Online publication July 1, 2011

臨床研究における統計学的解析
―推定と検定の正しい使い方―

赤澤　宏平　Md. Aminul Hoque　張　楠　凌　一葦　齋藤　翔太

要　旨：臨床研究の計画や収集されたデータの解析を行う際に，統計学の基本的な知識は必須のものとなる。本稿では，治療効果の大きさや患者群の特徴を数値で示すための推定の方法を解説する。平均値と標準偏差，中央値とパーセント点の使い分けに注意が必要である。また，治療効果が本当にあったのか，有害事象の発現率は有意に高かったのか，などを立証するための統計学的な有意差検定の手順とその選定方法を説明する。連続データの分布の 2 群比較について，5 種類の検定手法の使い分けを述べることにする。

（J Jpn Coll Angiol, 2011, **51**: 167–173）

Key words: estimation, statistical test, statistical analysis, clinical trial

はじめに

臨床医は，より質の高い医療を実践するために，自らが行っている医療についてその科学的根拠を収集する必要がある。収集され統合された最新の科学的根拠は，日常の診療に生かされるはずである。たとえば，実際の医療現場では，①治療法の優劣評価，②医薬品の有害事象の発生頻度，③疾患発症などリスクに関与する因子，④疾患の発症，生存余命，治癒等の確率予測などを，患者やコメディカル・スタッフに説明する場面がしばしばあるであろう。この場合，大規模臨床試験やバイアスの少ないレトロスペクティブ・スタディの論文からの引用が役立つはずである。これらの論文では，集積されたデータをその研究目的とデータの性質に合った統計手法で解析している。したがって，臨床医は論文上のこれらの推定値や検定結果を誤りなく理解し解釈する素養を求められる。

その一方で，臨床医は自らの臨床経験やデータの積み重ねにより，新たな科学的根拠を創出する場合もある。臨床医が学術論文をまとめる際に，正しい結論を少ないバイアスでしかも効率よく導く研究デザインの立案が必要となるが，その過程で統計学の基本的な知識や統計解析のノウハウが用いられることも忘れてはならない。学術雑誌「脈管学」でも多くの科学的根拠が発表されてきた。「脈管学」の最近数年分の原著論文を参照すると，記述統計量の算出や単一因子解析に加えて相関分析，分散分析，多変量解析など，研究の目的に合致した解析手法が用いられている。

このように，医学統計学は科学的な根拠を作り出すことに貢献してきた。脈管学の臨床研究の科学的根拠を国際的なレベルで創出するためには，統計学の基礎理論の概要を把握し，研究デザインおよび統計解析手法の使い方と解釈を理解しておく必要がある。

本稿では，洗練された研究デザインの立案や収集されたデータの適切な統計解析に役立つ，(1)患者群の特徴や治療効果の大きさなどを数値で示す推定の方法，ならびに，(2)有意差検定の基本的な手順と解析方法，について解説する。

1. 推　定

医学・医療における統計解析の目的のひとつは，患者属性の特徴，注目する治療法の治療効果の大きさなどを数量化することである。そのためには，年齢分布，疾患の重症度，抗腫瘍効果および有害事象の発生率を数値で示す必要がある。このように，臨床研究における重要な指標を統計学的な数値で表すことを**推定**という。

新潟大学医歯学総合病院医療情報部

2011 年 3 月 7 日受理

Continuous data

Dichotomous data

Ordinal data

		Clinical pathway	
		No use	Use
No. of cases		90	96
Age(years)		64.5±10.6	64.8±11.3
Gender	male	58	66
	female	32	30
Hypertension		88(97.8%)	93(96.9%)
Smoking		57(63.3%)	61(63.5%)
Extent of dissection (DeBakey)	Ⅲa	24(26.7%)	21(20.9%)
	Ⅲb	66(73.3%)	75(78.1%)

THE JOURNAL of JAPANESE COLLEGE of ANGIOLOGY Vol.47,2007

Figure 1 Various estimates for continuous, dichotomous and ordinal variables of patients characteristics.

