
総 説

臨床試験デザイン

手良向 聡*

京都府立医科大学大学院医学研究科生物統計学

Clinical Trial Design

Satoshi Teramukai

Department of Biostatistics,

Kyoto Prefectural University of Medicine Graduate School of Medical Science

抄 録

臨床試験はデザイン・計画の段階から始まり、試験実施・データ管理・モニタリングを経て、データ解析・報告書作成に至る。この各ステップが統計的方法を必要としている。再現性によって結果を保証することが可能な基礎実験と異なり、同じデザインで繰り返すことが困難な臨床試験においては、デザインと手続きの妥当性から結果を保証するしかない。計画段階から生物統計学の専門家が参画していれば、質の高い臨床試験を実施できる可能性は高い。近年、臨床試験にベイズ流の方法が有用であるという報告は着実に増えている。効率のかつ倫理的な試験デザインの開発は、資源を有効に活用するという観点から今後ますます重要になるであろう。本稿では、ランダム化対照試験の標準的方法、および探索的臨床試験のデザインとして有用であるベイズ流の方法について述べる。

キーワード：技術評価，統計的仮説，ランダム化，統計的考察，ベイズ流統計学。

Abstract

Every clinical trial starts from the design and planning stage, moves to trial conduct, data management and monitoring, and finally to the data analysis and conclusions. Each step along the way calls for statistical methods. While basic research which can be guaranteed a result by reproducibility, clinical trials which cannot be repeated with the same design must be guaranteed by the validity of design and procedure. If a biostatistician take part in the planning stage, we can do the high quality of clinical trials. In recent years, reports insisting on the usefulness of Bayesian statistics in clinical studies have steadily increased. The development of efficient and ethical design will become important in the future, from the viewpoint of the best use of resources. In this paper, I will describe the standard methods for randomized clinical trials, and Bayesian methods which could be useful for the design of exploratory clinical trials.

Key Words: Technology assessment, Statistical hypothesis, Randomization, Statistical consideration, Bayesian statistics.

平成26年 9月16日受付

*連絡先 手良向聡 〒602-8566 京都市上京区河原町通広小路上ル梶井町465

steramu@koto.kpu-m.ac.jp

はじめに

臨床試験はデザイン・計画の段階から始まり、試験実施・データ管理・モニタリングを経て、データ解析・報告書作成に至る。この各ステップが統計的方法を必要としている。臨床試験は、良いデザインおよび正しい遂行がなければ、台無し（例えば、結論の出ないまたは誤った知見を導く）になり、さらには大惨事（例えば、不必要に多数の患者が毒性または死を被ることを引き起こす）になる可能性さえある。臨床試験は効率的かつ倫理的であるべきで、資源を節約し、より多くの患者に恩恵を与え、より迅速に正しい結論を引き出し、結果として不必要な毒性をより少なくすべきである¹⁾。本稿では、ランダム化対照試験の標準的方法、および探索的臨床試験のデザインとして有用であるベイズ流（Bayesian）の方法について述べる。

医学・医療と技術評価

医学は普遍性のある真実を追求する科学の一分野である一方、医療は多様性のある個人に対して最適な技術を選択して適用することが要求される場である。技術評価は、主に統計学に基づく科学的方法を駆使して医療技術を相対的に評価し、医学から医療への橋渡しを行う（図1）。

統計学を医学・医療の領域に導入する際には、2つの大きなギャップを認識しておく必要がある。1つは、「決定論」と「非決定論（確率論）」のギャップである。1800年代半ばにクロード・ベルナールが「統計学に立脚している限り、医学は永久に推測科学に止まるであろう」と決定論的な考え方を主張して以来、医学の世界では決定論的な思想が支配的である。もう1つのギャップは、意思決定の主体に関わる問題であり、「対集団の確率」と「対個人の確率」とのギャップである。たとえば、ある医薬品を承認すべきかどうかという判断は、その国の人々という集団に対するベネフィットとリスクのバランスで決定される。その決定は「対集団の確率」に基づく一方、医療の場で診断や治療を行う際には、個人に対するベネフィットとリスクを評価しなければならない。たとえば、胎児診断を行って、医師が「胎児に異常がある確率は80%」と言ったとき、その80%は集団での頻度に基づいたものである。しかしながら、それを聞いた母親の「子供には異常があるか（100%）、ないか（0%）のどちらか」という感覚で、この確率を解釈することはそれほど容易ではない。これは「確率とは何か」という哲学的課題につながっている。

