ティングの開催，他部門との調整，同意説明補助，被験者スケジュール管理，症例報告書作成支援，モニタリング・監査対応，被験者ケア・治験相談窓口

④コンサルテーション部門：自主臨床試験のプロトコール等の作成支援，未承認薬等の臨床使用における申請支援，試験薬のマスク化（コード化），セミナー・講習会の企画，その他のコンサルテーション業務

なお，審査については，部全体で支援することとしている．また，現在，治験コーディネータ（Clinical Research Coordinator：CRC）は，医薬品および医療機器の治験の全てに配置するようにしている．

2）自主臨床試験の実施支援体制

未承認薬や適応外の薬剤を臨床試験で使用する場合は，試験薬の管理・調剤は臨床試験部で行うようにしている．また，まだ1試験ではあるが，二重盲検比較試験における薬剤を，臨床試験部でマスク化（コード化）し，被験者割付を行っている．IRBによる実施状況管理の側面からは，全ての自主臨床試験および未承認薬等の臨床使用を対象に行っており，臨床試験部用の同意書の回収，一部変更申請，重篤な有害事象報告，実施状況報告（年1回），終了・中止報告などの提出を求めている．また，安全性情報についても，緊急安全性情報や安全性等に係る添付文書改訂などの情報を試験責任医師に提供し，「新たな安全性情報の報告」の病院長への提出や，実施計画書や同意説明文書の改訂について検討してもらうようにしている．

コーディネータ業務については，現在4試験を対象に試行している．その内容は，スタートアップミーティングの開催，同意説明補助，被験者来院日対応となっており，被験者に対する倫理性確保に重点を置いている．

さらに，「医師主導の治験」では医療機関が品質保証を行うことになっていることから，その導入の問題点を探るため，1試験を対象に臨床試験部がモニタリングを試行する予定である．

おわりに

治験や自主臨床試験のグローバリゼーションや「医師主導の治験」の導入にも対応できるように，「Quality & Speed」をキーワードに支援体制の構築を図ってきた．まだ試行段階の部分も多いが，かなりの対応が可能になったと考えている．治験のみならず，自主臨床試験にも企業との連携により行うものが増えてきており，明らかに企業が主体となって行う「自主臨床試験」は受託研究契約を締結して実施している．今後，産学連携のあり方については，経費の負担や「利益の衝突」などの点についてさらに検討が必要と思われる．

一方，患者さんに対する治験・臨床試験の啓蒙や治験情報の提供の拡大，実施率をさらに向上させるための進捗管理，臨床試験の計画・実施者に対する教育研修について順次計画を進めている所である．当院では，治験の導入についてもコンサルテーションを実施しているので，当院へ申請を考慮される場合はご利用いただきたい．

なお，本稿に記載した内容は当院臨床試験部，臨床試験部運営委員会，治験審査委員会が主体となり，医学部倫理委員会等と連携して改革を進めてきたものである．また，医師主導の臨床試験の指針や手引き等の策定にあたっては，学内外の多くの方からの貴重な意見をいただいた．ここに謝辞を申し上げる．

文　献

1) 金井文彦，荒川義弘，小俣政男．自主臨床試験の実施計画書作成の手引き（東京大学医学部附属病院版）について．臨床薬理 2003；34：101-2.

1) GCP および ICH ガイドラインに準拠した試験デザイン

東京大学医学部附属病院臨床試験部　荒川　義弘
Yoshihiro Arakawa
（副部長，助教授）

Summary

世界標準である GCP に則って臨床試験を計画・実施することで，質の高い試験を実施することができる。日本においては治験のみが GCP の対象であるが，研究者主導の自主臨床試験も GCP を準用することが望まれる。本稿では，GCP や他のガイドラインの基本的事項について記載するので，試験のデザインの一助にしていただきたい。

Key Words　治験，臨床試験，GCP，ICH ガイドライン，エビデンスの形成

新しい臨床試験の枠組

EBM (Evidence-Based Medicine) の重要性が唱えられるようになり，研究者主導の臨床試験（自主臨床試験）が日本でも盛んになりつつある。また，改正 GCP の施行（2003 年 7 月）により，従来，企業しか行うことができなかった治験（医薬品の承認申請のために行う臨床試験）が，医師主導でも実施で

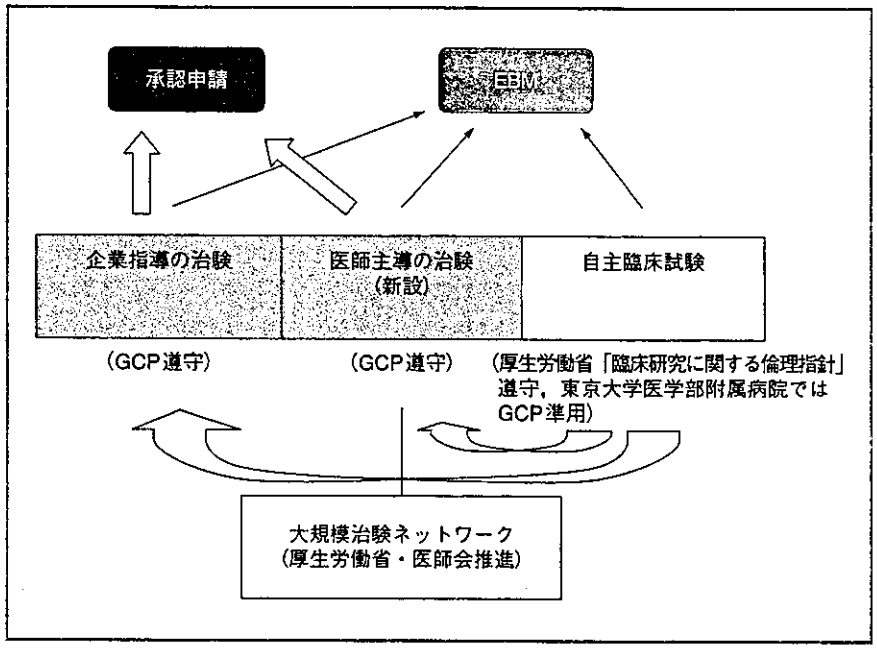


図 1　臨床試験の新しい枠組

Surgery Frontier 11(4)：77-81，2004

表1　臨床研究の倫理的原則

・ニュルンベルグ綱領(1947年，ニュルンベルグ裁判)
(http：//www.ushmm.org/research/doctors/codeptx.htm)
・ヘルシンキ宣言(1964年，世界医師会；最終改訂2000年)
(http：//www.med.or.jp/wma/helsinki02_j.html)
・ベルモントリポート(1979年，米国)
(http：//www.fda.gov/oc/ohrt/irbs/belmont.html)
・臨床研究の倫理指針(2003年7月，厚生労働省)
(http：//www.imcj.go.jp/rinri/index.html)

きるようになった(図1)。日本医師会治験促進センターは，厚生労働省の委託を受け，参加施設公募により大規模治験ネットワークを形成し，医師主導の治験の制度を利用して採算が取れないなどの理由で企業主導では開発されない医薬品の適応の承認取得を目指している。

一方，自主臨床試験の結果から治験に発展することも少なくない。東京大学医学部附属病院でも現在進行中の治験のいくつかは，同病院での自主臨床試験の結果に基づいて組まれたものである。社会的にも，アンジェス(株)による閉塞性動脈硬化症に対する遺伝子治療の治験が大阪大学での自主臨床試験の結果に基づいていることは有名である。