統計解析の目的のふたつめは，新治療法の治療効果が従来の治療法よりも格段に優れているのか，それとも偶然誤差の範囲内なのかを判定することにある。この判定のことを**検定**という。推定と検定の考え方を理解しこれらを目的に合わせて正しく使うことが臨床医にとっても必要となる。そこで，まずは推定を理解することにしよう。

1-1. 患者特性，治療効果，有害事象の人数分布を特徴づけるための推定

推定とは，患者特性，治療効果，有害事象の発生率，診断精度などを適切な統計量(数値)で表すことをいう。統計量とは，平均値，標準偏差，中央値(25%点，75%点)，割合，生存率，オッズ比，ハザード比などのことであり，とくに，推定のために用いた統計量のことを**推定値**という。

Fig. 1 は，クリニカルパスの非適用群と適用群とで患者属性に大きな違いがないことを示すための推定結果である。非適用群と適用群の 2 群間で，背景因子(年齢，性別など)の人数分布に違いがあるのかどうかを調べている。表中の適用群の年齢分布は，平均値と標準偏差が 64.8±11.3 歳であるが，これは横軸に年齢，縦軸に度数をとったヒストグラムが左右対称のひとつ山で，その中心位置が 64.8 歳，53 歳(= 平均値 − 標準偏差)から 76 歳(= 平均値 + 標準偏差)の間に症例が約 64 例(全体の約 66%)入っていることを示唆している。つまり，平均値と標準偏差は“正規分布の中心位置と広がり”を表す推定値といえる(**Fig. 2** 上)。

年齢分布の場合，ひとつ山の左右対称の正規分布で近似できることが多いが，その他の医学データでは必ずしも正規分布で近似できない。たとえば，健常者 100 名の AST(GOT)値 25，42，34，80，…の分布(ヒストグラム)は左右対称とならず，30 あたりにピーク(推定値としては最頻値と呼ぶ)がきて右に 80，100，150 などと尾をひく分布となる(**Fig. 2** 下)。こういう分布に対しては平均値 ± 標準偏差で分布の特徴づけを行うことができず，むしろ中央値と(25%点，75%点)などのパーセント点を用いて分布の位置と広がりを表す。

患者特性の人数分布以外に，医学的な治療や介入(**Fig. 1** のクリニカルパスなど)の成績を推定する場合にも，その大きさを適当な推定値を用いて示すことができる。**Table 1** は，クリニカルパスの適用により非適用に比べて，肺炎発症率，人工呼吸器の使用率および酸素投与量がどのように変化したかを割合と平均値 ± 標準偏差で表したものである。適用群の方が改善していることが読み取れる一方で，酸素投与量に関しては，平均値に比べて標準偏差が極めて大きな値であり，正規分布に従うとはいえない可能性も示唆している。

1-2. データの性質による推定値の使い分け

医学・医療で収集されるデータを統計学的に分類すると 4 種類ある。①**連続データ**(年齢，生化学検査の値など)，② **2 値データ**(性別，既往歴の有無など)，③**順序データ**(臨床進行期，自覚症状など)，④ **Time to event**