臨床試験は、20世紀を代表する英国の統計学

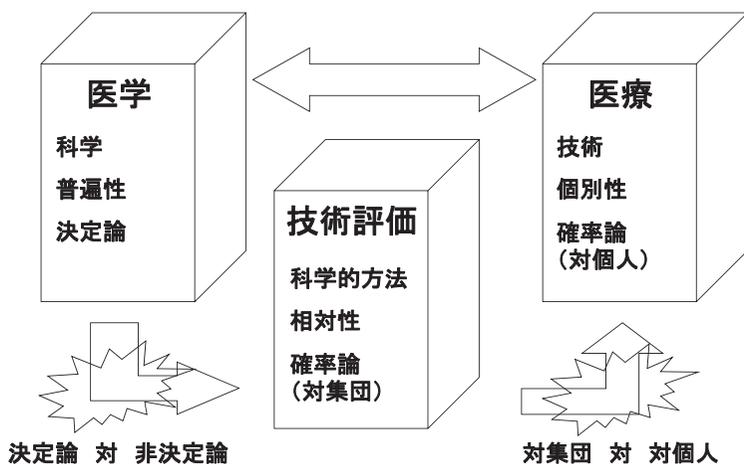


図1 医学・医療と技術評価

者 R.A. フィッシャー (1890~1962) が創始した統計的実験 (技術的実験とも呼ばれる) の方法論を基礎としている。科学的実験は、人工的に作り出された純粋な条件のもとでの因果関係を確定しようとするのに対して、統計的実験は以下の特徴を有する²⁾。

- ・実験の場は、現実の応用の場に近い状況に設定される
- ・結果の分析には誤差の存在を前提にしなければならない
- ・いくつかの因子を同時に変化させて結果を見る必要があることがある
- ・目的は、何らかの基準によって現実の場において最も良い結果が得られるような条件を求めることである

様々な種類の誤差を伴うデータを扱うためには統計的方法が不可欠である。また、再現性によって結果を保証することが可能な基礎研究と異なり、同じデザインで繰り返すことが困難な臨床試験においては、デザインと手続きの妥当性から結果を保証するしかない。

探索的試験と検証的試験

臨床試験の性格は検証的試験と探索的試験に大別される。検証的試験は有効性または安全性の確固たる証拠を提示するための試験と位置付けられる。ただし、個々のいかなる試験も検証的側面と探索的側面の両方を持つ。試験実施計画書 (以下、プロトコル) には、各試験について検証的な証明として用いられる側面と、探索的解析のためにデータを提供する側面とを、明確に区別しておくべきである³⁾。探索的解析から得られた結果は仮説に過ぎず、その仮説は検証的試験によって確認しなければならない。

統計的仮説

統計的仮説の代表的なものは、優越性仮説と非劣性仮説である。優越性仮説を証明しようとする試験 (優越性試験) とは、試験治療の効果が対照治療 (活性対照またはプラセボ対照) よりも「臨床的に優れること」を示すことが目的の試験である。一方、非劣性仮説を証明しよう

とする試験 (非劣性試験) とは、試験治療の効果が対照治療よりも「臨床的に劣らないこと」を示すことが目的の試験である。多くの場合は優越性仮説が設定されるが、対照治療 (通常、活性対照) に比べて安全性あるいは経済性に優れていることが見込まれる場合に、このような非劣性仮説が許容されることがある。