このように，エビデンスの形成を目指した新しい臨床試験の枠組ができてきている(図1)。エビデンスに貢献するためには“質の高い臨床試験”である必要がある。“質の高い臨床試験”とは倫理的原則を守りつつ，科学的デザインに基づき信頼性のあるデータを産み出す臨床試験のことであり，そのための国際的ルールとガイドラインが策定されてきている。「医薬品の臨床試験の実施の基準」(Good Clinical Practice：GCP)は，その国際的ルールであり，欧米では治験だけでなく自主臨床試験をも対象としているが，日本では治験のみが対象である。治験であれ，自主臨床試験であれ，ヒトを対象とする臨床試験である以上，倫理的かつ質の高い臨床試験でなければならない。また，一流誌に採用されるためにも，質の高い臨床試験であることが要求される。臨床試験の結果を投稿論文にまとめる際にもコンソート声明のようなガイドラインがある。ルールやガイドラインに従うことは，型にはまって行うことで質を上げることができるという側面と，互いにルールを守ることで信頼できるデータであることを示すことができるという側面がある。このような背景から，東京大学医学部附属病院では2002年度より自主臨床試験にもGCPを準用した指針，手引き，審査と指導体制を整備し，運用を開始している(http://www.h.u-tokyo.ac.jp/gcp/doctors/home 2.htm)。

臨床研究の倫理的原則

臨床試験を計画し，実施するためには，臨床研究の倫理的原則についてよく理解しておく必要がある。主なものを表1に示す。ニュルンベルグ綱領はナチの非人道的な人体実験を裁く裁判の判例の中で「許容できる医学実験」として示されたもので，自由意思による同意など10の項目に現在の研究倫理の原点が凝縮されている。ヘルシンキ宣言は，世界医師会から出され，改訂を重ねてきている。ヘルシンキ宣言は，GCPでもその遵守が義務付けられている。また，厚生労働省の「臨床研究の倫理指針」もヘルシンキ宣言をもとに策定された。したがって，これらは臨床研究に携わるものが熟知していなければならないものである。

ICH：日米EUの臨床試験の規制要件の標準化

日米EUのそれぞれの規制当局と製薬業界の代表6者が協議し，医薬品の承認や市販後の安全性調査に関わる規制要件の標準化を図り，非臨床試験のデータの共通化や臨床試験データの相互利用が可能となった。これにより無駄な試験が削減され，迅速な申請が可能となった。ICH (The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use)とは，この協議のことであり，現在もさまざまなトピックスについて標準ガイドラインの策定を目指し協議が継続されている。各ガイドラインは

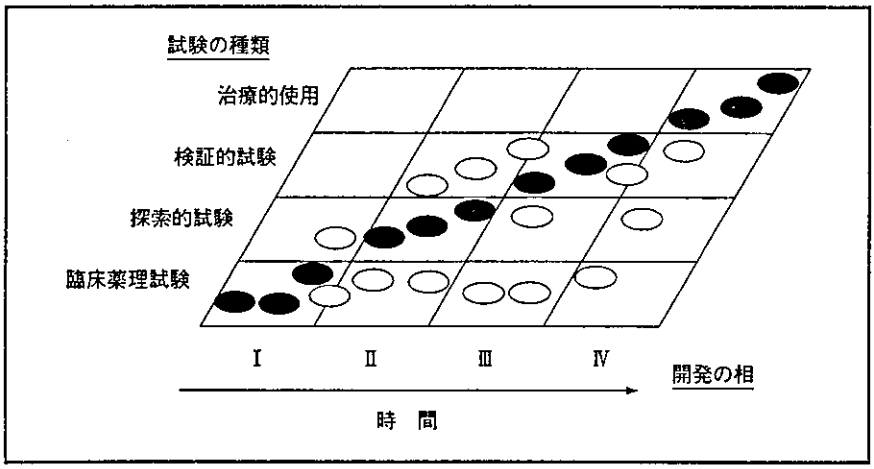


図2　開発の相と試験の種類の関係

黒丸はその相で最も一般的に実施される試験，白丸はその相で比較的まれに実施される試験。

（厚生労働省：臨床試験の一般指針，1998より改変引用）

表3　GCPに準拠したプロトコールや手順書等に記載されるべき項目

科学性	・背景，目的，方法，評価項目，統計解析，参考資料・文献 ・IRBによる審査
倫理性・安全性	・医師の要件・資務，利益の衝突 ・除外・中止基準等 ・自由意思によるインフォームド・コンセント，個人情報保護 ・重篤有害事象報告，健康被害の補償 ・被験者の費用負担，被験者への支払い ・ヘルシンキ宣言・GCPへの対応 ・定期実施状況報告，終了報告 ・IRBによる審査・調査
信頼性	・被験者登録，データ管理，モニタリング・監査，手順書の整備 ・試験薬の管理 ・逸脱報告 ・記録の保存，結果の公表 ・IRBによる審査・調査

第Ⅰ相，第Ⅱ相，第Ⅲ相というのは，本来，臨床開発の段階を示す言葉であり，試験の種類を表す言葉ではない。しかし，しばしば第Ⅰ相で実施されることの多い臨床薬理試験のことを第Ⅰ相試験とよぶ。実際には，腎障害患者や相互作用が懸念される薬剤を服用中の患者など特定の患者を対象とした臨床薬理試験が第Ⅲ相に必要に応じて実施されることがある。開発の相と試験の種類の関係は，「臨床試験の一般指針」（ICH-E 8）に示されている（図2）。

第Ⅱ相は，少数例の患者を対象に有効性と安全性を探索的に評価する段階であり，探索的な用量反応試験が組まれることが多い。抗がん剤の治験では，第Ⅱ相の試験のデータをもって医薬品としての承認申請がなされる。

第Ⅲ相は，それまでのデータをもとに，より幅広い患者を対象に検証的に試験を実施する段階である。検証的試験では，主要評価項目を1ないし2に限定して設定し，この項目で意図する結果が得られなければ失敗（仮説は検証できない）とみなすことになっている。これにより多重検定の問題を回避することができる。

GCPに準拠した試験デザイン：プロトコールの作成

GCPの骨子は，科学性，倫理性，信頼性である。これらに該当するGCP上の項目をあげてみると，表3のようになる。なお，これらを審査する治験審査委員会のことを米国での名称にならってIRB（Institutional Review Board）とよぶことが多い。

試験をデザインするときは，まず目的を明確にし，科学性，倫理性および実施可能性を考慮してアウトラインを作成する。この際，ICHのEfficacyの項目の各ガイドライン（表2）を参照することが推奨される。次に，さらに信頼性等の項目や細部を加えて試験実施計画書（プロトコール）を完成させる。

東京大学医学部附属病院では自主臨床試験の申請者に対してプロトコール作成等のコンサルテーションを行っている。経験の浅い研究者に多い事例と

して，目的，主要評価項目，方法の間の不整合があげられる。まず単純でかつ望ましいデザインであるランダム化並行群間比較試験を出発点にして，倫理性や実施可能性を考慮して試験をデザインするとよい。客観的な評価項目であれば，必ずしも盲検試験にする必要はない。客観的な評価項目とはいえない場合は，どちらの群の患者か知らない第三者が評価する方法(PROBEデザイン：Prospective Randomized Open Blinded-Endpoint design)もしばしばとられる。

