランダム化に基づく正確な検定

得られたパターンの出現確率

$n_{11} C_{n_{11}}^{n_{11}} \cdot n_{12} C_{n_{12}}^{n_{12}} / n C_{n_{11}}^{n_{11}} \cdot n C_{n_{12}}^{n_{12}} = \frac{1}{2^{n_{11}+n_{12}}} = \frac{1}{2^{41}}$

より極端なパターンの出現確率を加える: 片側 2群に分ける

$n_{11}=28$ 確率 0.00082258
 29,30,...,41 合計 0.001

確率が同じかそれ以下のパターンを加える(13以下か28以上): 両側 0.002

Fisherの直接(正確)確率検定 Fisher's Exact test

この場合の確率とは? 仮想的な実験の繰り返しの下で相対頻度

内容

- ◆なぜ統計学が必要か? バラツキの存在
- ◆確率を用いた判断とは? P値と検定(test)
- ◆「良い」推測とは? 正確性と精密性
- ◆データの適切な解釈とは?

認知症予防に「運動・栄養・昼寝」... 厚生省研究班データ

読売新聞朝刊2006年5月27日

よく運動し、栄養に気をつけて、昼寝した方が認知症の発症率が下がることが、厚生労働省の研究班(主任研究者=朝田隆・筑波大教授)の研究でわかった。生活習慣の改善による認知症予防の成果が確認されたのは初めてで、注目される。

研究は、表向き利根町の65歳以上を対象に2001年から2005年にかけて行われた。希望者約400人に運動や栄養、昼寝の改善を指導し、指導しなかった1500人と比較した。

具体的には、週3~5回、1回20~60分、音楽に合わせてステップを踏む簡単な有酸素運動を行った。また魚の脂質に含まれるDHA(ドコサヘキサエン酸)やEPA(エイコサペンタエン酸)などを含む栄養補助剤を毎日取るとともに、30分以内の昼寝をした。

その結果、生活習慣を指導したグループでは認知症の発症率が3.1%だったのに対し、しなかったグループは4.3%にのぼった。また、記憶能力のテストでも、指導したグループの成績が約16%向上した。今後さらに統計的分析を進める。

認知症予防については、海外でさまざまな研究がなされており、魚を食べたり運動をしたりすることなどが望ましいとされてきた。しかし、生活習慣改善を行う「介入研究」ではなく、生活習慣を観察し、数年にわたって認知症の発症率などを見る「観察研究」が主だった。

認知症予防に「運動・栄養・昼寝」... 厚生省研究班データ

読売新聞朝刊2006年5月27日

よく運動し、栄養に気をつけて、昼寝した方が認知症の発症率が下がることが、厚生労働省の研究班(主任研究者=朝田隆・筑波大教授)の研究でわかった。生活習慣の改善による認知症予防の成果が確認されたのは初めてで、注目される。

研究は、表向き利根町の65歳以上を対象に2001年から2005年にかけて行われた。希望者約400人に運動や栄養、昼寝の改善を指導し、指導しなかった1500人と比較した。

割り付けは? 比較可能性は?

具体的には、週3~5回、1回20~60分、音楽に合わせてステップを踏む簡単な有酸素運動を行った。また魚の脂質に含まれるDHA(ドコサヘキサエン酸)やEPA(エイコサペンタエン酸)などを含む栄養補助剤を毎日取るとともに、30分以内の昼寝をした。

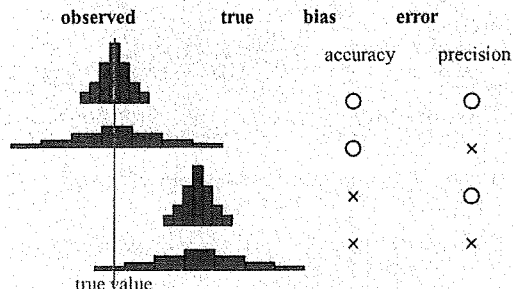
その結果、生活習慣を指導したグループでは認知症の発症率が3.1%だったのに対し、しなかったグループは4.3%にのぼった。また、記憶能力のテストでも、指導したグループの成績が約16%向上した。今後さらに統計的分析を進める。

統計的特徴はない? 臨床的にも意味あるか?