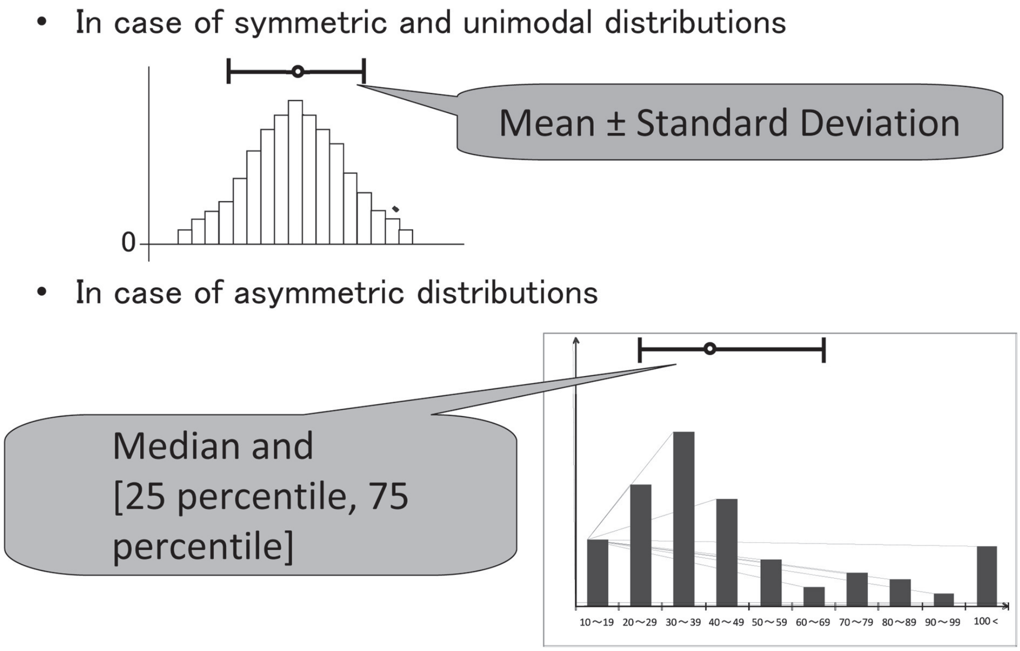

Figure 2　Summary of how to use estimates of mean±standard deviation and median [25 percentile, 75 percentile].

Table 1　Estimation for representing the effects of clinical pathway on several clinical indicators

Endpoints	Clinical pathway		P-value
	No use (n=90)	Use (n=96)	
Incidence of pneumonia	23.3%	4.2%	<0.001
Ventilator required	11.1%	4.2%	0.080
Quantity of oxygen	1,219±1,039	311.3±430	<0.001

（The Journal of Japanese College of Angiology Vol. 47, 2007）

データ(生存時間データなど)である。これらのデータの性質により，使われる推定値も違ってくる。**Fig. 3** はそれぞれに適した推定値をまとめたものである。2 値データや順序データでは，群全体の症例数に対する各カテゴリーの人数の割合を用いて人数分布の特徴づけを行う(**Fig. 1** の高血圧ありの割合，切除範囲)。

平均値 ± 標準偏差は多くの人が知っているなじみの推定値であるが，実は使用に際して注意を要する。前述のとおり，臨床試験で収集されるデータの分布が正規分布(左右対称でひとつ山の分布)に近いときにのみ，平均値と標準偏差は用いることができる。それ以外は中央値[25%点，75%点]を用いなければならない。健常者の体重や身長などは，経験的に正規分布に従う(ヒストグラムが正規分布の形になる)が，臨床試験などで得られる生化学検査値や血液検査値は正規分布に従わないことがしばしばある。

2. 検　　定
―新しい治療法は本当に有効か？を判定する―

ここでは，無作為化臨床試験の事例に基づき，統計学的検定の手順についてまとめてみる。臨床試験で使われる検定には，**有意差検定**と**非劣性検定**の 2 種類がある。ごく簡単に定義すると，有意差検定は新しい治療法が従来の治療法に比べて極めて高い治療効果が得られ