比較対照群の設定が困難な場合がある(ICH-E 10参照)。実施可能性を考慮すれば，標準治療への上乗せ効果をみるのもよい方法であるが，検出力が低下することは否めない。従前の標準的な治療法が無効あるいは適用できない患者を対象にして比較試験を組むこともある。歴史的対照(Historical control)や疫学的発現率等を比較対照にすることもあるが，妥当性についての十分な検討が必要である。主要評価項目に発現率の低い事象(例：心血管系のイベント)を設定する場合は，多数の症例や長期間の評価が必要であり，探索的な臨床試験には向かない。このような場合は代替評価項目(例：血圧，血清コレステロール)を使用する。東京大学医学部附属病院では独自に「自主臨床試験の実施計画書作成の手引き」(http://www.h.u-tokyo.ac.jp/gcp/doctors/home 2.htm)を作成して公開しているので，テンプレートとして使用することをお勧めする。

被験者登録割付とデータマネジメントは，試験の実施者(データの作成者)とは別の者が行うのが，信頼性の確保の上で重要である。生物統計家のアドバイスが受けられるのであれば，症例データの固定後に行う統計解析だけでなく，プロトコール作成時にもアドバイスを受けるとよい。試験の結果を総括報告書にまとめるときはICH-E 3を参照し，投稿論文にまとめるときはCONSORT声明(http://www.consort-statement.org/)を参照するとよい。

おわりに

米国では治験も自主臨床試験も明確な区別はないといわれる。信頼性のあるデータであれば，それをエビデンスとして新たな治験が組まれたり，承認申請に利用されたりする。投稿論文でも，データねつ造の問題が取りざたされるようになってきていることから，近い将来自主臨床試験でも信頼性保証が要求されるようになるかもしれない。そのようなことを鑑みて試験をデザインすべき時代になってきている。

Available online at www.sciencedirect.com

SCIENCE DIRECT

Controlled Clinical Trials 25 (2004) 572–584

Controlled Clinical Trials

www.elsevier.com/locate/conclintrial

Statistical comparison of random allocation methods in cancer clinical trials

Atsushi Hagino[a,b,*], Chikuma Hamada[a], Isao Yoshimura[a], Yasuo Ohashi[c], Junichi Sakamoto[d], Hiroaki Nakazato[e]

[a]*Department of Industrial Management and Enginneering, Faculty of Engineering, Tokyo University of Science, Tokyo, Japan*
[b]*Department of Clinical Development, Mochida Pharmaceutical Co., Ltd., Tokyo, Japan*
[c]*Department of Biostatistics/Epidemiology and Preventive Health Science, School of Health Science and Nursing, The University of Tokyo, Tokyo, Japan*
[d]*Department of Epidemiological and Clinical Research Information Management, Kyoto University, Kyoto, Japan*
[e]*Yokoyama Gastrointestinal Hospital, Aichi, Japan*

Received 31 December 2003; accepted 30 August 2004

Abstract

The selection of a trial design is an important issue in the planning of clinical trials. One of the most important considerations in trial design is the method of treatment allocation and appropriate analysis plan corresponding to the design.

In this article, we conducted computer simulations using the actual data from 2158 rectal cancer patients enrolled in the surgery-alone group from seven randomized controlled trials in Japan to compare the performance of allocation methods, simple randomization, stratified randomization and minimization in relatively small-scale trials (total number of two groups are 50, 100, 150 or 200 patients). The degree of imbalance in prognostic factors between groups was evaluated by changing the allocation probability of minimization from 1.00 to 0.70 by 0.05.

The simulation demonstrated that minimization provides the best performance to ensure balance in the number of patients between groups and prognostic factors. Moreover, to achieve the 1 percentile for the p-value of chi-square test around 0.50 with respect to balance in prognostic factors, the allocation probability of minimization was required to be set to 0.95 for 50, 0.80 for 100, 0.75 for 150 and 0.70 for 200 patients. When the sample size was larger, sufficient balance could be achieved even if reducing allocation probability. The simulation using actual data demonstrated that unadjusted tests for the allocation factors resulted in conservative type I errors when

* Corresponding author. Department of Clinical Development, Mochida Pharmaceutical Co., Ltd., 1-22, Yotsuya, Shinjuku-ku, Tokyo 160-0004, Japan. Tel.: +81 3 3359 3195; fax: +81 3 3225 4376.
E-mail address: haginoa@mochida.co.jp (A. Hagino).

0197-2456/$ - see front matter
doi:10.1016/j.cct.2004.08.004

dynamic allocation, such as minimization, was used. In contrast, adjusted tests for allocation factors as covariates improved type I errors closer to the nominal significance level and they provided slightly higher power. In conclusion, both the statistical and clinical validity of minimization was demonstrated in our study.

Keywords: Minimization method; Random allocation; Trial design; Prognostic factor; Simulation study

1. Introduction

To conduct clinical trials ethically and scientifically, various issues must be considered at the time of protocol planning. Above all, one of the most important elements of the design is that the methods of treatment allocation and allocation factors should be utilized as covariates in the analysis plan.

In clinical trials, random allocation is usually conducted to compare the efficacy and safety between (or among) treatment groups. Random allocation has three important implications, namely, "elimination of selection bias between groups", "assurance of blinding" and "justification of randomization based tests". To achieve these goals, various allocation methods are currently applied. In circumstances when important prognostic factors exist, stratified randomization and minimization are the methods of choice to achieve balance in these factors between treatment groups [1]. Such allocation methods are very important to ensure balance in important prognostic factors for smaller scale trials or at the time of interim data monitoring for larger trials. Thus, during the planning stages in clinical trials, when deciding on allocation methods, it is useful to consider the following questions: Are there important prognostic factors? How many important prognostic factors? How large is the sample size of the trial? What type of allocation method was used in similar trials?

Minimization [2,3] can be classified as a dynamic allocation method as the allocation depends on prognostic factors of patients already recruited. Important prognostic factors are identified before the trial starts, and the assignment of a new patient to a treatment group is determined to minimize the differences between the groups in terms of these factors. Minimization differs from stratified randomization from the viewpoint of minimizing the total imbalance of all factors together instead of considering balance in each stratum.

Stratified randomization and minimization are allocation methods that aggressively achieve balance in the important prognostic factors between treatment groups, but these allocation methods have disadvantages. Stratified randomization has restrictions in the number of prognostic factors, while balance in the joint distribution is expected to be achieved among any combinations of levels of the selected factors. Minimization has few restrictions in the number of prognostic factors. However, this method ensures balance in marginal distributions and, thus, does not ensure balance among the combinations of the levels of factors. In addition, deterministic allocation is not desirable from the standpoint of predictability and principle of randomness [4]; therefore, the selection of appropriate allocation probability should be considered. To deal with these problems, allocation probabilities set to 3/4 or 2/3 have been recommended [5,6]. Other special features and potential problems with these allocation methods have been discussed in the comprehensive reviews [7,8].