認知症予防については、海外でさまざまな研究がなされており、魚を食べたり運動をしたりすることなどが望ましいとされてきた。しかし、生活習慣改善を行う「介入研究」ではなく、生活習慣を観察し、数年にわたって認知症の発症率などを見る「観察研究」が主だった。

精密度と正確度 Precision and accuracy

観測値・研究成果 = 真の値 + バイアス + 誤差的バラツキ



(臨床・疫学) 研究の目標

- ◆ 誤差的バラツキを小さくすること (精度を高くすること)
Clarity
- ◆ 偏り(バイアス)を小さくすること
Comparability
- ◆ 広い対象に適用できる結論を得ること
Generalizability

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 (%)
Shapiro and Heston	13	45
Hall and Good	19	63
Rocklin et al.	47	55
Allahr et al.	17	47
Cerna et al.	13	44
Evry	12	41
Field	37	41
Deh	12	36
Wells and Jackson	37	35
Frigolin and Humphrey	12	33
Hurley	150	31
ECOG	48	26
Tabby	271	21
Hymas et al.	20	20
Moore et al.	20	19
Amfeld	141	17
Hayo	358	17
Eliass	27	12
Kneazdy	21	9
Kneazdy et al.	11	9
Olsen 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ラツキ
たと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
- ◆ 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
- ◆ 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
確率変数としてのモデル化: 確率論の応用が可能となる
さらに確率化により「誤差化」する (無作為抽出、ランダム化)

Random sampling vs randomization

ランダム化(無作為化) randomization 内的妥当性の確保
無作為抽出 random sampling 外的妥当性



無作為抽出: 多くの統計理論において理論上重要な概念であるが、臨床試験においてほとんど現実味をもたない。私見であるが、不用で誤解を招く概念
Senn, S(1997), *Statistical Issues in Drug Development*, Wiley: glossary

ランダム化によって、
実際は観測されない二つの介入結果の
差(の平均)を、正確に推定することができる! ?

ランダム化の意義

例：仮想的な10人の患者
治療に対する潜在的反応

A	B	人数	推定したい奏効率の差
+	+	2人	(5-3)/10 = 20%
+	-	3人	
-	+	1人	
-	-	4人	

ランダム割付け 10Cs = 252パターン 最良は100-20=80%、最悪は0-40=-40%
この分布は？

真の値20%からのバイアスは0、誤差的バラツキは大
誤差的バラツキを小さくするには？ サンプルサイズ大きく

並べ替え分布

ランダム化の原理

$X_{1A}, X_{2A}, \dots, X_{1B}, X_{2B}, \dots, X_{nB}$ をA群(n)とB群(n)にランダム化
 X_{iA} のAとBの下でのアウトカムは、 X_{iA} と X_{iB}
これらは通常は同時には観察されない(counter-factualな想定)

平均値の差は？
ランダム化によって AかBかが各対象に割り振られる
この割り付けの確率変数を δ_i とする(Aなら1、Bなら0)とすると

実験の結果の各群の平均値の差は $D = \sum 1/n (X_{iA} \delta_i - X_{iB} (1 - \delta_i))$
この、全てのランダム化パターンの下での期待値は $D = \sum 1/2n (X_{iA} - X_{iB})$ 真値

介入効果がない、すなわちA群とB群の差がなければ、期待値は0
分散は $V = \frac{\sum (X_{iA} - X_{iB})^2}{n(2n-1)}$