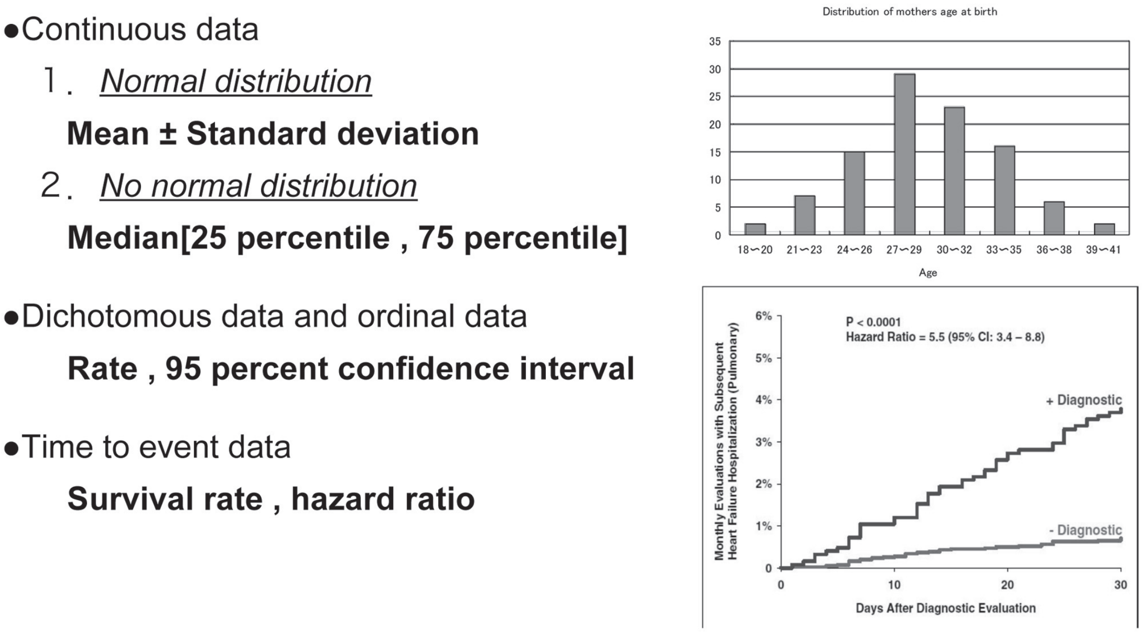

Figure 3 Estimates for continuous, dichotomous, ordinal, and time-to-event data.

ることを立証するための検定であり，非劣性検定は，新治療法が従来の治療より劣ることはないことを示すための検定である。有意差検定と非劣性検定のいずれを用いるかは，臨床試験を計画する段階で検討すべきものであり，それにより試験デザインも違ってくる。

本稿では有意差検定に限定して説明する。**有意差検定**は，新治療法が従来の治療法に比べて確かな効果がある，もしくは，有害事象が著しく減少することを立証するときに用いる。実際の検定では，比較する群の数，収集されるデータの性質によって，その用途にあった検定手法が用いられる。2群の連続データの分布に差があるかどうかを検定するためのt–検定，2群以上の比率の有意差検定を行うためのカイ2乗検定などは，臨床試験で頻繁に用いられる検定手法の例である。いろいろな検定手法があるが，その手順はほぼすべて同じである。検定に関する要点を以下にまとめた[1]。

2-1. 臨床試験の目的を統計学的な仮説におきかえる

検定では，2つの仮説，**帰無仮説**と**対立仮説**を立てる。2つの仮説のうち，どちらが正しいのかを臨床試験データに基づき判定する。新治療法が従来の治療法に比べて奏効率が高いかどうかを判定したい場合に，2つの仮説は以下のように表される。

帰無仮説：新治療法の奏効率＝従来の治療法の奏効率

対立仮説：新治療法の奏効率≠従来の治療法の奏効率

この2つの仮説の意味を補足すると，対立仮説は「新治療法は，従来の治療法に比べて，偶然誤差とはいえない，新治療法を行ったことによる確かな奏効率の向上がある」という意味である。そして，この「新治療法を行ったことによる確かな奏効率の向上」のことを**有意な差**という。一方の帰無仮説は，「新治療法と従来の治療法の奏効率に多少の差はあったにせよ，それは症例の選択の偏りや未知の要因による偶然誤差であり，新治療法による差ではない」という意味である。

2-2. 臨床試験データから得られる新治療の有効性の証拠P-値

どちらの仮説が正しいのかを判定するために，使われる統計用語は検定統計量，P-値，有意水準であるが，これらの詳しい説明は本稿では割愛する。ごく簡単にこれらを説明すると，**検定統計量**とは，新治療群と従来の治療群との治療効果（検査値，奏効率，生存率など）の違いや有害事象の程度の差を標準化したもので，臨床試験データから直接計算される。**P-値**は帰無仮説が正しいとしたときに検定統計量が得られる「得られやすさ」を表した指標であり，検定統計量が求まると検定統計量の理論

Table 2　Summary of statistical tests: two hypotheses, level of significance, and p-value