Recently, the Committee for Proprietary Medicinal Products (CPMP) of the European Agency for the Evaluation of Medicinal Products (EMEA) has issued "Points to consider on adjustment for baseline covariates", which states "even if deterministic schemes are avoided, such methods remain highly

controversial and strongly advised to avoid such methods, and if they are used, the reasons should be justified on solid clinical and statistical grounds". It also noted that "dynamic allocation is strongly discouraged, and if used it is imperative that all factors used in the allocation scheme be included as covariates in the analysis" [9]. This indicates widespread concerns regarding the use of dynamic allocation, while at the same time, indicating the requirement of statistical and clinical rationale of minimization, if it is used.

We evaluated performance of minimization, focusing on relatively small-scale cancer clinical trials (two-arm comparative trials), as an example, by computer simulation. In the simulations, the balance in prognostic factors, type I error and power of statistical tests of several allocation methods were evaluated.

2. Data sources

As a hypothetical population, the data of 2158 patients with rectal cancer in the surgery-alone group who had been enrolled in seven randomized colorectal cancer trials conducted in Japan was selected.

The following six variables, sex, age, Dukes classification, histological depth of tumor invasion, lymph node metastasis and histological stage, were important prognostic factors for rectal cancer. Histological depth of tumor invasion and lymph node metastasis were based on criteria specified in "General Rules for Clinical and Pathological Studies on Cancer of Colon, Rectum and Anus" [10]. On the other hand, Dukes classification and histological stage were reclassified by combining the two factors, histological depth of tumor invasion and lymph node metastasis [10,11]. Therefore, independent prognostic factors essentially consisted of four factors: sex, age, histological depth of tumor invasion and lymph node metastasis. Table 1 shows the distribution of these prognostic factors in 2158 patients.

The rate of overall survival at 5 years was 64.9% in all 2158 patients [672 patients (31.1%) died and 85 patients (3.9%) were censored in follow-up period].

The degree of association between these prognostic factors and the endpoint (overall survival) was evaluated by Cox regression. Table 2 shows the result of Cox regression.

Table 1
Distribution of prognostic factors in 2158 patients

Prognostic factors		Number of patients (n=2158)
Sex	Male	1304
	Female	854
Age	≤49	376
	50–59	687
	60–69	783
	≥70	312
Histological depth of tumor invasion	≤pm	556
	ss/a1	758
	≥s/a2	844
Lymph node metastasis	n (−)	1242
	n1 (+)	528
	≥n2 (+)	388

Histological depth of tumor: ≤pm: m, sm and pm; ≥s/a2: s/a2 and si/ai. Lymph node metastasis: ≥n2 (+): n2 (+), n3 (+) and n4 (+).

Table 2
Influence of prognostic factors on overall survival

Prognostic factors		Number of patients	Influence of prognostic factors	
		n=2158	HR (p-value)[a]	HR (p-value)[b]
Sex	Male (%)	1304 (60.4%)	–	–
	Female (%)	854 (39.6%)	0.896 (p=0.1666)	0.866 (p=0.0713)
Age	≤49 (%)	376 (17.4%)	–	–
	50–59 (%)	687 (31.8%)	1.163 (p=0.1996)	1.241 (p=0.0673)
	60–69 (%)	783 (36.3%)	1.112 (p=0.3614)	1.176 (p=0.1626)
	≥70 (%)	312 (14.5%)	1.295 (p=0.0615)	1.418 (p=0.0117)
Histological depth of tumor invasion	≤pm (%)	556 (25.8%)	–	–
	ss/a1 (%)	758 (35.1%)	1.763 (p<0.0001)	1.434 (p=0.0037)
	≥s/a2 (%)	844 (39.1%)	2.873 (p<0.0001)	2.089 (p<0.0001)
Lymph node metastasis	n (−) (%)	1242 (57.6%)	–	–
	n1 (+) (%)	528 (24.5%)	2.159 (p<0.0001)	1.953 (p<0.0001)
	≥n2 (+) (%)	388 (18.0%)	3.632 (p<0.0001)	3.171 (p<0.0001)

[a] Crude hazard ratio and p-value using Cox regression.
[b] Adjusted hazard ratio by other three factors and p-value using Cox regression.

The results suggested that, as consistent with common knowledge in this field [12,13], histological depth of tumor invasion and lymph node metastasis had a particularly strong influence on overall survival.

3. Simulation methods

From the 2158 patients of the hypothetical population, 10,000 data sets of 50, 100, 150 and 200 patients (combining two groups) were repeatedly sampled with replacement to provide a data set for simulation. The sampling was conducted using the SURVEYSELECT procedure [14]. Allocation into two treatment groups, active (A) or placebo (P), was conducted repeatedly 10,000 times using three types of allocation methods: simple randomization, stratified randomization and minimization.

Simple randomization (SR) was conducted using pseudo Bernoulli random numbers. Two types of stratified randomization were performed: one is stratified randomization with four factors (STR4), sex, age, histological depth of tumor invasion and lymph node metastasis, and the other stratified randomization with the later two factors (STR2). In both cases, to ensure balance in the number of patients between groups within strata, the block size was set to four (an example of block AAPP). The influence of allocation probability and number of allocation factors on performance of minimization was evaluated, as well as balance in the simultaneous distribution of prognostic factors.

From the 10,000 sets derived for each trial size using the three types of allocation methods, 1000 sets were used to compare the differences in the number of patients between groups, balance in prognostic factors between groups and balance in the simultaneous distribution. The absolute value of the difference of the patient number between groups was the indicator for imbalance, and its 50 and 99 percentiles in the derived 1000 sets were evaluated, while the p-value of the chi-square test for the contingency table for each prognostic factor and the groups was calculated, and its 50 and 1 percentiles were used to compare the degree of balance among the three allocation methods. Interactions between prognostic

factors often arise in practical situations. Based on the features of allocation methods, stratified randomization is an allocation method that achieves balance in the simultaneous distribution of multiple prognostic factors, while minimization is an allocation method that achieves balance in the marginal distribution of each prognostic factor and does not ensure balance in the simultaneous distribution. Therefore, it has been suggested that if interactions exist among prognostic factors, stratified randomization is preferred over minimization [15]. Therefore, the p-value of the chi-square test of the contingency table formed by simultaneous distribution of multiple prognostic factors and group was calculated to evaluate balance in the simultaneous distribution.

Finally, the entire 10,000 sets of simulation data were used to evaluate the performance of statistical tests. Several statistical tests for actual overall survival time were conducted for evaluating the size of type I error after allocation, and the proportion achieving the chi-square test statistics greater than 3.841 (upper 5% point of chi-square distribution with df=1) was calculated to determine whether the nominal significance level (5%) was maintained. The applied tests were the log-rank test, the stratified log-rank test and the hazard ratio test using Cox regression. To compare the statistical power, based on an accelerated model in which the survival time increases or decreases due to the effect of treatment, overall survival is prolonged to 1.6- or 2.0-fold (censored at 5 years if it is longer than 5 years) for patients allocated to group A. In other words, if a patient died after 2 years from randomization, it was presumed that the patient died after 3.2 years ($2\times1.6=3.2$), and if a patient died after 4 years, it was presumed that the patient was censored at 5 years because it is longer than 5 years ($4\times1.6=6.4$). In special cases, that is, exponential and Weibull distributions, the accelerated model is equivalent to the proportional hazard model, and the hazard ratios are multiplied 1/1.6- or 1/2.0-fold, respectively.

4. Results

4.1. Imbalance in the number of patients between groups

The degree of imbalance in the number of patients between groups was compared among the three allocation methods. The allocation probability of minimization was changed from 1.00 (deterministic allocation) to 0.70 by 0.05. Table 3 shows the degree of imbalance in the number of patients between groups of 50-patient trials based on 1000 simulations.

