

がんの臨床研究論文を読む のに必要な統計学

山本 精一郎

国立がんセンターがん対策情報センター

今回の対象

- medical oncologistががんの臨床研究に関する論文を読むのに必要な知識、主に統計学的側面について知ってもらう
 - 自ら臨床研究を行う研究者向けの内容ではない
- 臨床分野のみとする
 - 時間の都合上、診断評価の論文は対象としない
 - がんの疫学（がんの原因の探求）の論文は対象としない
 - 基礎研究の論文も対象とはしない
- 論文を読むには統計だけでは不十分なので研究デザインについても合わせて説明
- 本講義の内容は臨床腫瘍学会教育セッションで話している内容です。

典型的ながん臨床論文の種類

■ 臨床試験

- ☐ Phase1

- ☐ Phase2

- ☐ phase3

■ 観察研究

- ☐ レトロで治療成績を評価

- ☐ 予後因子解析

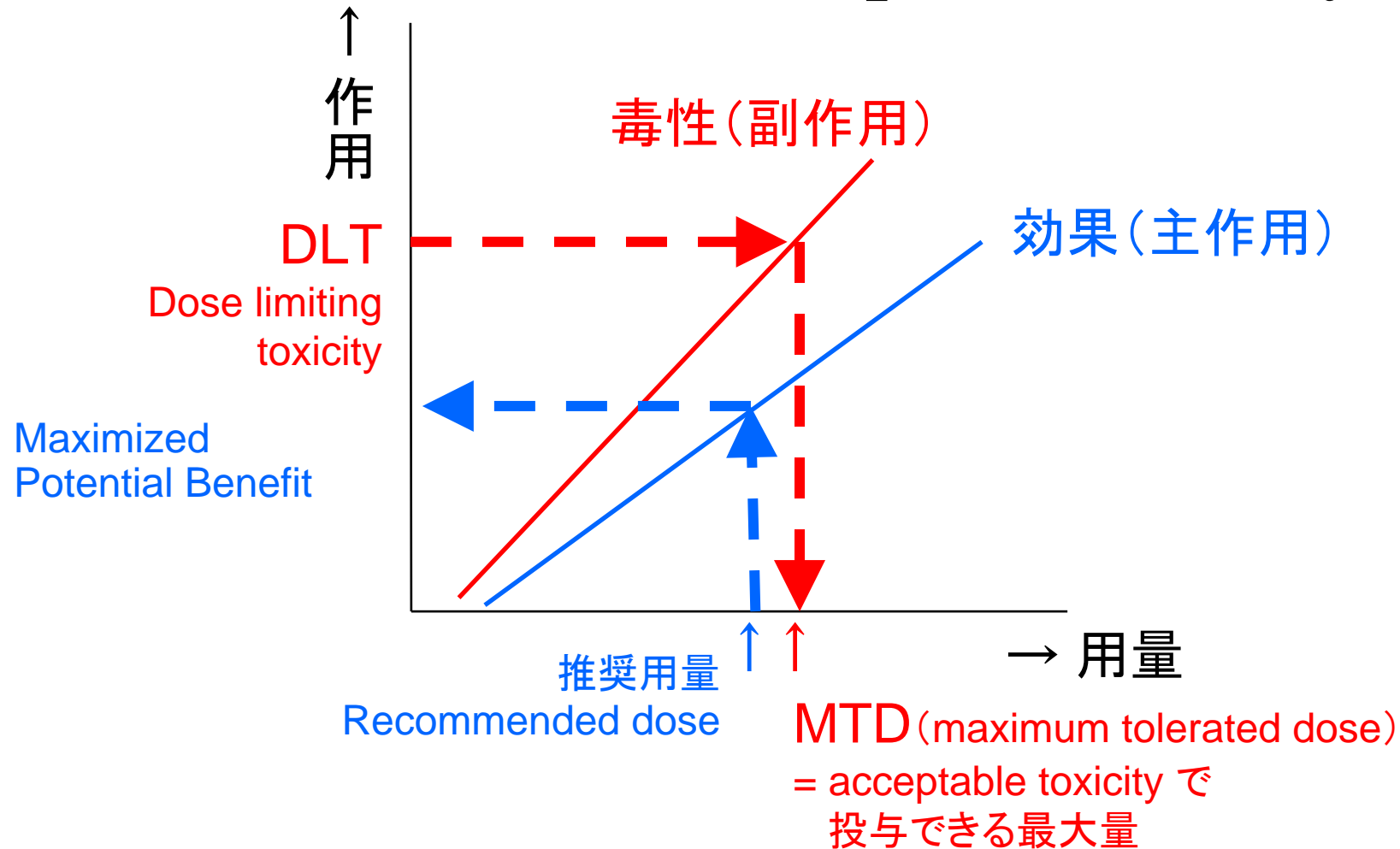
上記の内容を例題論文を挙げながら説明
(例題は私が実際携わった研究から選択)
用語はなるべく英語のままで