非劣性試験を計画する際には、非劣性マージン (臨床的に意味のある最小の差: Δ) の決定、データの質などについて十分な注意が必要である。ハザード比を治療効果の尺度とした臨床試験の場合、優越性試験では、ハザード比の 95% 信頼区間の上限が 1 より小さければ、有意水準 5% で試験治療が優れると判断される。一方、非劣性試験では、その 95% 信頼区間の上限が $1+\Delta$ よりも小さければ、有意水準 5% で試験治療は対照治療に Δ 以上は劣らない、と判断される (図 2)。

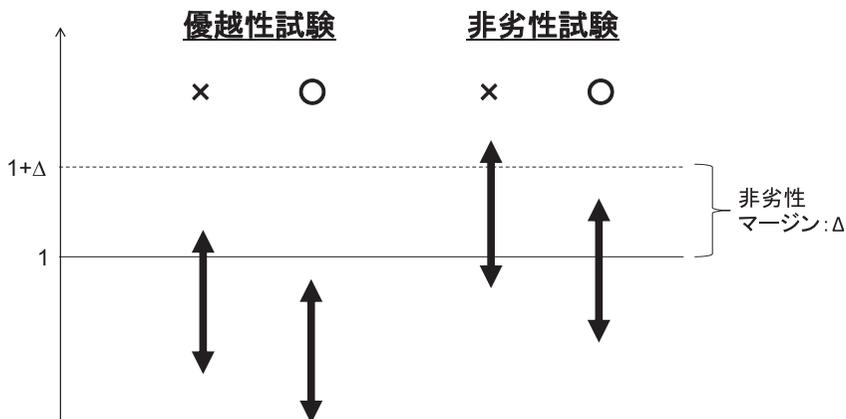
ランダム化

ランダム化対照試験では、登録の際にランダム化という操作が必要になる (図 3)。ランダム化を行うことにより、試験群と対照群の比較可能性 (内的妥当性とも呼ばれる) が保証される。ランダム化は「実験計画法」を確立したフィッシャーの偉大な発明の一つである。実験に伴う誤差には以下の 2 種類がある。ちなみに、フィッシャーは臨床試験ではなく、農事実験に従事していた。

- ・偶然誤差…測定誤差のようにある確率分布に従うと想定できる誤差であり、繰り返し測定を行えばその大きさについて推定可能である
- ・系統誤差またはバイアス…圃場の肥沃度や日当たりの不均一性のように確率変動と見なせない誤差であり、繰り返しには関係なく結果を歪める原因となる