Table 3
Degree of imbalance in the number of patients between groups of 50-patient trials based on 1000 simulations

Summary statistics	Absolute value of the difference in the number of patients between two groups									
	MIN (allocation probability)							STR2	STR4	SR
	1	0.95	0.90	0.85	0.80	0.75	0.70			
99 Percentile	2.0	2.0	2.0	4.0	4.0	4.0	4.0	10.0	16.0	18.0
50 percentile	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	4.0	4.0
Mean	0.5	0.6	0.7	0.8	0.9	1.0	1.1	2.9	4.9	5.6

Figures in the table show the summary statistics of the absolute value of the difference of patients between groups for each allocation method in 1000 simulations (N=50).
MIN: minimization included all four factors; STR2: stratified randomization included two factors; STR4: stratified randomization included all four factors; SR: simple randomization.

Using the SR, the 50 percentile was 4 (in actual case A=27 and P=23), and the 99 percentile (in the worst scenario) was 18 (A=34 and P=16).

The 50 percentile was 2, and the 99 percentile was 10 using the STR2. On the other hand, the 50 and 99 percentiles of STR4 were 4 and 16, respectively. In this case, the number of strata 72 (2×4×3×3) was relatively large, compared with trial size 50, therefore, a decrease in the number of patients per stratum resulted in large differences in the number of patients between groups.

Decreasing the allocation probability of minimization provided a slightly worse balance in the number of patients; however, even with the allocation probability of 0.70, balance still remained good, contrasting to the STR2, the STR4 and the SR. The difference in the number of patients between groups using minimization with the allocation probability of 0.70 was 0 (50 percentile) and 4 (99 percentile).

Although similar tendencies were confirmed with trial sizes of 100, 150 and 200 patients, the degree of imbalance of the SR, the STR2 and the STR4 was improved. The 50 and the 99 percentiles are shown in Table 4.

4.2. Balance in prognostic factors

Table 5 shows the degree of imbalance in prognostic factors in the case of 50-patient trials based on 1000 simulations.

Using the SR, under the null hypothesis, the p-value is expected to be uniformly distributed between 0 and 1; therefore, results of the simulation showed that the 50 percentile of p-values for all prognostic factors was around 0.50, and similarly, the 1 percentile of p-values was about 0.01.

It was confirmed that smaller allocation probability in minimization offered greater imbalance in prognostic factors.

Balance in the allocation factors achieved with the STR2 was similar to that achieved with minimization at the allocation probability of 0.70; however, balance in nonstratified factors such as sex and age were comparable with that obtained from the SR. In contrast, the STR4 had nonignorable imbalance, due to too many strata compared with a given trial size. It is difficult to apply stratified randomization to achieve good balance in such small-scale trials.

When selecting the allocation probability in minimization, if the 1 percentile of p-value for the chi-square test requires about 0.50 as a criterion to achieve a strongly acceptable degree of

Table 4
Degree of imbalance in the number of patients between groups of 100- to 200-patient trials based on 1000 simulations

Trial size	Absolute value of the difference in the number of patients between two groups							
	MIN0.7		STR2		STR4		SR	
	50 Percentile	99 Percentile	50 Percentile	99 Percentile	50 Percentile	99 Percentile	50 Percentile	99 Percentile
100	2	4	2	10	6	20	6	26
150	2	4	2	10	6	23	8	30
200	0	4	2	10	6	24	10	35

Figures in table show 50 and 99 percentiles of the absolute value of the difference of patients between groups for each allocation method, in 1000 simulations (N=100, 150 and 200, respectively).

MIN0.7: minimization included all four factors (allocation probability=0.70); STR2: stratified randomization included two factors; STR4: stratified randomization included all four factors; SR: simple randomization.

Table 5
Balance in the marginal distribution of prognostic factors of 50-patient trials based on 1000 simulations

Prognostic factor	MIN (allocation probability)							STR2	STR4	SR
	1	0.95	0.90	0.85	0.80	0.75	0.70			
50 Percentile of p-value of chi-square test for the contingency table of each factor and group										
Sex	0.7815	0.7766	0.7773	0.7766	0.7766	0.7745	0.7708	0.4741	0.5557	0.5218
Age	0.9438	0.9381	0.9297	0.9225	0.8975	0.8732	0.8397	0.4919	0.5745	0.4919
Histological depth of tumor invasion	0.9344	0.9331	0.9279	0.9256	0.9098	0.8629	0.8387	0.8151	0.5909	0.4948
Lymph node metastasis	0.9195	0.9162	0.9082	0.8905	0.8559	0.8437	0.8226	0.7702	0.5331	0.4723
1 Percentile of p-value of chi-square test for the contingency table of each factor and group										
Sex	0.5443	0.5107	0.3954	0.3705	0.3737	0.2575	0.2410	0.0138	0.0225	0.0121
Age	0.5086	0.5019	0.4725	0.3728	0.2917	0.2647	0.2309	0.0140	0.0265	0.0139
Histological depth of tumor invasion	0.6125	0.5831	0.5242	0.4897	0.4415	0.3236	0.2971	0.2789	0.0267	0.0109
Lymph node metastasis	0.5318	0.4867	0.5023	0.3621	0.2903	0.2275	0.2525	0.1405	0.0275	0.0141

Figures in table show 50 and 1 percentiles of p-value of chi-square test from 1000 simulation data sets (N=50) for each allocation method.
MIN: minimization included all four factors; STR2: stratified randomization included two factors; STR4: stratified randomization included all four factors; SR: simple randomization.

balance, even in the possible worst case, the allocation probability should be set to 0.95 in 50-patient trials.

Similar balance ranking was observed among the allocation methods in trials with 100, 150 and 200 patients; however, the degree of imbalance of the STR2 and the STR4 improved as trial size increased. Based on the above criterion 1 percentile of p-value is about 0.50, the allocation probability required for a given trial size would be 0.80 for 100, 0.75 for 150, and 0.70 for 200 patients. If the number of patients increases, it is possible to use a smaller allocation probability to avoid predictability, while keeping a good balance.

4.3. Balance in simultaneous distribution

The difference of 3×3 simultaneous distribution (histological depth of tumor invasion and lymph node metastasis) between groups were examined using the p-value of the chi-square test as an indicator, and the 50 and the 1 percentiles of p-value were calculated. Table 6 shows the results of 200-patient trials based on 1000 simulations.

As expected, the STR2 achieved the best comparability in the combination (nine levels) of histological depth of tumor invasion and lymph node metastasis. Minimization did not provide a good balance in simultaneous distribution, even when increasing allocation probability or the trial size.

Table 7 shows a typical pattern in which minimization did not work well; that is, marginal distribution was comparable between groups. However, nonignorable imbalance was observed in the simultaneous distribution. On the other hand, stratified randomization guarantees a good balance in the simultaneous distribution.

Fig. 1 shows the hazard ratios for each level in 2158 patients of the hypothetical population.

Histological depth of tumor invasion correlated with lymph node metastasis, but there was no strong interaction in the overall survival (Wald test, chi-square=3.086, df=4). If interaction cannot be ignored, it is important to ensure balance in the simultaneous distribution. To achieve balance in the simultaneous distribution similar to stratified randomization, minimization should be applied by combining these two factors into one allocation factor with nine levels.