この分散を用いて検定してもよいし正確な分布を用いて検定してもよい。割り付け方法を変えれば、この分散(分布)も変化

標準正規分布(平均0、分散1)の密度関数

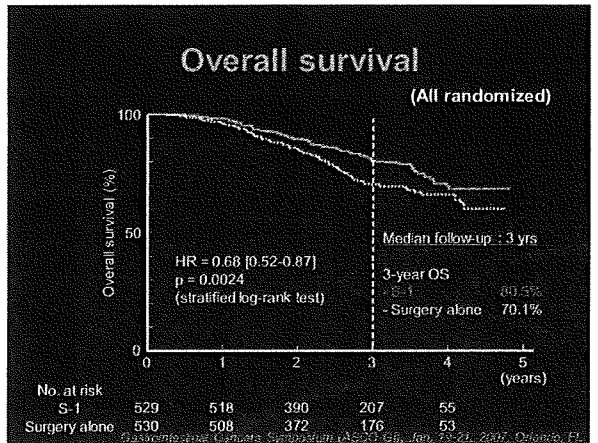
1.96:これを超えれば5%有意

真値に収束する 大数の法則

分布の形が正規分布に収束する 中心極限定理

内容

- ◆ なぜ統計学が必要か？ バラツキの存在
- ◆ 確率を用いた判断とは？ P値と検定(test)
- ◆ 「良い」推測とは？ 正確性と精密性
- ◆ データの適切な解釈とは？



生存時間解析の3種の神器

「Kaplan-Meier法で生存曲線を推定し、Log-rank検定で群間比較を行い、Cox回帰(多変量解析)による調整解析も行い結果の安定性を確認した。」

Kaplan-Meier法
Log-rank検定
Cox回帰

ランダム化試験なら基本的解析はログランク検定 しかし、単なる有意性有無だけでなく、治療効果を推定したい(割付に用いた)予後因子の予後予測性は？治療効果予後因子の予測性は(治療との交互作用)？多数の因子の影響を同時に考慮できないか？

ハザード

- ◆ 発症率 罹患率 morbidity, incidence rate
- ◆ 死亡率 mortality, death rate

- ◆ ハザード hazard
ともこれまで起きていない事象(イベント)が新たに起きるスピードとして
が用いられる

ハザードとその推定

- ◆ 発症率、死亡率 ハザード hazard
[time]⁻¹
単位時間あたりの新たなイベントの発生
- ◆ ハザードの直接推定(保険統計法) actuarial method
国の人口動態統計など



- ◆ 臨床では、サンプルサイズ不足のため(累積)生存率・発症率推定
通常Kaplan-Meier法(これを1から引いた関数が累積発症率)

オッズ、ハザード、リスク

リスク(累積発症率)

$$p(t) = 1 - \exp\left(-\int_{u=0}^t \lambda(u) du\right)$$

λが小さく一定なら $\approx \lambda t$

オッズ $\frac{p}{1-p}$ p小さければ $\approx p$

慢性疾患では(イベント発生少ないが再発でも)
オッズ比~ハザード比~リスク比

ハザード(比)と累積死亡率: 数値例

	1-year	2-year	3-year	累積死亡率	まとめたハザード
Group1	2/1000	2/998	2/996	0.006	6/2991=0.002006
Group2	4/1000	4/996	4/992	0.012	12/2982=0.004024
ハザード比	2.002	2.006	2.010		2.0061
				リスク比 2.000	

ハザード小さく一定 ⇒ ハザード比 ≒ リスク比

ハザード(比)と累積死亡率: 数値例

	1-year	2-year	3-year	累積死亡率	まとめたハザード
Group1	2/1000	3/998	4/995	0.009	9/2988.5=0.003012
Group2	4/1000	6/996	8/990	0.018	18/2977=0.006046
ハザード比	2.002	2.007	2.014		2.0074
				リスク比 2.000	

ハザード小さく一定 ⇒ ハザード比 ≒ リスク比
比例ハザード性

Q:統計手法とくに生存時間(time-to-event)解析に関する質問

- ◆ 図表などでハザード比とp値の両方が記載されていることが多いが、このときのp値はハザード比とは全く別に算出されたものなのか？
- ◆ PFS、DFS、TTP、OSなどの図で、特に断り書きの無いp値はどの時点でのp値なのか？

p値は通常はログランク検定による

イベント発生時点毎に群間のハザード比を求めそれを総合よって特定の時点というより、すべてのイベント時点での平均ハザード比はログランク検定から間接的に求めるかCox回帰で推定を行っている。両者はほぼ等しい