ランダム化の目的は、一言で言うと「系統誤差を偶然誤差に転化すること」である。臨床試験におけるその意義は、

- ・予後因子が既知か未知かにかかわらず、予後因子の分布が類似したグループを作る



矢印は、ハザード比(試験薬のハザード率/対照薬のハザード率)の95%信頼区間

図2 優越性試験と非劣性試験の判断規準

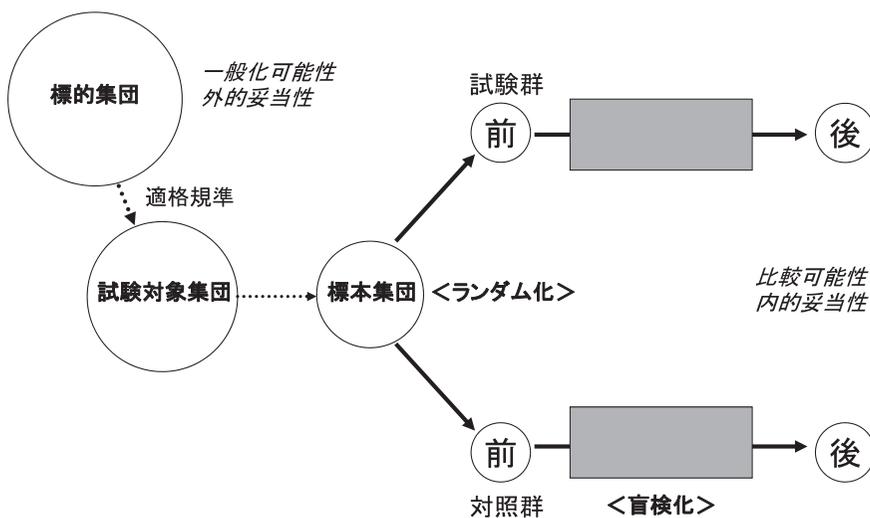


図3 ランダム化対照試験

・データ解析の際に治療効果の定量的な評価のための正しい統計的根拠を与えることである³⁾。

臨床試験で利用されるランダム化の方法は、固定的割付(単純ランダム化、並べ替えブロックランダム化など)と適応的割付(偏コインデザイン、最小化法など)とに分類される。重要な予後因子の分布を治療群間で均等にするためには、それらの因子で層を作り、その層ごとに

ランダム化(層別ランダム化)を行う必要がある。ランダム化の方法には多くの選択肢があり、試験統計家がそれぞれの利点と欠点を考慮して決定すべきである。

統計的考察

試験統計家がプロトコルに記載すべき統計的事項として、「標本サイズの設定根拠」、「解析対象集団」、「解析項目・方法」、「中間モニタリン

グ」がある。

1. 標本サイズの設定根拠

臨床試験に参加していただく対象数（標本サイズ）は科学性、倫理性、実施可能性のバランスを考慮して決める必要があり、多すぎても少なすぎてもいけない。標本サイズは主要評価項目に関する情報に基づいて計算される。その時点における情報を最大限利用するものの、一時的な仮定に基づく概算であることに注意が必要である。例えば、ある癌の補助化学療法の臨床試験において、標準治療を受ける患者の5年生存率を予測する場合、利用できる情報にはかなり大きなばらつきがある。これらの前提条件を慎重に検討した上で、通常は仮説検定に基づく決定方式に従って、帰無仮説、対立仮説、検定統計量、有意水準、第Ⅰ種の過誤、第Ⅱ種の過誤または検出力などが試験の性格（探索的または検証的）を考慮して決定される。

2. 解析対象集団

解析対象集団は、ITT (intention-to-treat) の原則に従って定義すべきである。これは、「被験者が実際に受けた治療ではなく、被験者を治療しようとした意図 (intention to treat) に基づいて評価する」という原則である³⁾。従って、登録されたすべての被験者を解析対象とすることが原則であるが、登録後に判明した不適格例、試験治療を全く受けなかった例を対象から除くことは一般的に許容される。いずれにしても、有効性および安全性に関する主要な解析対象集団の定義をプロトコルに明記し、報告時には解析対象から除外した対象数とその理由を明記する必要がある。

全体集団のある一部のグループを対象とした解析をサブグループ解析（サブセット解析、部分集団解析）と呼ぶ。通常、患者特性やベースライン情報に基づいてサブグループ化を行う。「試験治療を完遂した集団」のように介入後の情報に基づいて対象を選択する場合には、比較可能性など別の問題が生じるため、サブグループ解析とは区別しておく必要がある。ランダム化の際に層別に用いた因子によるサブグループ解析は、ランダム化に基づく比較可能性の条件を

満たしている。そうでない場合は比較可能性の条件を満たさない。ただ、その条件を無視したとしても、多数の検定の実施により第Ⅰ種の過誤確率が上昇すること、一方ではサブグループ内の標本サイズ不足により検出力が低下することが問題となる。対応策としては、1) 関心のある少数のサブグループを事前にプロトコルに記載する、および2) 交互作用の検定が有意な場合にのみサブグループでの検定を行うことが推奨されている。交互作用には、量的交互作用と質的交互作用の2種類があり、交互作用が検出されたときは有意差の有無だけでなく、医学的な意義と解釈について十分な吟味が必要である。いずれにしても、サブグループ解析は、探索的解析の代表的なものであり、その目的は仮説の生成である。

3. 解析項目・方法

臨床試験で利用される標準的な統計解析手法について表1に示す。ランダム化対照試験では、ランダム化によって比較可能性が保証されているので、通常は観察研究のように複雑な回帰モデルを用いて交絡因子を調整する必要はない。