phase1臨床試験論文を読む のに必要な統計

細胞障害性抗がん剤p1デザインのrationale

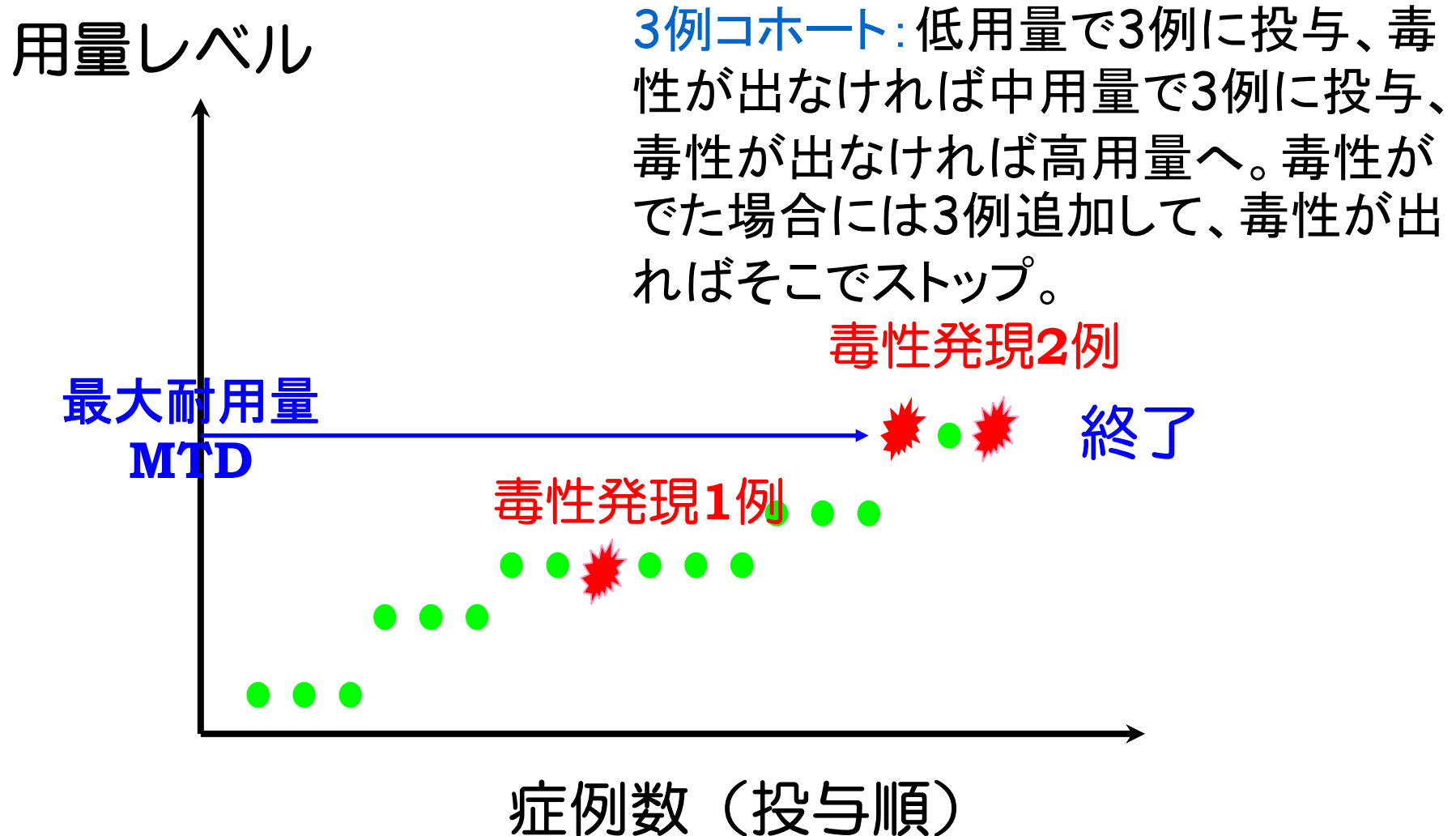
“More is better” with acceptable toxicity



毒性が有効性の surrogate endpoint

第Ⅰ相臨床試験の典型的なデザイン

目的：毒性による第Ⅱ相へのスクリーニング



Phase1論文を読むにはあまり統計は必要ない

■ Phase1臨床試験の目的

- 試験治療の最適用量を決める
- 毒性によるスクリーニング

■ 主なデザイン

- 3例コホート
 - 3例ずつ同一用量を投与し、毒性で評価した最大耐用量となるまで用量を増加
 - やるためにも読むためにも統計は必要ない
- CRM (continuous reassessment method)
 - 1例ごとに統計的に最適と思われる用量を投与し、用量反応曲線を推定
 - ベイズ流統計学の手法を用いているが、3例コホートに比べ最適用量付近が選択されやすい優れた方法であることを知っていればよい



phase2臨床試験論文を読む のに必要な統計

Phase2論文を読むのに必要な統計知識も最小限でいい

■ Phase2臨床試験の目的

- Surrogate endpointにより、有効性でスクリーニング
- Phase3で標準治療とランダム化比較する候補となるか

■ 主なデザイン

- 閾値(threshold)、期待値(expected)を用いたデザイン
- ランダム化phase2 selectionデザイン
- その他のデザインも統計的な考え方はサンプルサイズ設計と判断の目安にしか用いない

閾値、期待値を用いたデザイン

- 閾値(threshold)、期待値(expected)を用いたデザイン
 - Primary endpoint（response rateなど）に対して、それ以上開発を続けてもしようがない値を閾値とし、期待される値のもとで閾値以上ということが証明できるようにサンプルサイズを設計
 - 途中で既に無効が証明された場合には登録を終了する2 stage designが用いられることも多い
 - 閾値以下であるということが棄却されれば統計的には有効性が期待できる

閾値、期待値を用いたデザイン

- 論文を読む場合に必要なのは、試験治療の対照としている成績との比較可能性をチェック
 - ヒストリカルコントロールとの比較
 - 実際に試験にエントリーされた対象者は、ヒストリカルコントロールの対象者とどのくらい比較可能か
 - とくにtime to event（生存や無増悪生存期間）をエンドポイントとする場合には、ヒストリカルコントロールがよほどstableでないと比較できない
 - Phase2でランダム化比較していれば比較可能性は担保されている
 - サンプルサイズが小さいので検証的でなく、あくまでphase3へのスクリーニングと考える

Randomized P2 Selection Design

- 患者を2つ以上の試験治療にランダムに割り付け、もっとも高い奏効割合が得られた治療が次の研究に進む
 - 2つ以上の試験治療があるときに優先順位をつけるデザイン
 - 非常によい群があった場合、その群を高い確率で選択できるようにサンプルサイズを決定
 - 例) 真のresponseがA35% v B20%であれば、90%の確率で35%のほうを選択するためには29例が必要
 - 検証的試験ではありえない
 - どちらも同じ場合には50%の確率でAを選択!!!
 - あまり変わらない場合には、ちゃんと選べない確率も高い

A Phase II Study of Irinotecan in Combination with 120-h Infusion of 5-Fluorouracil in Patients with Metastatic Colorectal Carcinoma: Japan Clinical Oncology Group Study (JCOG9703)

Atsushi Ohtsu¹, Narikazu Boku¹, Takashi Yoshioka², Ichinosuke Hyodo³, Kuniaki Shirao⁴, Yasuhiro Shimada⁴, Soh Saitoh⁵, Akira Nakamura⁶, Noboru Yamamichi⁷, Seiichiro Yamamoto⁸ and Shigeaki Yoshida¹

Purpose: To evaluate the antitumor effect and feasibility of a combination of irinotecan (CPT-11) and 5-day infusional 5-fluorouracil (5-FU) in a sequential schedule based on our previous combination phase I studies in patients with metastatic colorectal cancer.