Table 6
Balance in simultaneous distribution of 200-patient trials based on 1000 simulations

Simultaneous distribution[a]	p-Value of chi-square test for the contingency table of simultaneous distribution and groups									
	MIN (allocation probability)							STR2	STR4	SR
	1	0.95	0.90	0.85	0.80	0.75	0.70			
50 Percentile	0.8882	0.8748	0.8869	0.8909	0.8873	0.8677	0.8642	0.9973	0.8665	0.4741
1 Percentile	0.1497	0.0837	0.1048	0.1192	0.1127	0.1123	0.1072	0.8569	0.2331	0.0081

Figures in table show 50 and 1 percentiles of p-value of chi-square test from 1000 simulation data sets (N=200) for each allocation method.

MIN: minimization included all four factors; STR2: stratified randomization included two factors; STR4: stratified randomization included all four factors; SR: simple randomization.

[a] Combinations (nine levels) of the depth of tumor invasion and the lymph node metastasis.

Table 7
Balance between groups in the combination of levels of a 200-patient trial

		Lymph node metastasis			Group A	Group P
		n (−)	n1 (+)	≥n2 (+)		
STR2						
Histological depth of	≤pm	44 [A: 23, P: 21]	2 [A: 1, P: 1]	10 [A: 6, P: 4]	29	27
tumor invasion	ss/a1	34 [A: 17, P: 17]	21 [A: 10, P: 11]	8 [A: 4, P: 4]	31	32
	≥s/a2	45 [A: 22, P: 23]	18 [A: 9, P: 9]	18 [A: 9, P: 9]	40	41
Group A		61	20	19	100	
Group P		62	21	17		100
Min0.70						
Histological depth of	≤pm	44 [A: 23, P: 21]	2 [A: 2, P: 0]	10 [A: 2, P: 8]	27	29
tumor invasion	ss/a1	34 [A: 12, P: 22]	21 [A: 13, P: 8]	8 [A: 7, P: 1]	32	31
	≥s/a2	45 [A: 26, P: 19]	18 [A: 6, P: 12]	18 [A: 8, P: 10]	40	41
Group A		61	21	17	99	
Group P		62	20	19		101

Figures in table show the example of result of allocation from one of 1000 simulations (N=200) for each allocation method for STR2 and MIN0.7. In brackets is the number of allocated patients in each group.
MIN0.7: minimization included all four factors (allocation probability=0.70); STR2: stratified randomization included two factors.

4.4. Type I error and power

Stratified log-rank test and Cox regression were conducted adjusting for two factors: histological depth of tumor invasion and lymph node metastasis. Table 8 shows the results of 200-patient trials based on 10,000 simulations.

When stratified randomization or minimization was used, it was apparent that the result of unadjusted test for allocation factors (log-rank test) turned out to be conservative. In contrast, analysis with adjustment for the allocation factors as covariates such as stratified log-rank test and Cox regression, provided type I error close to the nominal significance level and, as a result, improved the statistical power. Parallel affinities were noticed with trial sizes of 50, 100 and 150 patients.

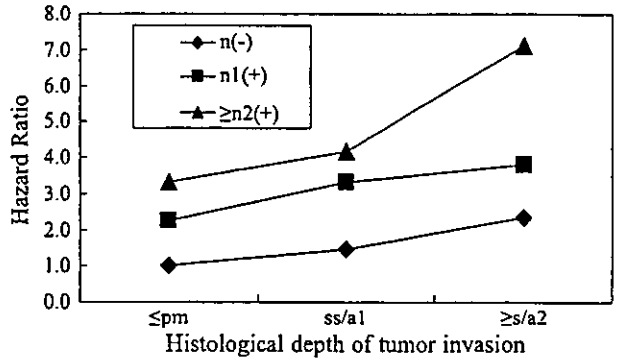


Fig. 1. Hazard ratio (n (−), pm or less as 1) for each level of combination (nine levels) of histological depth of tumor invasion and lymph node metastasis in 2158 patients. Histological depth of tumor: ≤pm: m, sm and pm; ≥s/a2: s/a2 and si/ai. Lymph node metastasis: ≥n2 (+): n2 (+), n3 (+), and n4 (+).

Table 8
Type I error and power of 200-patient trials based on 10,000 simulations

Active group (×survival time)		Test	Allocation methods		
			MIN (detaministic)	STR2	SR
H_0	(×1.0)	Log-rank	0.0352	0.0334	0.0505
		Stratified log-rank	0.0466	0.0446	0.0504
		Cox regression	0.0494	0.0475	0.0548
H_1	(×1.6)	Log-rank	0.4155	0.4265	0.4302
		Stratified log-rank	0.4790	0.4904	0.4778
		Cox regression	0.4799	0.4853	0.4812
	(×2.0)	Log-rank	0.7793	0.7852	0.7610
		Stratified log-rank	0.8215	0.8303	0.8054
		Cox regression	0.8220	0.8271	0.8118

The table shows the actual type I error at a nominal significance level of 0.05. The table also shows the power (H_1) attained with the overall survival prolonged for allocated to group A. Both cases in the 10,000 simulations (N=200).
MIN: minimization included two factors (allocation probability=1); STR2: stratified randomization included two factors; SR: simple randomization.

It is evident that these results could be predicted qualitatively based on the feature of the analysis methods; however, this study confirmed the extent of the difference among allocation methods in a quantitative manner, based on actual clinical trial data. The simulation also revealed that type I error of the minimization could be sufficiently maintained both with and without adjustment analysis. This suggested the statistical validity of minimization.

5. Discussion

We conducted simulations using the actual data from clinical trials of rectal cancer and they provided the following results:

(1) When four allocation factors exist, stratified randomization does not perform well in small-scale trials (about 50 patients). However, even in such cases, minimization can achieve sound balance in the number of patients and in the distribution of prognostic factors between groups. It can be concluded that the results indicate the usefulness of minimization, which can achieve balance even in smaller scale trials.
(2) Minimization can ensure comparability between groups even using smaller allocation probability (instead of deterministic allocation) to prevent predictability by increasing the sample size.
(3) Minimization can ensure balance in the marginal distribution of prognostic factors but does not ensure balance in the simultaneous distribution. Therefore, minimization can be difficult if interactions between prognostic factors can be predicted. In such cases, stratified randomization or minimization in which levels are reclassified based on simultaneous distribution of multiple prognostic factors should be applied.
(4) Simple unadjusted tests have conservative type I errors when minimization is conducted. On the other hand, because adjusted tests for allocation factors as covariates can achieve an approximate nominal significance level, they have the elevated power by up to 5–6%.

Three main criteria were used to assess performance of each allocation method:

(1) balance in the number of patients between treatment arms,
(2) balance in the distribution of prognostic factors,
(3) performance of statistical tests.

The reason why balance in the number of patients allocated to each treatment arm is desirable in a randomized trial is that, for a fixed number of patients, statistical power is maximized and the width of confidence interval is minimized when an equal number of patients is allocated to each arm, although moderate imbalances produce negligible loss of statistical performance.

As confirmed in our study, if known prognostic factors are balanced in the allocation process by means of stratification or minimization, it is obvious that these prognostic factors are more evenly distributed than with simple randomization, especially in small trials. Assuring a balance in the distribution of prognostic factors provides an unbiased estimation of hazard ratio and the rationale for the use of simple statistical methods without adjusting for prognostic factors.