4. 中間モニタリング

中間モニタリング（中間解析、中間評価とも呼ばれる）の目的は、

- ・試験治療の優越性が疑いなく立証された場合
- ・適切な試験治療の差を示す見込みがないことが判明した場合
- ・許容できない有害事象が明らかになった場合

に試験を早期に中止することである³⁾。検定の多重性を考慮した中止規則の設定には多くの方法（グループ逐次法など）が開発されてきているが、実際に臨床試験を中止すべきかどうかという判断は純粹に統計的な問題ではなく、臨床的のみならず社会的な影響も考慮する必要がある。そのような判断を公正に行う場として、中間モニタリングを実施する際には、当該臨床試験に関与しない第三者からなるデータモニタリング委員会（効果安全性評価委員会とも呼ばれ

表1 変数の型別の標準的な統計解析手法

目的	連続変数	分類変数	時間-イベント変数
分布の記述	ヒストグラム、箱ヒゲ図、散布図	ヒストグラム、分割表	生存曲線 (Kaplan-Meier法)
要約統計量	平均、分散、中央値、パーセント点、相関係数	頻度、一致度、相関係数	x年生生存率、中央生存期間
検定(単純)	t検定、分散分析、Wilcoxon検定	χ^2 検定、Fisher正確検定	ログランク検定
検定(層別)	共分散分析	Mantel-Haenszel検定	層別ログランク検定
回帰モデル	重回帰分析	ロジスティック回帰分析	Cox回帰分析

る)を設置しなければならない。

探索的臨床試験のデザイン —ベイズ流の方法

1950年頃に臨床試験の方法論がほぼ確立して以来、統計的評価の方法として、フィッシャーあるいはネイマン・ピアソンによる頻度流(frequentist)の仮説検定・推定が主に用いられてきた。データ解析へのベイズ流統計学の適用は、物理学をはじめとする多くの自然科学分野および社会科学分野で広く行われており、医学・生物学分野においても、ベイズ流の統計モデルを適用したデータ解析の事例は多く存在する。しかしながら、臨床試験のデザインにベイズ流の方法を適用した事例としては、抗がん剤の第I相試験(最大耐用量を決定するための試験)でのCRM(continual reassessment method)、ランダム化試験の中間解析でのベイズ流予測確率の利用などがあるが、未だそれほど多くない。

大学等の研究機関が主体となって実施するトランスレーショナルリサーチおよび臨床試験の対象疾患は、難治性かつ重篤であり、そのうえ患者数が限られているという特徴がある。このような状況では、基礎研究で認められたコンセプトを実証するためのPOC(proof of concept)試験と呼ばれる探索的試験の実施が主であり、

疾患の重篤性などを考えると同時対照を設定すること自体困難な場合が多い。また、被験者のリスクを最小にするために臨床試験の途中で結果をモニタリングしながら意思決定を行うというような柔軟な対応も必要である。さらに、被験者数を最小にするために、過去に得られた証拠や情報(事前情報)を十分に生かすことも重要となる。これらを鑑みると、予期しない事態が発生して試験途中でデザイン(標本サイズや中間モニタリングの時期・方法など)の変更を行う場合などに、頻度流接近法に基づく方法は柔軟性の観点から不十分であり、新しい方法の開発が必要となる。近年、ベイズ流接近法は柔軟性と効率性の面から有望と考えられている。臨床試験におけるベイズ流接近法の主な特長は以下の通りである。

- ① 解釈が容易な「確率」だけを用いて整合性のある推測と意思決定を行うことができる
- ② 標本サイズに関わらず事前分布を事後分布に更新して推測ができる
- ③ 予測分布を用いて試験結果を予測することができる