Patients and Methods: Forty chemotherapy-naïve patients with metastatic colorectal cancer received 90-min infusion of CPT-11 at a dose of 150 mg/m² on days 1 and 15 and 120-h protracted infusion of 5-FU at 600 mg/m²/day on days 3–7, which were repeated every 4 weeks.

Results: The median number of actually administered courses was five, ranging from one to 14. There were 16 (40%) patients who developed grade 3 or 4 neutropenia. Grade 3 or 4 nausea/vomiting and diarrhea were seen in three (8%) and seven (18%) patients, respectively. Only one early death not related to treatment occurred during the study. There was one complete response and 17 partial responses with a response rate of 45% (95% confidence interval: 29.3–61.5%). With a median follow-up period of 22.5 months for survivors, the median survival and median progression-free survival times were 15.9 and 7.0 months, respectively.

Conclusions: Although the toxicities were modest, this sequentially combined regimen is active and feasible in patients with metastatic colorectal cancer.

Jpn J Clin Oncol 2003;33,28-32
13

STATISTICAL CONSIDERATIONS

The sample size for the study was calculated from an expected response rate of 40% and a minimum of 20% with an α and β error of 0.1, using Simon's two-stage minimax design (9). The estimated sample size was 36 and adding 10% of expected ineligible cases, then a total of 40 patients including 19 patients for the first stage were required. Overall survival was calculated from the date of registration to death due to any cause or to the last contact date, using the Kaplan–Meier method. Progression-free survival was analyzed from the date of registration to date of documented disease progression or, if patients died without disease progression, to date of death.

まとめ

- Phase3へ進むかどうか（閾値・期待値）、どのレジメンを選択するか（選択デザイン）の決定には、毒性の情報や他の開発薬剤の状況なども考慮する必要がある。
- 統計的判断は有効性の足切りの目安として使うことが一般的
 - シングルアームだとだいたい数十名(<50)
 - 次にphase3がやれず、やむを得ず検証的な意味を持たせたい場合は>100
 - ランダム化比較していると~100名
 - ランダム化selection designだと<100
- いずれにしろ、検証的な結果ではないことを理解する

臨床試験phase3論文に必要な統計

JCOG9511を例に

IRINOTECAN PLUS CISPLATIN COMPARED WITH ETOPOSIDE
PLUS CISPLATIN FOR EXTENSIVE SMALL-CELL LUNG CANCER

KAZUMASA NODA, M.D., YUTAKA NISHIWAKI, M.D., MASAOKI KAWAHARA, M.D., SHUNICHI NEGORO, M.D.,
TAKAHICO SUGIURA, M.D., AKIRA YOKOYAMA, M.D., MASAHIRO FUKUOKA, M.D., KIYOSHI MORI, M.D.,
KOSHIRO WATANABE, M.D., TOMOHIDE TAMURA, M.D., SEIICHIRO YAMAMOTO, PH.D., AND NAGAHIRO SAIJO, M.D.,
FOR THE JAPAN CLINICAL ONCOLOGY GROUP*

Noda et al. *New Eng J Med* 2002;346:85-91

まずabstractで研究の概略と質を把握

■ 研究仮説を探す

□ 比較する治療法は何か、どちらが標準か

- 下手な標準を選んでいる試験は解釈ができない

□ primary endpointは何か

- 全生存期間（any cause of deathまでの期間）がもっともよい
 - 死因別死亡はあまり使われない
- Progression-free survivalやtime-to-treatment failureの場合はそれがその領域で確立されているエンドポイントでなければ質を落として評価

まずabstractで研究の概略と質を把握（続き）

■ デザインは

□ 優越性 or 非劣性

- Toxic newなら優越性、less toxic newなら非劣性
- 非劣性の場合、臨床的に意味のあるsecondary endpointが定義されていて、それが標準治療に勝っていることが必要

■ 試験が完遂しているか

- 予定通りのサンプルサイズで終了しているか、途中終了の場合、理由は何か
- 追跡は十分か

ABSTRACT

Background Irinotecan hydrochloride, a topoisomerase I inhibitor, is effective against small-cell lung cancer. In a phase 2 study of irinotecan plus cisplatin in patients with extensive small-cell lung cancer, there was a high response rate and a promising median survival time.

Methods We conducted a multicenter, randomized, phase 3 study in which we compared irinotecan plus cisplatin with etoposide plus cisplatin in patients with extensive (metastatic) small-cell lung cancer.

Results The planned size of the study population was 230 patients, but enrollment was terminated early because an interim analysis found a statistically significant difference in survival between the patients assigned to receive irinotecan and cisplatin and those assigned to receive etoposide and cisplatin; as a result, only 154 patients were enrolled. The median survival was 12.8 months in the irinotecan-plus-cisplatin group and 9.4 months in the etoposide-plus-cisplatin group ($P=0.002$ by the unadjusted log-rank test). At two years, the proportion of patients surviving was 19.5 percent in the irinotecan-plus-cisplatin group and 5.2 percent in the etoposide-plus-cisplatin group. Severe or life-threatening myelosuppression was more frequent in the etoposide-plus-cisplatin group than in the irinotecan-plus-cisplatin group, and severe or life-threatening diarrhea was more frequent in the irinotecan-plus-cisplatin group than in the etoposide-plus-cisplatin group.

Conclusions Irinotecan plus cisplatin is an effective treatment for metastatic small-cell lung cancer. (N Engl J Med 2002;346:85-91.)

- 研究仮説は何か？
- Primary endpointは？
 - IPはEPよりも生存期間を延長するか
- デザインは？
 - IPのEPに対する優越性
- 試験が予定通り完遂しているか
 - 中間解析によりCPの優越性が証明されたため途中中止、追跡は十分でない

Introduction & Methods

■ Introduction

- 研究を行うに至った背景、特に対象と標準治療の決め方、なぜその新治療が有効であると考えられるかの説明
- それらを踏まえ、どのようなデザインが適当かを自分で考えてみる

■ Methods

- どんなグループが行っているか
- Eligibilityの記載は正確か
- 統計記述は十分か
 - Sample size設計の根拠
 - 優越性：期待する治療効果が大きすぎないか→検出力不足
 - 非劣性：許容域が広すぎないか→有意にでやすい
 - 中間解析のやり方を含め、解析方法がきちんとかかれているか
- 多少でもデータマネジメントについて記述があるか

Study Design and Statistical Analysis

This trial was designed as a multicenter, prospective, randomized phase 3 study. The study protocol was approved by the Clinical Trial Review Committee of JCOG and the institutional review board of each participating institution before the initiation of the study, and all the patients provided written informed consent before randomization in accordance with the policies of JCOG in effect in 1995, when enrollment began. The primary end point was overall survival, and the secondary end points were the rates of complete and overall response, progression-free survival, sites of relapse, and toxicity. The sample size initially planned was 230 patients from 54 participating sites, with 115 patients in each group. The planned duration of accrual was 3 years, and the planned follow-up time was 1.5 years. This sample size was designed to provide the study with 80 percent power to detect an improvement of 9 months in the median survival of the patients in the etoposide-plus-cisplatin group and an improvement of 13 months in the median survival of patients in the irinotecan-plus-cisplatin group (hazard ratio, 0.69) with a one-sided type I error of 0.05.