Key words of significance tests

1. Two statistical hypothesis
   (Ex.) Null hypothesis (H_0): $\mu_A = \mu_B$
   Alternative hypothesis (H_A): $\mu_A \neq \mu_B$
   (μ_A: mean of A group, μ_B: mean of B group)

2. Level of significance α
   Usually, α=0.05

3. Decision making of rejecting the null hypothesis
   When $P<\alpha$, reject null hypothesis and accept alternative hypothesis
   When $P\geq\alpha$, accept null hypothesis

的な分布よりすぐに求まる。P- 値は 0 以上 1 以下である。P- 値がきわめて小さな値のとき，帰無仮説が正しいとしたときに得られる確率が低い，別の言い方をすれば，対立仮説が正しい確率が高まることになる。

2-3. 帰無仮説，対立仮説いずれを採択するかの判定基準としての有意水準

有意水準は帰無仮説，対立仮説のどちらが正しいかを判定する基準値であり，医学研究ならびに臨床試験では多くの場合 0.05 を用いる。最終的に，P- 値が有意水準より小さいとき，「帰無仮説が正しいと仮定したときに，きわめてまれにしか起こらない現象が起きた」として対立仮説を採択する。逆に，P- 値が有意水準より大きいとき帰無仮説を採択する。

上述の 2-1 から 2-3 までのまとめを **Table 2** に示した。

3. 検定手法の選定

医学研究で用いられる検定手法は，論文に出てくるものだけでも数十種類程度に大別でき，データの性質により使い分けが必要となる。使い分けを考える際に，以下の視点でみると理解しやすい。

(1)検定に用いるデータは連続データ，順序データ，2 値データ，Time to event データ？

連続データ，順序データ，2 値データ，Time to event データにより，解析手法が全く異なるので，統計解析の専門書やソフトウエアのマニュアルなどを参照する際には，まず，このデータの性質に着目するとよい。この中で，順序データを計測値(あるいは評価尺度)とするときの群間比較，たとえば，5 つのカテゴリーからなる病期分類について新治療群と対照治療群の 2 群間で有意差があるかどうか，を検定する場合には，後述のノンパラメトリック検定を用いる。また，Time to event データは，ある事象が発生するまでの時間を指すが，ある事象には，死亡，再発以外にも，治癒，退院，疾患の発症などがありうる。Time to event データと連続データの違いは，連続データは収縮期血圧のように計測値が確定するのに対して，Time to event データは，観察期間中に研究の終了，症例の転居などにより，「その時点までは事象が発生していないことが確認されたが，その後いつ事象が起こったか不明である」データ，すなわち，打ち切りデータが存在する点にある。つまり，Time to event データは値の確定しない不完全データを含むという点が重要である[2)]。

(2)比較する群はいくつか？

既存の臨床試験の奏効率と今回実施する臨床試験の奏効率を比較する場合，1 群のみ基準値との比較となるので，比較する群はなし(もしくは 1 群)となる。自家静脈グラフトを用いたバイパス手術施行患者を対象に，HMG-CoA 還元酵素阻害薬の投与群と非投与群とで静脈グラフト開存に与える影響を調べたいときには 2 群比較となる。また，3 群以上での比較を行う臨床試験もある。

(3)対応のあるデータか，それとも，独立したデータか？

対応のあるデータには，自己対合データと年齢，性別などが一致した症例の対(ペア)とがある。自己対合型データは，治療前の検査値と治療後の検査値のペアがその例である。一方の独立したデータは，A 群と B 群に無作為に割り付けた患者群の検査値や効果の有無に関するデータなどがある。