また、ベイズ流デザインの動作特性が頻度流に評価できることも1つの利点である。

ベイズ流デザインの例

すべての被験者に同一の試験治療を行う単群臨床試験は、探索的な臨床試験の大部分を占めている。その主目的は、治療効果に対する確定的な証拠を得ることではなく、さらに研究を継続すべき有望な治療をスクリーニングすることである。単群臨床試験デザインの多くは、効果が認められない場合は試験を早期中止することが望ましい致命的な疾患の領域で開発されてきた。抗がん剤の第Ⅱ相単群臨床試験については、1960年代から頻度流の方法が開発され、1990年代以降ベイズ流接近法を用いたデザインがいくつか提案されている。その中には、効用／損失関数を明示的に用いるベイズ流決定理論に基づく手法も含まれる。

ここで、被験者20名にある試験治療を行い、14名に「効果あり（成功）」、6名に「効果なし（失敗）」という結果が得られた、二値（成功または失敗）評価項目の単群臨床試験という単純な事例を用いてベイズ流の方法を概説する。まず、事前情報が存在しないと仮定し、事前分布を一様分布（Beta(1,1)と表現されるベータ分布）と定める。次に、ベイズの定理を用いて、

事前分布と観察データ（実際には尤度 [ゆゑど] と呼ばれる形に変換されたもの）を結合し、事後分布 Beta(15,7) (=Beta(1+14, 1+6)) を得る (図4)。このように更新された事後分布から、この治療の成功確率の平均は0.68(=15/(15+7))、成功確率が0.5以下の確率は分布下面積から0.039と得られる。この結果から、引き続いて5名の被験者に同じ治療を行ったときに何名の被験者に成功が観察されるかという予測分布を得ることもできる (図5)。

おわりに

フィッシャーは1938年に次のように述べている：「同じだけの時間と労力をかけたとしてもデータ収集の過程、または実験計画を厳密に検討しているか否かによって、得られる収穫は10倍から12倍にもなる。実験終了後に統計学者に相談を持ちかけるのは、統計学者に、単に死後診察を行って下さいと頼むようなものである。統計学者はおそらく何が原因で実験が失敗したかという実験の死因について意見を述べてくれるだけであろう」⁴⁾。

臨床試験に統計的方法は必須であり、計画段階から生物統計学の専門家が参画していれば、

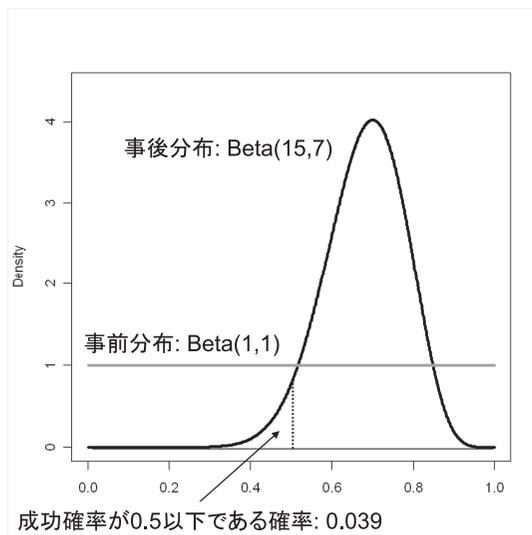
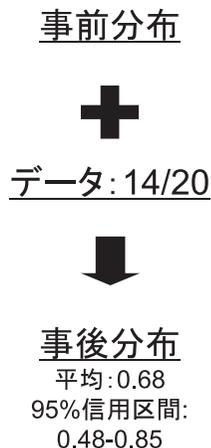