All comparisons of patients' characteristics, prognostic variables, response rates, and rates of toxic effects were performed with Fisher's exact test, except for age, for which the t-test was used. Survival was measured as the date of randomization to the date of death or the date of the most recent follow-up. Progression-free survival was measured as the date of randomization to the date of the first observation of disease progression or the date of death from any cause if there had been no progression. If there was no progression and if the patient had not died, data on progression-free survival were censored as of the date that the absence of progression was confirmed. If a patient died without information on progression, data on progression-free survival were censored as of the last date on which progression could be ruled out by review of follow-up forms. Survival curves were calculated by the Kaplan-Meier method¹⁰ and compared with use of the log-rank test.

Two interim analyses were planned, with adjustment for multiple comparisons taken into account by the method of Lan and DeMets.¹¹ The O'Brien-Fleming type alpha spending function was used. The first interim analysis was planned for the date on which half the planned number of patients had been enrolled, and the second for the date on which all the patients had been enrolled. The boundaries were calculated with the use of computer programs provided by Reboussin et al.¹² The current study was designed and conducted on the basis of one-sided testing, but the results are presented with two-sided P values. Unadjusted P values are reported because of the conservative spending function used.

All patient-information forms were collected and managed at the data center. In-house interim monitoring was performed at the data center to ensure the submission of data, the eligibility of the patients, compliance with the protocol, safety, and progress of the study on schedule. The monitoring reports were submitted to and reviewed by an independent monitoring committee semiannually.

■ 統計記述は十分か

- Sample size設計の根拠
- 中間解析のやり方
- 解析方法
- データマネジメントの記述

研究の流れ

- 試験の流れ図があるか、なければ描いてみるとよい
 - CONSORT声明に書かれているような図
 - 途中で脱落した人がどのくらいいるか
 - 治療のコンタミがどの程度あるか
 - ちゃんとITT解析しているか

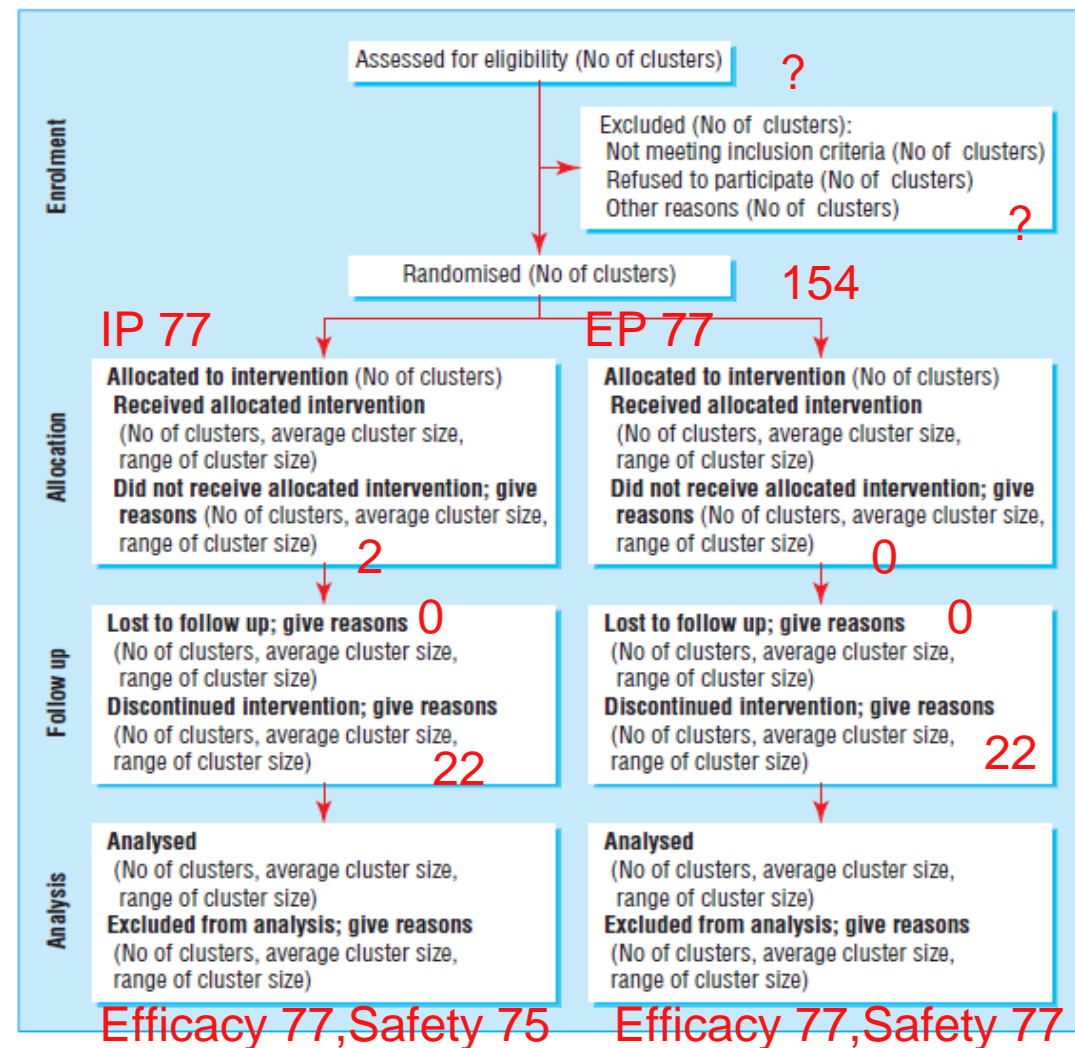


Fig 1 Recommended format for flow diagram of the progress of clusters and individuals through the phases of a randomised trial

Results:背景因子の比較

- あまり大きなアンバランスがないか確認、もしあれば調整した解析や結果の解釈につながるか考えるため

TABLE 1. CHARACTERISTICS OF THE PATIENTS.