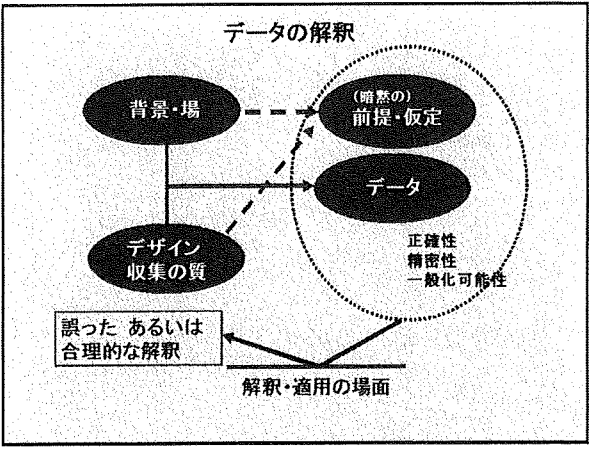
さらに、データの解釈と意思決定

- ◆ 胃痛の術後に薬剤S1を投与したところ、死亡率が30%減少するという大規模臨床試験の結果が発表された。福島大学の木村先生は、昨日手術した田中さんにS1を投与すべきであろうか。

EBM

臨床試験結果の信憑性(正確性と精密性)は？
 しっかりしたデザインと質・量か？
 一般化可能性は？ さらに目の前の患者に適用できるか？
 対象患者は福島大学に来る患者と同質？
 論文で行なわれている手術と同様の手術が可能？
 田中さんに特殊な事情はないか？

田中さんの希望は？ 説明の上の同意
Informed consent



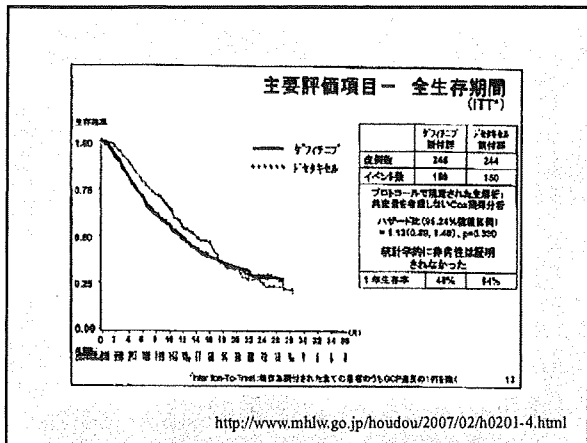
適性使用の問題
 短期・大量に処方
 抗癌剤認可の問題
 第III相試験はネガティブ
 医療制度(がん治療)の問題
 専門家以外の処方

PMSの問題の複雑さ

イレッサ(ゲフィチニブ)臨床試験とその背景

- ◆ 2002年7月5日、優れた第II相試験結果を下に世界に先駆けてわが国で承認
- ◆ 発売当時から副作用である間質性肺炎については添付文書で注意。しかし...「夢の抗癌剤」としての過剰な期待から非専門家による処方・治療も含め一気に使用が広がり、結果として間質性肺炎による死亡例が続発
- ◆ 厚生労働省は13例の死亡例報告を受け、10月15日にアストラゼネカ社に緊急安全性情報発出を指示。12月25日には第1回のゲフィチニブ安全性問題検討
- ◆ 2003年5月5日には米FDAがゲフィチニブを認可したものの、2005年1月4日、全体での延命効果は存在しないというプラセボ対照のISEL試験結果(AACR2005)を受け、アストラゼネカ社は欧州医薬品審査庁への申請を取り下げ
- ◆ しかし、ISEL試験では東洋人に対しては延命効果があることが示唆され、他の試験においても女性・腺癌・非喫煙という特定の部分集団に奏効するという示唆
- ◆ この示唆はゲフィチニブがEGFRレセプター突然変異例に奏効するという独立な研究結果(Lynch et al. NEJM 2004;350:2129-39, Paerzi et al. Science 2004; 304:1497-500)と符合(突然変異例が女性・腺癌・非喫煙に多かった)
- ◆ ドセタキセルを対照とした市販後試験結果の概要は、2007年2月1日に開催された厚生労働省の安全対策調査会で検討 (<http://www.mhwh.go.jp/houdou/2007/02/h0201-4.html>)