図4 事前分布から事後分布へ

事後分布



予測分布
次の5名のうち、
何名の成功が
観察されるか？

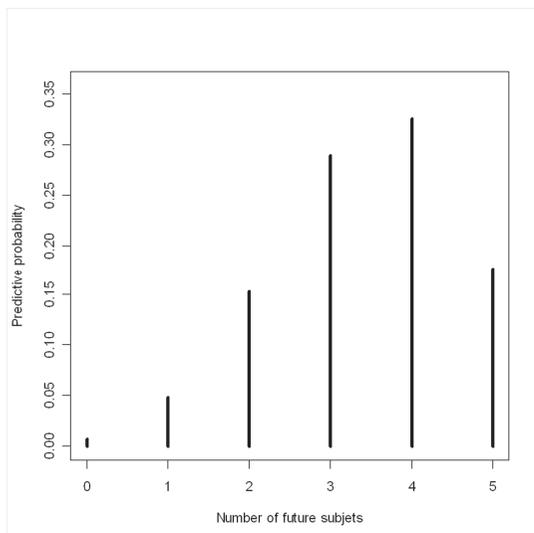


図5 事後分布から予測分布へ

質の高い臨床試験を実施できる可能性は高い。近年、臨床試験にベイズ流の方法が有用であるという報告は着実に増えている。効率的かつ倫理的な試験デザインの開発は、資源を有効に活

用するという観点からも今後ますます重要になるであろう。

開示すべき潜在的利益相反状態はない。

文

献

- 1) 手良向聡, 大門貴志. 臨床試験デザイン—ベイズ流・頻度流の適応的方法. メディカル・パブリケーションズ. 2014.
- 2) 竹内 啓. 統計学的な考え方—デザイン・推測・意思決定—. 統計学の基礎Ⅱ. 岩波書店. 2003: pp1-53.
- 3) 厚生省医薬安全局審査管理課長. 「臨床試験のための統計的原則」について. 臨床評価 1999; 27: 161-206.
- 4) C.R. ラオ. 統計学とは何か—偶然を生かす. 筑摩書房. 2010.

著者プロフィール



手良向 聡 Satoshi Teramukai

所属・職：京都府立医科大学大学院医学研究科生物統計学・教授

略歴：1985年3月 神戸大学理学部卒業
1985年4月 三共株式会社医薬情報部
1998年1月 アムジェン株式会社開発本部
2002年4月 京都大学医学部附属病院探索医療センター助手
2006年12月 京都大学医学部附属病院探索医療センター准教授
2013年4月 金沢大学附属病院先端医療開発センター特任教授
2014年6月 京都府立医科大学大学院医学研究科生物統計学教授，現在に至る

専門分野：臨床統計学，臨床研究方法論

- 主な業績：1. Teramukai S, Daimon T, Zohar S. A Bayesian predictive sample size selection design for single-arm exploratory clinical trials. *Stat Med* 2012; 31: 4243-4254.
2. Teramukai S, Kitano T, Kishida Y, Kawahara M, Kubota K, Komuta K, Minato K, Mio T, Fujita Y, Yonei T, Nakano K, Tsuboi M, Shibata K, Furuse K, Fukushima M. Pretreatment neutrophil count as an independent prognostic factor in advanced non-small-cell lung cancer: An analysis of Japan Multinational Trial Organisation LC00-03. *Eur J Cancer* 2009; 45: 1950-1958.
3. Kubota K, Kawahara M, Ogawara M, Nishiwaki Y, Komuta K, Minato K, Fujita Y, Teramukai S, Fukushima M, Furuse K, on behalf of the Japan Multi-National Trial Organisation. Vinorelbine plus gemcitabine followed by docetaxel versus carboplatin plus paclitaxel in patients with advanced non-small-cell lung cancer: A randomised, open-label, phase III study. *Lancet Oncol* 2008; 9: 1135-1142.
4. Egawa H, Teramukai S, Haga H, Tanabe M, Tanaka K, Fukushima M, Shimazu M. Present status of ABO-incompatible living donor liver transplantation in Japan. *Hepatology* 2008; 47: 143-152.
5. Teramukai S, Ochiai K, Tada H, Fukushima M. PIEPOC: A new prognostic index for advanced epithelial ovarian cancer — Japan Multinational Trial Organization OC01-01. *J Clin Oncol* 2007; 25: 3302-3306.
6. Teramukai S, Nishiyama H, Matsui Y, Ogawa O, Fukushima M. Evaluation for surrogacy of end points by using data from observational studies: Tumor down-staging for evaluating neoadjuvant chemotherapy in invasive bladder cancer. *Clin Cancer Res* 2006; 12: 139-143.