CHARACTERISTIC	IRINOTECAN PLUS CISPLATIN (N=77)	ETOPOSIDE PLUS CISPLATIN (N=77)	P VALUE
Age (yr)			0.12
Median	63	63	
Range	30–70	41–70	
Sex			0.25
Male	63	69	
Female	14	8	
ECOG performance status*			0.61
0	10	9	
1	61	58	
2	6	10	

しかし、背景因子ごとにp値を計算する統計的な根拠が無いため、p値に重きを置く必要はない。

ランダム化で期待しているのは、“同じ治療をするとした場合”の結果であって、背景因子毎のバランスではない

Results:

- Toxicity
 - サンプルサイズ計算をしていないのでp値は参考程度にしか過ぎない、有意かどうかだけで判断すべきでない
- 治療のcompliance
 - 経験上、定義とデータを取るのが難しいと思われるので、そのことを理解して解釈する
- 有効性のsecondary endpoint
 - Phase3では、responseはあくまで参考値
 - Eligibilityに入っていないければ全員について計測できない
 - central reviewなどもしないことが多い
 - QOLはそれほど重視されていないのが現状

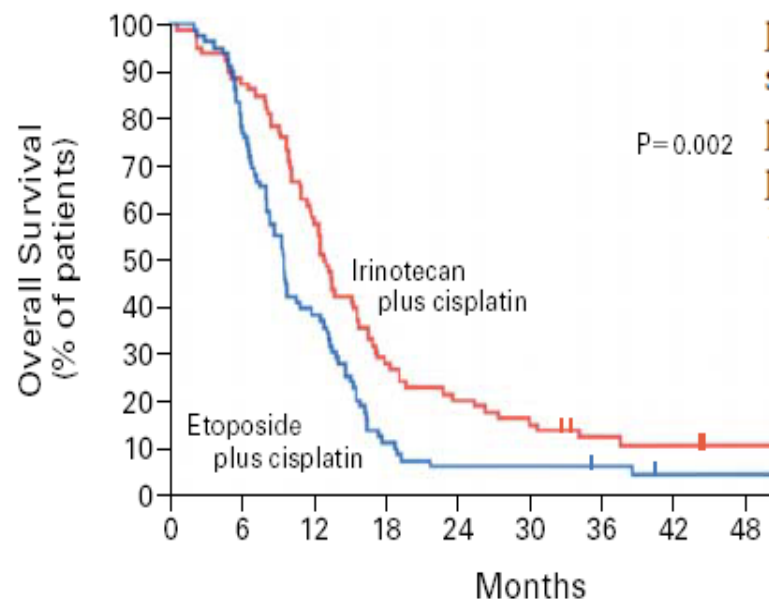
Results: Primary endpointの比較

- Primary endpointの比較
 - 全生存期間がもっともよい
 - PFSがprimary endpointである場合でも、OSは大きな意味を持っている
- 用いられる統計手法
 - Kaplan-Meier curveによる生存曲線の比較
 - 打ち切りがどのくらいあるか見る
 - ひげはついていない場合もある
 - Logrank検定による比較
 - Generalized Wilcoxon検定は初期のイベントに大きな重みをおいている
 - 有意であるかどうかとともに、効果の大きさ（ハザード比やMSTの差）についても着目する
 - Cox回帰
 - ハザード比を出すのに用いる（単変量、多変量ともに）
 - 多変量調整した結果はあくまで補助的なものとする

Discussion:結果の解釈：優越性試験

- 試験がしっかり計画、実施、解析された場合
 - Primary endpointで有意な場合、新治療が勝っていると判断
 - 効果の大きさについても考慮
 - Primary endpointで有意でない場合、全体では新治療が勝っているとは言えない
 - 探索的にサブグループアナリシスをして、次につながる仮説を立てる
- 試験がしっかり計画、実施、解析されていない場合
 - サンプルサイズが小さい、途中でリクルートできず終わったなどは、有意であっても偶然の可能性が高い
 - 探索的な試験として考えるべき

結果と解釈



Overall Survival

As of March 2001, when the final analysis was conducted, the median overall survival was 12.8 months (95 percent confidence interval, 11.7 to 15.2) in the

irinotecan-plus-cisplatin group and 9.4 months (95 percent confidence interval, 8.1 to 10.8) in the etoposide-plus-cisplatin group; 70 patients in the irinotecan-plus-cisplatin group and 74 in the etoposide-plus-cisplatin group died ($P=0.002$ by the log-rank test) (Fig. 1). The rate of overall survival in the irinotecan-plus-

DISCUSSION

We consider that the trend toward a higher complete-response rate in the etoposide-plus-cisplatin group than in the irinotecan-plus-cisplatin group is due to chance. Although it is possible that these results occurred by chance, we believe that the decision to terminate the trial early was based on generally accepted scientific and ethical principles and that, despite the small sample size, we can conclude that the combination of irinotecan and cisplatin is an attractive option for patients with metastatic small-cell lung cancer who have a good performance status.

Discussion:結果の解釈：非劣性試験

- 試験がしっかり計画、実施、解析された場合
 - Primary endpointで有意な場合かつ、2ndary endpointで新治療の有効性が示されている場合、新治療が優れているor標準治療のオプションになりえると判断
 - 非劣性の許容域が広すぎないことに注意（5生率で10%やハザード比1.5など）
 - Primary endpointで有意な場合かつ、2ndary endpointで新治療の有効性が示されていない場合、標準治療のオプションになりえるかは判断が難しい
 - Primary endpointで有意でない場合、新治療は標準治療やオプションとなれない
 - ただし、信頼区間で解釈して許容域を再定義することができる（5生率で3%劣っていないことを許容域として棄却できなかったが、5%は劣っていないことは言える場合がある）
- 試験がしっかり計画、実施、解析されていない場合
 - 非劣性試験は、きっちり行わないと有意=非劣性が証明されやすいので注意
 - 治療の不遵守で両群が同じ治療をしたり、測定の変異があると差が消えていく

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**.