When marked imbalances are found, these can be adjusted in the analysis; however, a variety of possible models are available, according to the difference of mathematical formulation and combination of prognostic factors. Moreover, the validity of adjusted analysis depends on the correctness of model assumptions that cannot be confirmed. Therefore, it is much better to balance major prognostic factors at the design stage and to apply an unadjusted simple statistical method, although, even in the balanced case, adjusted analysis gives greater power.

In this study, small-scale clinical trials were evaluated. The total number of patients was between 50 and 200 in the two groups. In actual comparative cancer clinical trials, the trial size is often larger than these sizes, especially in the adjuvant setting, cancer trial usually imply the enrollment of many hundreds (and sometimes more than thousand of patients) when endpoint is survival. However, if the trial size is large enough, simple randomization works sufficiently to obtain a good balance, and special allocation techniques are not required. And cancer trials assessing biological treatments potentially associated with dramatic effects on advanced disease or using the other endpoint, such as event-free survival, QOL and response rate, can require much fewer patients. In addition, it is usually required to conduct interim analysis in long-term and large-scale cancer clinical trials. Interim analysis is often conducted at the time when a third or a half of planned total event is accumulated. Therefore, sample size at the interim analysis is much smaller than that at the final analysis. Moreover, it is very important that the decision-making process is based on only minimum information to avoid any possible biases when the results of interim analysis are leaked. In addition, decision making based on interim analysis must be conducted in a very limited time frame.

Therefore, it is preferable to achieve strict balance in the number of patients and prognostic factors between groups at the time of the interim analysis to avoid complex analysis, such as adjusted analysis. Even in large-scale trials, it is important to ensure balance at the time of interim analysis.

Although four allocation factors were considered in our study, in some clinical trials, more allocation factors must be considered. When more allocation factors exist, other investigators have evaluated the performance of minimization in trial sizes of 1000 patients and prognostic factors involving 12 variables, and they demonstrated that minimization achieved balance and maintained nominal significance level

[16]. The study used actual stroke-patient data; however, the sampling method from the population was different from our study.

The covariate, age, was dealt as a categorical variable after categorizing into four levels in minimization of our study. However, age is essentially measured in a continuous way. Extensions of minimization to balance of the means and standard deviations of continuous prognostic factors between groups have also been proposed [17,18]. If a prognostic factor is a continuous variable and clinically appropriate categorization is difficult, the minimization which provides balance of the means and standard deviations are the method of choice.

Finally, the problem of dynamic allocation, such as minimization, as indicated by CPMP is discussed, based on our simulation results.

Predictability is one of the most important issues in the conduct of clinical trials. However, even when using reducing allocation probability, it is possible to achieve good balance between groups in a moderate-scale trial; therefore, this does not directly obstruct the application of minimization. It is indicated that balance in prognostic factors and the number of patients can be achieved in small clinical trials and that the significance level can be maintained, thus, the clinical and statistical justification of minimization was demonstrated. Indication by CPMP that "when dynamic allocation is used, the allocation factors should be considered in the analysis" was confirmed to be appropriate. The results of the simulation revealed that adjustments for the allocation factors can bring closer to the nominal significance level, with a pay-off being an improvement in the power by about 5%. Thus, when using minimization, it is necessary to specify the adjusted method in the statistical analysis plan.

In this investigation, where we used the data from rectal cancer patients in clinical trials of other fields, there are often multiple prognostic factors that reflect the severity of disease, corresponding to the histological depth of tumor invasion and lymph node metastasis investigated here, besides basic demographic variables such as sex and age. In conclusion, this investigation demonstrated that in small-scale clinical trials where multiple prognostic factors exist, minimization is a useful method to achieve balance in prognostic factors.

Acknowledgments

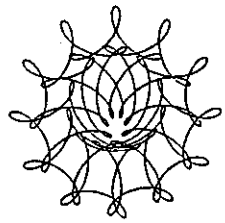
We are grateful to Wataru Kashiwagi (Taiho Pharmaceutical) and two anonymous referees for their useful comments. This research was (partially) supported by JSPS Grant-in-Aid for Scientific Research No. 16200022.

References

[1] Rosenberger WF, Lachin JM. Randomization in clinical trials, theory and practice. New York: John Wiley & Sons; 2002.
[2] Taves DR. Minimization: a new methods of assigning patients to treatment and control groups. Clin Pharmacol Ther 1974;15:443–53.
[3] Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 1975;31:103–15.
[4] ICH E9 Expert Working Group. Statistical principles for clinical trials. Stat Med 1999;18:1905–42.
[5] Japan Pharmaceutical Manufacturers Association. Drug allocation in clinical trials. Jpn Pharmacol Ther 1991;19: 2995–3013 [in Japanese].

[6] Pocock SJ. Clinical trials: a practical approach. New York: John Wiley & Sons; 1984.
[7] Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials: a review. Control Clin Trials 2002;23:662–74.
[8] McEntegart DJ. The pursuit of balance using stratified and dynamic randomization techniques: an overview. Drug Inf J 2003;37:293–308.
[9] The European Agency for the Evaluation of Medicinal Products. Points to consider on adjustment for baseline covariates. CPMP/EWP/2863/99 London, 2003. ⟨⟨http://www.emea.eu.int/htms/human/ewp/ewpptc.htm⟩⟩. Accessed December 11, 2003.
[10] Japanese Society for Cancer of the Colon and Rectum. General rules for clinical and pathological studies on cancer of the colon, rectum and anus. Sixth edition. Tokyo, Japan: Kanehara and Co.; 1998. In Japanese.
[11] Dukes CE. The classification of cancer of the rectum. J Pathol 1932;35:323–32.
[12] Minsky BD, Mies C, Recht A, Rich TA, Chaffery JT. Resectable adenocarcinoma of the rectosigmoid and rectum: 1. Patterns of failure and survival. Cancer 1988;61:1408–16.
[13] Minsky BD, Mies C, Recht A, Chaffery JT. Potentially curative surgery of the colon cancer: 1. Patterns of failure and survival. J Clin Oncol 1988;6:106–18.
[14] SAS Institute. SAS/STAT User's Guide, Version 8.
[15] Tu D, Shalay K, Pater J. Adjustment of treatment effect for covariates in clinical trials: statistical and regulatory issues. Drug Inf J 2000;34:511–23.
[16] Weir CJ, Lees KR. Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial. Stat Med 2003;22:705–26.
[17] Nishi T. Allocation to achieve balance in the means between treatment groups. Proceedings of the Biometric Society of Japan symposium; 2002. p. 43–8. In Japanese.
[18] Takaichi A, Nishi T. The allocation of patients using the extended minimization method. Proceedings of the 21th annual SAS Users Group International Japan conference; 2002. p. 467–71. In Japanese.