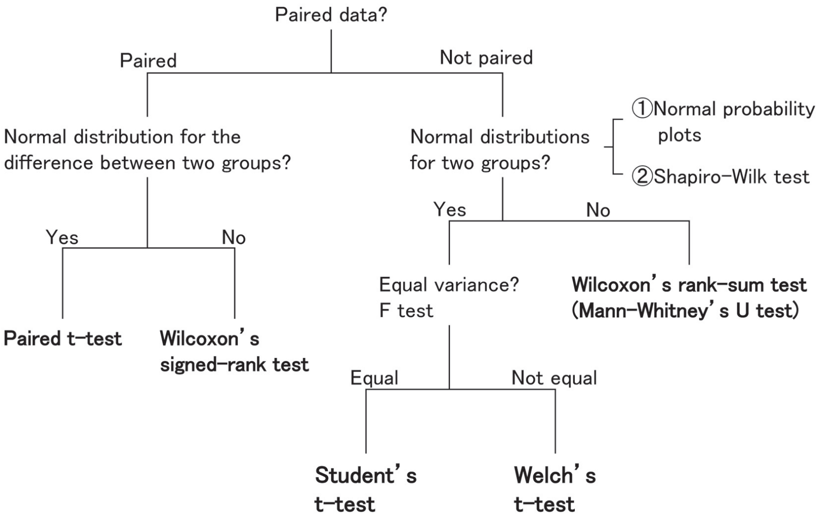

Figure 4　Flow chart for discriminating statistical tests to compare two distributions of continuous data.

(4)検定に用いるデータが連続データの場合，度数分布が正規分布に近いか？

横軸に連続データをいくつかに区分けしたカテゴリー，縦軸にその人数を取ったときの棒グラフ(度数分布)が，左右対称のひとつ山(正規分布)になっているかどうかが問題となる。臨床試験で採取される種々の検査値は必ずしも正規分に近いとはいえないこともある。このようなデータを用いた検定では，**ノンパラメトリックな検定**手法を用いるべきである。ノンパラメトリックな検定では，各群で連続データを小さい順に並べてそれらに順番を付与して，その順番の分布が群間で有意に異なるかどうかを調べることになる。

このような点を踏まえて，2群の連続データの分布の比較を行う場合の検定手法の使い分けを**Fig. 4**に示した。**Fig. 4**には3種類のt-検定が記述されているが，いずれも元のデータの人数分布が正規分布(左右対称の一つ山の分布)になっていることが前提条件となる。それに対して，非正規分布の2群の位置に有意な差があるかどうかを調べる検定手法として，Wilcoxonの順位和検定(もしくはMann-WhitenyのU検定)とWilcoxonの符号付順位和検定とがある。データに対応があるか否かにより，2つの使い分けが必要となるし，論文を読む際にも2つは別物であると理解したうえで読み解く必要がある。元のデータの人数分布が正規分布になっていないときの検定を，**ノンパラメトリック検定**と総称することがある。

5. まとめ

臨床研究で用いられる推定値の使い分けと検定の手順，検定手法について解説した。限られた紙面の都合で十分な説明はできていないが，論文を読み解く，あるいは，臨床データを解析する際のポイントとなるべき点はおおよそ網羅したつもりである。現在の日本の医学教育では，医学統計学を習得するための十分な時間が取られていない。したがって，統計学の基礎知識が欠如するのは仕方ないことである。ある程度の知識や技術を習得しつつその先の統計処理は，統計解析専門家に依頼することもひとつの選択肢であると考える。

文　献

1) 柳川　堯，荒木由布子：バイオ統計の基礎．近代科学社，東京，2010.
2) 赤澤宏平，柳川　堯：サバイバルデータの解析．近代科学社，東京，2010.

Statistical Methods in Medical Research: How to Use Estimates and Tests

Kouhei Akazawa, Md. Aminul Hoque, Nan Zhang, Yiwei Ling and Shota Saito

Department of Medical Informatics, Niigata University Medical and Dental Hospital, Niigata, Japan

Key words: estimation, statistical test, statistical analysis, clinical trial

Clinicians need to gather recent clinical evidence regarding their specialties to provide medicines of high quality. They should use integrated clinical evidence with respect to treatment effects, rates of adverse effects, risk factors of diseases, and prognoses of patients in their clinical practices. Most clinical evidence is derived through statistical analyses, so to correctly gather clinical evidence, clinicians have to learn basic information about statistical theories and methods and how to interpret them.

This article will mainly explain how to use statistical inferences such as estimations and tests and how to interpret the output of statistical analyses. We expect that our explanation will help clinicians understand medical statistics.

(J Jpn Coll Angiol, 2011, **51**: 167–173)

Online publication July 1, 2011