イレッサ公式サイト
http://www.iressa.jp/ekdgaku/eki_index.asp

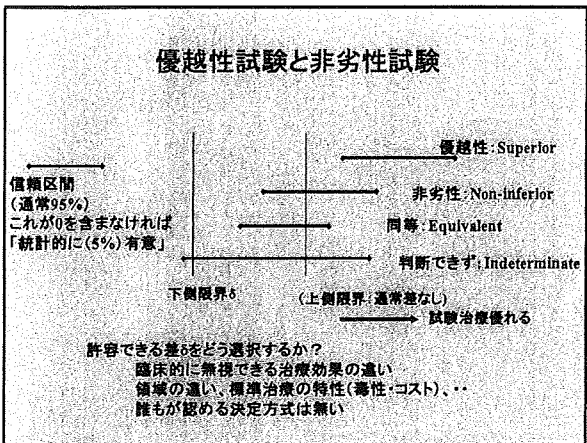


イレブテック公式サイト: 結果の概要

本試験の主要目的は、試験実施計画書で規定した統計学的な判定基準 (Cox 回帰分析を行い、ハザード比の信頼区間の上限が 1.25 以下であること) に従い、全生存期間におけるイレブテックのドセタキセルに対する非劣性を証明することです。

この規定された解析を行った結果、主要目的であるイレブテックのドセタキセルに対する非劣性の証明には至りませんでした (ハザード比 1.12、95.24% 信頼区間 0.89-1.40、 $p=0.330$)。

なお、この解析は、全生存期間におけるドセタキセル群に対するイレブテック群の効果が時間の経過にかかわらず一定であることを前提としていますが、第三者統計専門家が行った追加の解析では、当該効果が時間依存的に変化するという結果がでており、その前提が成り立っているとは言い難い結果でした。

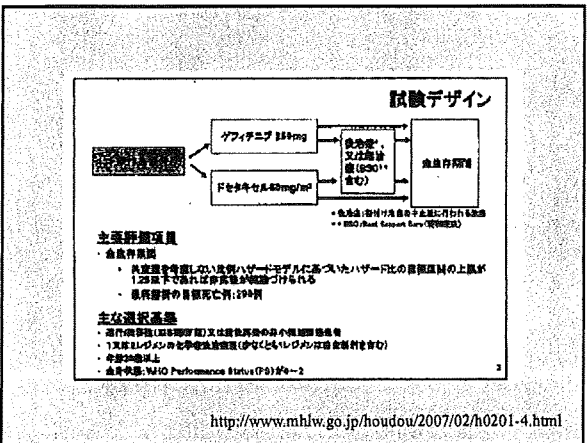


結果概要

1又は2レジメンの化学療法治療歴を有する、進行/転移性 (ⅢB期/Ⅳ期)又は術後再発の非小細胞肺癌患者を対象にゲフィチニブとドセタキセルの生存期間を比較する多施設共同非盲検無作為化並行群間比較第Ⅲ相臨床試験