臨床統計学
臨床試験を中心として



大橋 靖雄

臨床統計学（Clinical Biostatistics）は，健康・医療に関わる応用統計学 Biostatistics（生物統計学と訳される）の重要な1分野である．わが国でも生物統計学の重要性がようやく認識され，実務家・研究者の需要が立ち上がるとともに教育システムも生まれつつある．大学では1992年に東京大学に初めての講座が誕生し，1999年の北里大学薬学系研究科，2000年の京都大学医学系研究科と講座開設が続いた．2002年には東京理科大学大学院（経営工学）に社会人対象修士コースが生まれ，2004年には久留米大学，厚生労働省保健科学院にも教育コースが開設予定である．現時点で生物統計家の参画が最も必要とされているのは，臨床研究とくに臨床試験の計画と解析である．本稿では，歴史的な展開も含め，詳細な数理というよりは，このような専門家が必要となってきた背景の記述と主要な概念について述べてみたい．

1. 臨床医学と EBM，臨床試験

臨床医学の目的は，患者の疾患を正確に診断し適切に治療を行うことにある．治療目的は，可能な場合には治癒を目指し，治癒が不可能な場合には患者の QOL を可能な限り向上させるか維持することである．しかし，医療提供者側の知識・技術の不完全さ，主に患者・疾患の多様さに多く由来する治療効果の不確実さから，診断結果や治療結果には永遠に除去不可能な曖昧さが伴う．したがって，診断・治療によってもたらされるベネフィットと被るリスクは，ともに「可能性」として確率変数的性格を帯びる．これらに対する患者の重み付けも患者個々の価値判断を反映して異なるはずのものであり，診断・治療法の選択は，リスク・ベネフィット両者のバランスと資源の制約の中で，本来は充分な情報提示と理解，そして自発的意志を前提としたインフォームドコンセントによるべきである（とされる）．しかし，医療提供者と患者の有する情報の不均衡，意志・患者双方の意識の問題もあり，これまでの治療上の意志決定は，医師主導のパターナリズムの中で，曖昧な状況下で行われてきたといってよい．近年の情報公開あるいは患者の権利主張の流れは，このような意思決定プロセスに大きな変革を与えようとしている．厚生労働省は，ここ数年，疾患毎に標準治療をまとめたガイドライン策定を各関連学会に依頼し，数多くの疾患ガイドラインやその案が発表されてきた．これは，国民皆保険のもと，治療結果をマスとして評価するシステムが存在しないまま，医師の自由裁量のもと出来高払いで治療が行われてきた反省と，また上記の情報開示の要求に応えるためでもあった．そして，このガイドライン策定過程で（医療関係者には周知であったが）改めて次の事態が浮き彫りになった．

「わが国には客観的証拠（evidence）が無い！」

臨床医学系学会では，最近5年ほど Evidence Based Medicine（略して EBM）という言葉が大流行である．EBM とは，目の前の患者の問題点を一定の手順で定型化し，主に文献検索と抽出された文献の批判的吟味により過去の「証拠・根拠」を点検し，そこから有効な情報を引き出し，目の前の患者に対して，その特異性を充分考慮しつつ，ときにはこれまでの経験によって修正も加え実践することである．科学と経験の融合である．ここで，evidence を提供するのが患者を対象とした臨床研究成果である．

臨床研究は，患者の診断・治療経過をまとめた症例報告，多数の患者の治療実態下での観察研究，そして臨床試験に大まかに分けられる．薬・手術・放射線などの治

療（予防）手段を研究者側が前向きに制御して行われる実験的研究が臨床試験である．これらの手段をまとめて介入（intervention）と呼ぶ．複数の臨床試験を統計的に併合する「メタアナリシス（meta-analysis）」も，もとの臨床試験の質が高ければ，高い質の evidence を提供し，ガイドライン策定に大いに利用されている．

臨床試験はヒトを対象とした実験ではあるが，ナチスあるいはわが国の石井部隊（731 部隊）が戦時中に行ってきた人体実験とは，倫理性の確保という点で大いに異なる．このための国際合意の実施基準が後述の GCP（Good Clinical Practice）である．

さて，治療法の開発とくに新薬開発に関して一般国民が抱いている最大の誤解は，画期的な新薬が動物実験や試験管内の基礎研究から生れる，という「幻想」である．すべての新薬の認可あるいは既存薬剤の適応拡大には，適切に計画・管理実施され，通常は適切な比較対照を有する（これを adequately well-controlled と呼ぶ）臨床試験が必須である（この事情は医療機器についても同様であるが，簡単のため以下は「薬」「製薬会社」で代表させる）．

GCP に従った臨床試験は，図 1 に示すように入念な準備と複雑な手順，多くの人々の参加・協力により実施される．事前にプロトコルと呼ばれる研究計画書が策定される．これは研究の意義（科学的側面と倫理的側面）と実施マニュアルの両面を持つ基本文書であり，この中で試験の目的・対象者・介入手段・統計解析を含む評価方法が規定される．このプロトコルに従って試験が進行するために，研究スポンサー（後述する「治験」では製薬会社）あるいは実施施設（病院）は手順書（Standard Operational Procedure; SOP）を定め，人的配置を行う．製薬会社や臨床試験の請負会社（Contract Research Organization; CRO）の社員が施設を訪問して実施状況と報告の正確さを点検する活動がモニタリングであり，さらに第三者の監査等を通じて，得られたデータの信頼性の品質保証がなされる．統計解析を担当する統計家は計画段階から参画し，解析計画書を策定し解析実務を行い，膨大な量となる報告書作成に協力する．

新薬申請の場合は，このような臨床試験を通じて当該薬剤が安全性の点で許容可能であり，さらにもちろん患者の生存・QOL やこれらと直結していることが証明されている評価変数（エンドポイントと呼ばれる）の統計的処理を通じて有効であることの証明がなされて初めてデータが審査当局（日本の場合は厚生労働省）に提出される．そして，1 年ほどの専門家による審査を経て，認可・発売となる．新薬 1 薬剤の開発費用はうなぎ上りであり，失敗例も含めれば 1 薬剤 100–200 億円以上，うち臨床試験の費用は数 10 億円以上となる．数千例を超えるような大規模試験の場合には数 100 億円の費用を要することもある．

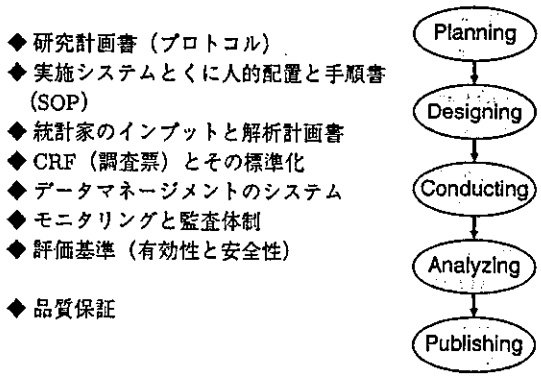


図 1　臨床試験に必要な準備．

2. わが国の evidence 不足

さて EBM に戻ろう．わが国の臨床医学は，evidence 作り，とくに臨床試験を通じた evidence に実はほとんど寄与していない．その理由は，以下のような，臨床試験実施のためのインフラストラクチャを欠いていたことにある．

(1) 医師研究者を支援するコーディネータ（Clinical Research Coordinator）
(2) 研究計画書（プロトコル）を書くことのできる医師研究者とそのための教育
(3) 生物統計家（試験統計家）あるいは生物統計学の教育
(4) 効率的なデータマネージメント・システムと中立的なデータセンター
(5) Peer review あるいは監査・査察によるデータの品質保証体制
(6) 医師研究者主導型の臨床研究に対する法的規制
(7) 結果を発信するプロフェッショナルである Medical writer の人材と教育
(8) 臨床試験の意義の理解と患者参加に対するインセンティブ