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

サブグループ アナリシス記述例

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812

JULY 31, 2008

VOL. 359 NO. 5

D2 Lymphadenectomy Alone or with Para-aortic Nodal Dissection for Gastric Cancer

Mitsuru Sasako, M.D., Takeshi Sano, M.D., Seiichiro Yamamoto, Ph.D., Yukinori Kurokawa, M.D.,
Atsushi Nashimoto, M.D., Akira Kurita, M.D., Masahiro Hiratsuka, M.D., Toshimasa Tsujinaka, M.D.,
Taira Kinoshita, M.D., Kuniyoshi Arai, M.D., Yoshitaka Yamamura, M.D., and Kunio Okajima, M.D.,
for the Japan Clinical Oncology Group

institution. Analyses of two prespecified subgroups (Borrmann macroscopic type and clinical T stage) and nine post hoc subgroups were also conducted to evaluate interactions between treatment and subgroup with the use of Cox regression; we report the result of all these analyses. No more than one significant interaction test result ($P < 0.05$) would be expected on the basis of chance alone as a result of multiple testing.

Phase3臨床試験論文：まとめ

- Abstract：研究の概略と質を把握
 - 研究仮説は何か？、Primary endpointは？デザインは？試験が予定通り完遂しているか
- Introduction
 - 研究を行うに至った背景、特に対象と標準治療の決め方、なぜその新治療が有効であると考えられるかの説明
- Methods
 - どんなグループ？Eligibilityの記載は正確か？統計記述は十分か？データマネジメントについて記述？研究流れ図が参考になる
- Results & Discussion
 - Overall survivalの結果が重要
 - 有意かどうか、どこまでいえるか（何が検証的で何が探索的か）を理解しておく

観察研究論文を読むのに 必要な統計

レトロで治療成績を評価

観察研究で治療効果を議論する論文(1)

- 始めに集団を設定して前向きに追跡した研究
 - きっちりデータが取れるが、ランダム化していないのであくまで探索的
- 研究ではなく、後でカルテから起こした調査
 - 擬似前向き研究
 - レトロの研究と呼ぶことが多い
 - ケース・コントロール研究と呼ぶ人がいるが間違い
 - データがきっちりあるか不明、事後的に対象者を都合よく選択している可能性もあるので、より質が低い

観察研究で治療効果を議論する論文(2)

- 目的：治療効果の（何らかの対照）との比較
 - ヒストリカルコントロールの成績
 - 非ランダム化対照群との比較
- 比較の方法：Endpointは全生存などtime-to-eventが多い
 - Kaplan-Meier curve, logrank検定, Cox回帰
 - 多変量調整したCox回帰がメインの結果
 - 予後因子、治療効果の選択に関わるような因子をConfounding factorsとしてできるだけ多く調整しているか
- いくら多変量調整しても、正しい結果かどうかはわからない
 - いずれにしろ探索的に解釈

NONRANDOMIZED COMPARISON BETWEEN DEFINITIVE CHEMORADIOOTHERAPY AND RADICAL SURGERY IN PATIENTS WITH T₂₋₃N_{any} M₀ SQUAMOUS CELL CARCINOMA OF THE ESOPHAGUS

SHUICHI HIRONAKA, M.D.,* ATSUSHI OHTSU, PH.D.,* NARIKAZU BOKU, M.D.,*
MANABU MUTO, M.D.,* FUMIO NAGASHIMA, M.D.,* HIROKI SAITO, M.D.,* SHIGEAKI YOSHIDA, M.D.,*
MITSUYO NISHIMURA, M.D.,† MASATORA HARUNO, M.D.,‡ SATOSHI ISHIKURA, M.D.,§
TAKASHI OGINO, M.D.,§ SEIICHIRO YAMAMOTO, PH.D.,|| AND ATSUSHI OCHIAI, PH.D.¶

Purpose: To compare the treatment results between radical surgery and definitive chemoradiotherapy for resectable squamous cell carcinoma of the esophagus and to identify useful clinicopathologic and biologic markers to select better treatment.

Methods and Materials: Between August 1992 and April 1999, 98 consecutive patients were selected for this study; 53 were treated with chemoradiotherapy and 45 with surgery. The patients in the chemoradiotherapy group received 5-fluorouracil combined with cisplatin plus 60 Gy of radiation, and those in the surgery group received an esophagectomy with radical node dissection. Biologic markers were investigated immunohistochemically using pretreatment biopsy specimens.

Results: The baseline clinical TNM stage was more advanced in the chemoradiotherapy group than in the surgery group. With a median follow-up period of 43 months, the 5-year survival rate was 46% in the chemoradiotherapy and 51% in the surgery group, without statistical significance ($p = 0.47$, log-rank test). Cox regression analysis for prognosis revealed that epidermal growth factor receptor positivity, high microvessel density, and cyclin D1 positivity yielded a low value for relative risk (0.66, 0.54, and 0.62, respectively), which favored chemoradiotherapy over surgery, without statistical significance.

Conclusion: This nonrandomized study showed a trend for the chemoradiotherapy in the treatment of esophageal carcinoma, but the results need to be confirmed by additional study. © 2003 Elsevier Inc.

Esophageal cancer, Chemoradiotherapy, Surgery, Biologic marker.