アストラゼネカ株式会社
平成19年2月1日
平成18年度第2回薬事・食品衛生審議会
医薬品等安全対策部会安全対策調査会

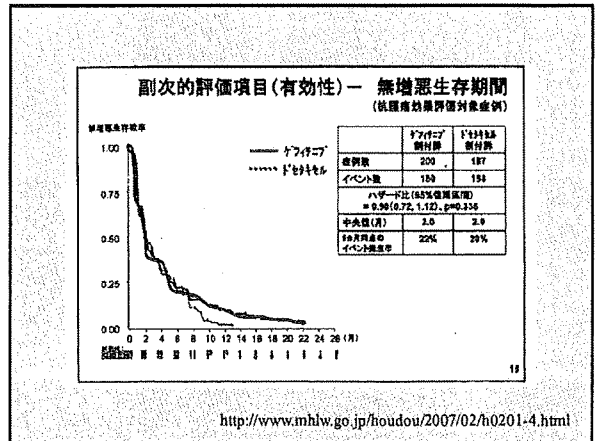
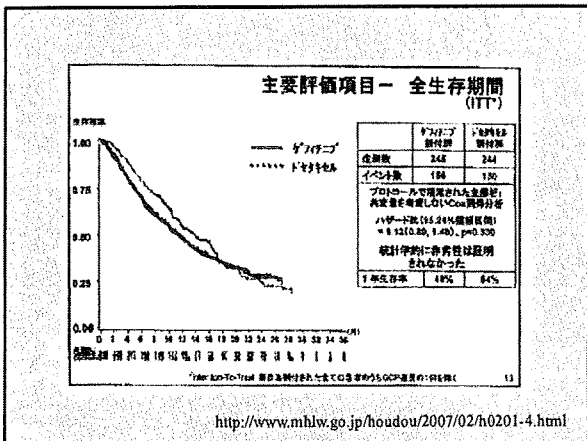
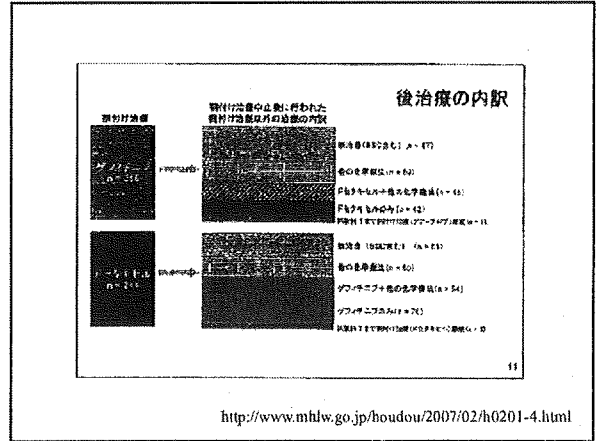
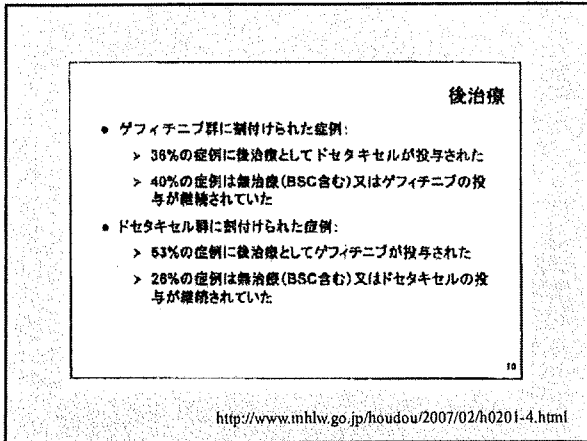
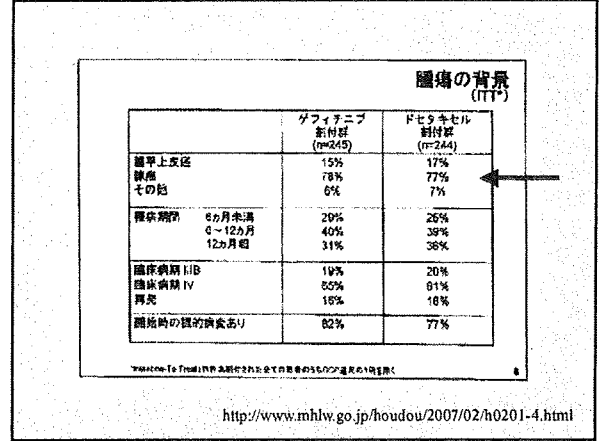
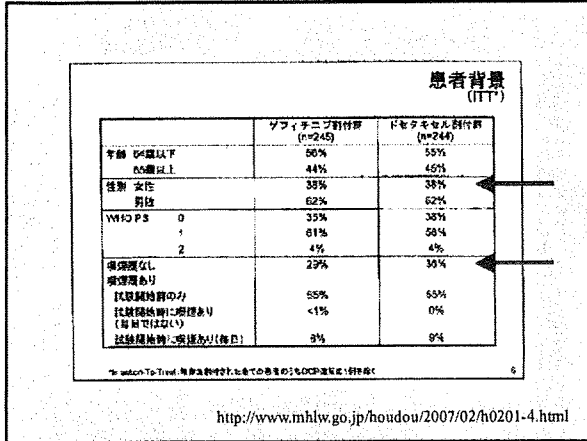
<http://www.mhlw.go.jp/houdou/2007/02/h0201-4.html>

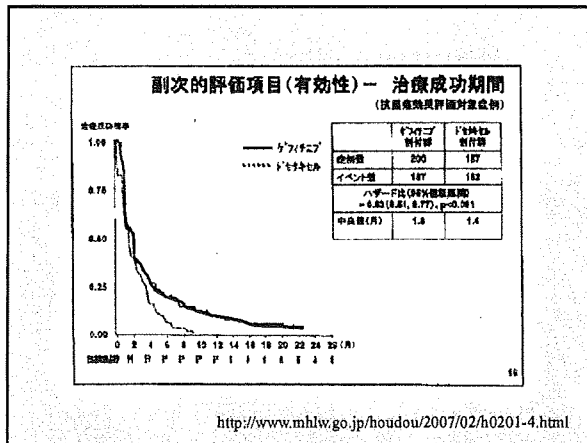


臨床試験の実施状況

- 実施施設: 全国計50施設
- 無作為割付症例数: 490例
- 登録期間: 2003年9月 ~ 2006年1月
- 全生存期間のデータカットオフ日: 2006年10月31日
- データカットオフ日までの死亡: 306例 (全症例の63%)
- 追跡期間 (割付日 ~ 死亡又は最終生存確認日) の中央値: 21ヵ月

<http://www.mhlw.go.jp/houdou/2007/02/h0201-4.html>



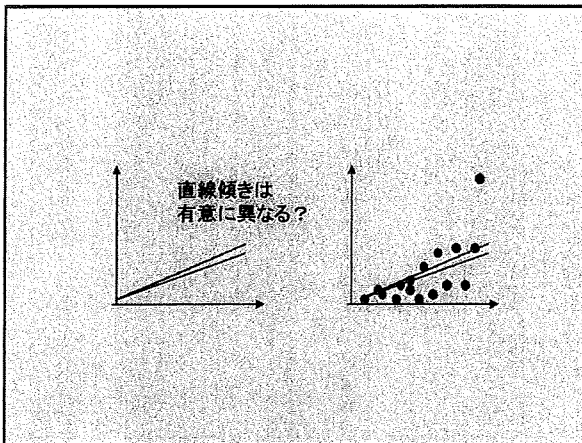
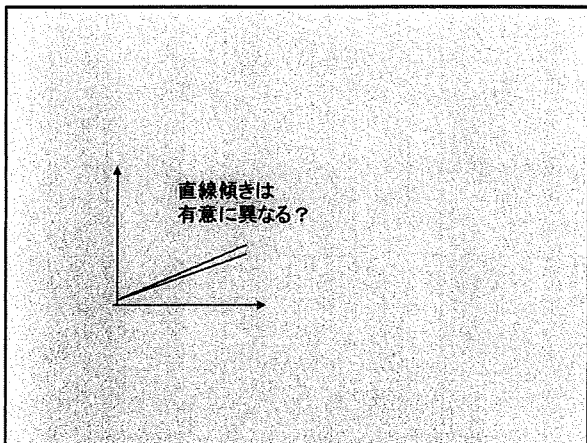


非劣性が検証できなかったとは

1. 劣っていた
2. 判断できるだけの情報量が存在しない
3. 非劣性を判断するために、2群の違いをハザード比に要約した。それが適切ではなかった!

比例ハザード性が成立しなかった!

なぜ?



なぜイレッサ臨床試験で比例ハザード性が成立しなかったのか?

- ◆ 再発では比較的比例ハザード性がよく成り立つが...
- ◆ イレッサが奏効する部分集団の存在
ハザード比の交差
- ◆ 進行癌の生存がエンドポイントであった
そもそも比例ハザード的でないのかも
さらに後治療が比例ハザード性をくずした
- ◆ さらに
後治療の選択に群間差(ドセタキセル群で早く移行の傾向)
登録された患者は必ずしもイレッサに有利ではなかった
- ◆ 振り返ってみればデザインに問題、しかし...
当時、NSCLCの非劣性試験は生存ハザードがエンドポイント
後治療が生存に寄与することは殆ど予想されていなかった
予測因子に関する情報は試験途中で判明