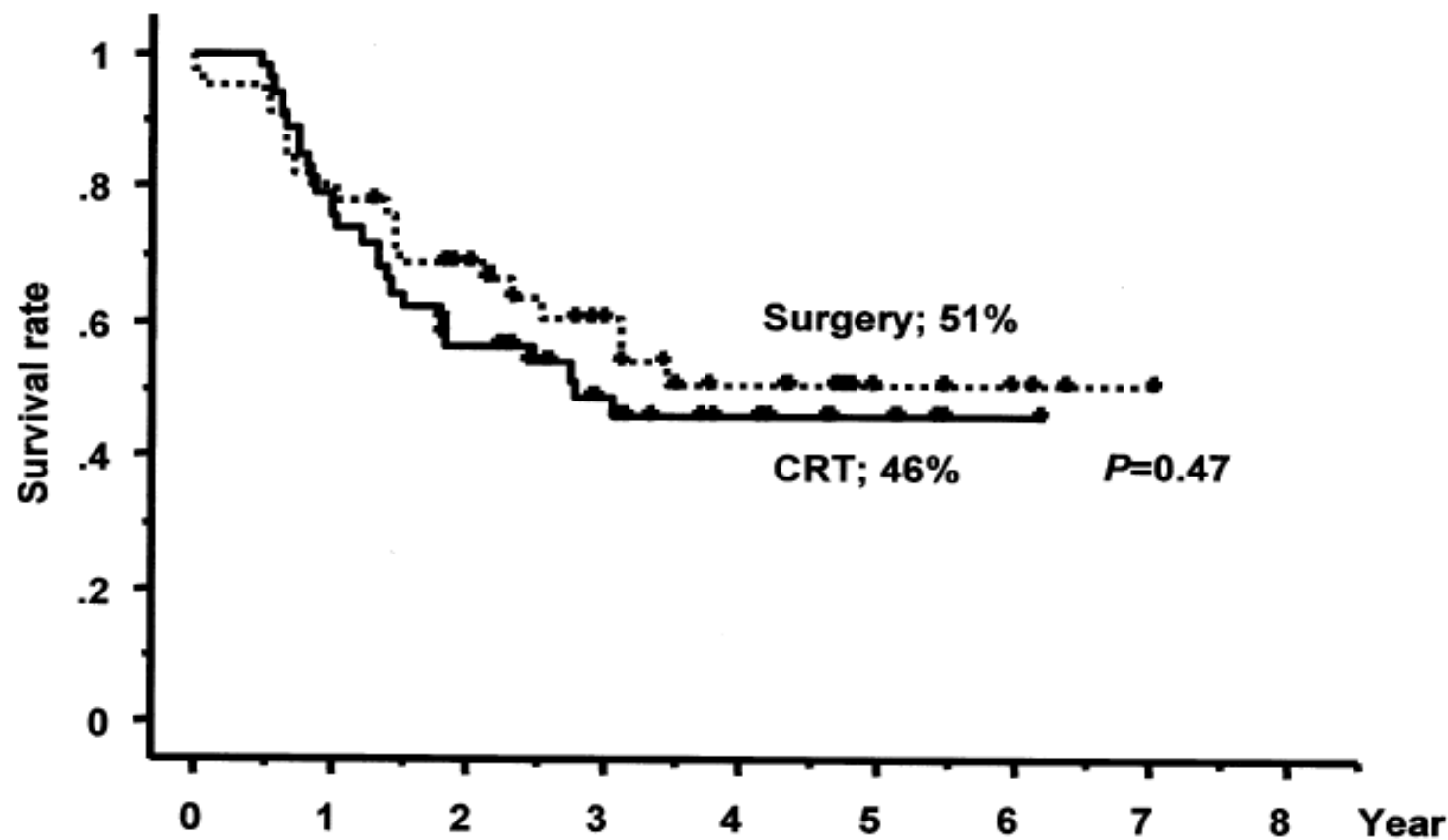


Fig. 1. Survival curves of chemoradiotherapy group (CRT, solid line) and surgery group (dotted line) with Kaplan-Meier method. Median survival was 34 months in chemoradiotherapy group and not reached in surgery group. No statistical significant difference was found in both groups in overall survival ($p = 0.47$).

Table 5. Univariate and multivariate analyses of relative risks of chemoradiotherapy against surgery on survival in 95 esophageal cancer patients

Category	n	Univariate analysis			Multivariate analysis*		
		RR [†]	95% CI	p	RR [†]	95% CI	p
N stage							
0	51	1.64	0.68–3.96	0.27	1.01	0.38–2.71	0.98
1	44	0.67	0.30–1.47	0.32	0.81	0.35–1.87	0.62
p53							
Negative	34	1.46	0.51–4.11	0.48	0.77	0.33–1.80	0.55
Positive	61	1.10	0.54–2.23	0.79	0.96	0.48–1.94	0.91
Ki-67							
≤64	34	1.35	0.49–3.73	0.57	0.77	0.31–1.93	0.58
>64	61	1.10	0.53–2.26	0.81	0.94	0.48–1.85	0.85
EGFR							
Negative	44	1.23	0.55–2.78	0.61	1.26	0.62–2.57	0.53
Positive	51	1.20	0.52–2.79	0.68	0.66	0.30–1.47	0.31
Cyclin D1							
Negative	31	3.86	0.86–17.4	0.08	3.25	0.68–15.5	0.14
Positive	64	0.91	0.45–1.84	0.80	0.62	0.30–1.30	0.21
MVD							
≤50	56	1.82	0.75–4.38	0.19	1.35	0.51–3.55	0.55
>50	39	0.74	0.34–1.64	0.46	0.54	0.22–1.31	0.17
VEGF							
Negative	51	1.24	0.53–2.94	0.62	1.04	0.43–2.50	0.94
Positive	44	0.97	0.43–2.19	0.94	0.77	0.33–1.79	0.54

Abbreviations: RR = relative risk; CI = confidence interval; EGFR = epidermal growth factor receptor; MVD = microvessel density; VEGF = vascular endothelial growth factor.

* Adjusted by N stage, p53, Ki-67, EGFR, cyclin D1, MVD, and VEGF.

† When relative risk >1, surgery had better survival than chemoradiotherapy.

観察研究論文を読むのに 必要な統計

予後因子解析

予後因子探索の研究

- 治療の効果を調べるのではなく、予後に関係する因子を調べていることを認識することが必要
- 予後を調べてどうするか
 - 予後を正確に予測したい
 - 当たればいいのだから、統計的な議論を中心にしてbiologyやclinicalな解釈は二番手に
 - GEPなどhigh throughputデータから統計的に予後予測モデルを立てる
 - 予後に影響する因子を調べたい
 - Biologyやclinicalな意味づけを一番にし、統計結果に振り回されない
 - 臨床的、臨床病理的な予後因子の中からclinicalにリーズナブルなモデルを立てる
- 結果の妥当性
 - 別のデータセットで確認しているかに尽きる

Prognostic Factors in Patients with Hepatocellular Carcinoma Treated by Transcatheter Arterial Embolization

Masafumi Ikeda¹, Shuichi Okada¹, Seiichiro Yamamoto², Tosiya Sato³, Hideki Ueno¹, Takuji Okusaka¹, Hitoshi Kuriyama¹, Kenichi Takayasu⁴, Hiroyoshi Furukawa⁴ and Ryoko Iwata⁴

Background: Transcatheter arterial embolization induces marked antitumor response in patients with hepatocellular carcinoma, but the survival benefit of transcatheter arterial embolization remains to be determined. This study investigated prognostic factors in patients with advanced hepatocellular carcinoma treated by transcatheter arterial embolization.

Methods: A total of 128 consecutive patients with non-resectable hepatocellular carcinoma, who had undergone transcatheter arterial embolization between May 1990 and August 1998, were analyzed to investigate prognostic factors.

Results: Median survival time and survival proportions at 1, 3 and 5 years were 3.3 years, 92.0, 54.6 and 23.4%, respectively. By multivariate analysis using the accelerated failure time model, age <60 years, hepatitis C virus antibody positivity, serum albumin >3.5 g/dl, absence of portal vein invasion and serum α -fetoprotein level <400 ng/ml were significantly associated with favorable survival. For clinical application, we also propose a prognostic equation with combination of specific prognostic factors, by which survival curves of each patient could be predicted directly.

Conclusion: The findings of the current study may be helpful in predicting the life expectancy of hepatocellular carcinoma patients treated by transcatheter arterial embolization and in designing future clinical trials of transcatheter arterial embolization for hepatocellular carcinoma.

Table 2. Median survival times and ratios of survival time by prognostic factors for patients with HCC treated by TAE in univariate and multivariate analyses

		<i>n</i>	Median survival time (years)	Univariate analysis ^a			Multivariate analysis ^c		
				Ratio of survival time	95% CI	<i>P</i>	Ratio of survival time	95% CI	<i>P</i>
<i>Host-related factors</i>									
Age (years)	≥60	90	3.0	0.90	(0.67–1.20)	0.47	0.69	(0.5–0.95)	0.02
	<60	38	3.8	— ^b			—		
Gender	F	26	3.3	0.90	(0.62–1.32)	0.60	0.81	(0.56–1.18)	0.27
	M	102	3.4	—			—		
Alcohol abuse ^d	+	30	3.8	1.02	(0.75–1.39)	0.90	1.10	(0.80–1.51)	0.57
	—	98	3.0	—			—		
HBs Ag	+	21	2.9	0.96	(0.67–1.38)	0.82	1.01	(0.62–1.64)	0.97
	—	107	3.3	—			—		
HCV Ab	+	101	3.4	1.27	(0.92–1.74)	0.14	1.77	(1.11–2.81)	0.02
	—	26	2.9	—			—		
Prior hepatectomy	+	62	3.8	1.25	(0.94–1.65)	0.12	0.95	(0.70–1.27)	0.71
	—	66	2.8	—			—		
Total bilirubin	≥2.0 mg/dl	7	2.7	0.93	(0.53–1.64)	0.81	1.13	(0.69–1.86)	0.63
	<2.0 mg/dl	121	3.3	—			—		
Albumin	>3.5 g/dl	93	3.5	1.30	(1.06–1.48)	0.02	1.37	(1.16–1.54)	<0.01
	≤3.5 g/dl	35	2.7	—			—		
GOT	>82 IU/l ^e	44	3.0	0.89	(0.67–1.18)	0.41	0.77	(0.52–1.14)	0.19
	≤82 IU/l	84	3.5	—			—		
GPT	>70 IU/l ^e	54	3.5	1.05	(0.79–1.39)	0.74	1.11	(0.75–1.63)	0.60
	≤70 IU/l	73	3.0	—			—		
Lactic dehydrogenase	≥500 IU/l	26	3.0	1.19	(0.81–1.74)	0.38	1.20	(0.84–1.70)	0.32
	<500 IU/l	102	3.3	—			—		
Cholinesterase	≥100 IU/l	115	3.4	0.64	(0.41–1.00)	0.05	0.91	(0.57–1.46)	0.71
	<100 IU/l	11	2.5	—			—		

		<i>n</i>	Median survival time (years)	Univariate analysis ^a			Multivariate analysis ^c		
				Ratio of survival time	95% CI	<i>P</i>	Ratio of survival time	95% CI	<i>P</i>
<i>Tumor-related factors</i>									
No. of tumors	Multiple	90	3.0	0.70	(0.50–0.98)	0.04	0.76	(0.51–1.13)	0.17
	Single	38	3.8	–			–		
Tumor distribution	Bilateral	55	2.9	0.68	(0.51–0.91)	0.01	0.78	(0.56–1.09)	0.15
	Unilateral	73	3.8	–			–		
Tumor size	>25%	18	1.0	0.54	(0.38–0.77)	<0.01	0.78	(0.56–1.10)	0.16
	≤25%	110	3.5	–			–		
Portal vein invasion	+	4	0.8	0.81	(0.42–1.54)	0.52	0.58	(0.30–1.11)	0.10
	–	124	3.3	–			–		
α-Fetoprotein	≥400 ng/ml	26	1.4	0.56	(0.40–0.77)	<0.01	0.64	(0.44–0.93)	0.02
	<400 ng/ml	101	3.5	–			–		
<i>Treatment-related factors</i>									
Anticancer agents	MMC + ADR	17	2.9	0.83	(1.18–9.86)	0.54	0.91	(0.64–1.28)	0.59
	SMANCS	24	3.5	1.08	(0.58–1.19)		0.86	(0.55–1.35)	0.52
	ADR	87	3.5	–			–		

HBs Ag, hepatitis B surface antigen; HCV Ab, hepatitis C virus antibody; GOT, glutamic oxaloacetic transaminase; GPT, glutamic pyruvic transaminase; ADR, doxorubicin; MMC, mitomycin; SMANCS, zidovudine. ^aRatio of survival time, its 95% confidence intervals and *P*-value estimated by univariate accelerated failure time models. ^bReference category. ^cAccelerated failure time models with all prognostic factors in the models. ^dEthanol intake ≥80 g/day for ≥5 years. ^e2.5 times upper normal limit.

論文を読むのに必要な統計：まとめ

■ 臨床試験の論文

- 典型的なPhase1, 2論文を読むのにあまり統計はいらない
 - 統計に惑わされずに読む
- Phase3には研究の質の評価と正しい解釈のために必要

■ 観察研究の論文、レトロの論文

- レトロで治療法を評価
 - 探索的な結果、さらに検証が必要だと思って読む
- 予後因子探索
 - 別のデータセットで確認しているか

参考資料

- *CONSORT statement: extension to cluster randomised trials* Marion K Campbell, Diana R Elbourne, Douglas G Altman for the CONSORT Group *BMJ* 2004;328:702–8
- Wang R et al. *Statistics in medicine -reporting of subgroup analyses in clinical trials. New Eng J Med, vol.357:pp2189-94. 2007.*
- 福田治彦、大橋靖雄. Japanese Journal of Clinical Oncology投稿に際しての統計解析結果のレポートに関するガイドライン (<http://www.jjco.org/UsefulInfo/guideline.html>)
- ICR臨床試験入門 website (<http://www.icrweb.jp/icr/>)
- Green S, Crowley J, Benedetti. (翻訳福田治彦他) 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
- 大橋靖雄, 荒川義弘編集. 臨床試験のすすめ方. 南江堂2006