

—特集 [臨床研究をはじめてみよう! (4)]—



医学統計解析のいろは

大塚 俊昭^{1,2}¹ 日本医科大学衛生学公衆衛生学² 日本医科大学付属病院臨床研究総合センター

はじめに—なぜ医学統計が必要なのか—

近年、医学・医療は著しい進歩を遂げる一方、未だ不完全でもある。実際、日常臨床の現場では、成書や文献を検索しても解決しない疑問にしばしば遭遇する。医師や医学研究者はこれらの疑問を解決し、ひいては医学・医療の進歩に貢献すべく研究を行い、その結果を論文等で公表しようとする。そして、研究の成果を公表するためには、得られたデータに対する正しい統計解析が必要とされる。しかし、新たな知見を研究成果として公表する際になぜ統計解析が必要なのか？それは、医学研究は科学であり、科学である以上その成果は客観的に解釈される必要があるためである。そこで本稿では、自身が臨床研究を実施する際に必要不可欠かつ最小限と思われる医学統計解析に関する知識について触れる。

基本的な統計関連語句

最初に、医学統計解析を実施する上で知っておく必要がある統計関連語句について解説する。

・平均（平均値）、中央値と正規分布

平均とはデータの中心位置を示す基本統計量であり、一般的には「算術平均（相加平均）」を指す。データの合計をデータの個数で割って得られる値と定義される。医学統計ではその他に「幾何平均」や「加重平均」などが用いられる場合があるが、ここでは割愛する。以下の記載における平均（平均値）とは全て算術平均を指すこととする。

中央値とは、平均値とともにデータの中心位置を示す基本統計量である。データを大きさの順に並べ替えたとき、ちょうど順番が真ん中となる値と定義される。データ数が奇数の場合はちょうど順番が真ん中となる値を求めることができるが、データ数が偶数の場合には順番が真ん中となる値は求められない。その場合は、真ん中を挟んで前後する2値の平均を求め、これを中

央値とする。

ともにデータの中心位置を示す統計量である平均値と中央値であるが、その使い分けはデータが正規性を有するか否か、すなわち正規分布しているときみなせるか否かで判断する。正規分布とは、左右対称の釣鐘型で左右に裾を引き、平均 $\pm 1 \times$ 標準偏差、 $\pm 2 \times$ 標準偏差、および $\pm 3 \times$ 標準偏差の範囲内にデータの68.3%、95.4%、および99.7%が含まれる形状の分布を指す（図1）。データが正規性を有する場合は平均値、データ分布が偏っている場合は中央値を用いる。解析対象となるデータの正規性は、その変数が一般的に正規性を有するか（あるいは偏った分布を示すか）を関連文献などで調査したり、度数分布図（ヒストグラム）を描出しその分布形状に基づいて正規性を判断するか、Shapiro-Wilk 検定やKolmogorov-Smirnov 検定などの統計学的検定により判断する。ただし、これら正規性検定の帰無仮説は「このデータは正規分布に従っている」であるため、サンプル数が大きな標本の場合、ヒストグラム上は正規分布とも判断し得る形状を示していても検定において帰無仮説が却下される、すなわち「正規分布とはいえない」と判断される場合があることに注意を払う必要がある。

・分散と標準偏差

分散とは、個々のデータの平均値からのばらつき具合を示す指標である。個々のデータと平均値との差を求め、その値を2乗し、それらを合計したものをデータの個数で割って得られる値と定義される。図2に分散の例を示す。サンプル数8個の標本において、個々の値と平均値との差を $a \sim h$ とすると、

$$\text{分散} = \frac{a^2 + b^2 + c^2 + d^2 + e^2 + f^2 + g^2 + h^2}{8} \dots\dots\dots (1)$$

と計算される。個々の値と平均値との差を2乗するのは、平均値との差をそのまま合計すれば平均値の性質上必ず0となるためである。

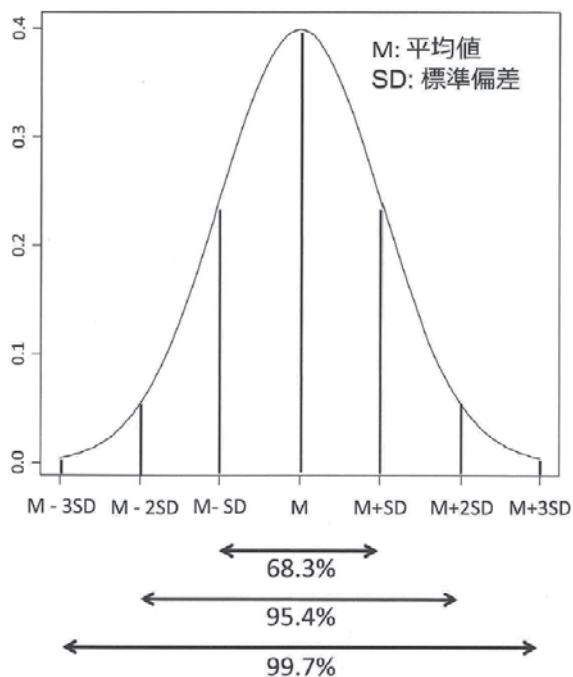


図1 正規分布

標準偏差とは、分散の平方根である(図2)。平方根をとる理由は、分散の単位がもとの単位の2乗となるために、平均値と単位を合わせる必要があるためである。すると、平均値〇〇mg/dL、標準偏差××mg/dLなどと同じ単位のもとの併記が可能となる。なお、正規性を示さない分布におけるばらつきの表現方法としては、25パーセンタイルから75パーセンタイル間の範囲である4分位範囲(Interquartile Range; IQR)で示すことが多い。

分散には標本分散と不偏分散がある。上記(1)式は標本分散、すなわち目の前で統計処理を行っている標本についての分散を表す。これに対して、(1)式の分母をデータ数-1として求めた分散を不偏分散と言い、母集団の分散(母分散)を推定するとき用いる。母集団とは、われわれが研究や調査を行いたいと考えている対象の全体のことである。これに対し、母集団から抽出した一部分の集団のことを標本と呼ぶ。多くの場合、母集団を対象とした研究はマンパワーや経済的理由等から実現困難であるため、われわれは通常、標本を対象とした調査を行う。ここで、母集団の分散(母分散)は、母集団が標本と比べて十分大きいときには標本の不偏分散にほぼ等しいことが知られている。統計解析の大きな目的の一つは標本調査によって母集団の特性を推定することなので、標本から母集団の分散を推定したい場合に標本の不偏分散を求めることになる。なお、平均については、標本平均と母集団の平均

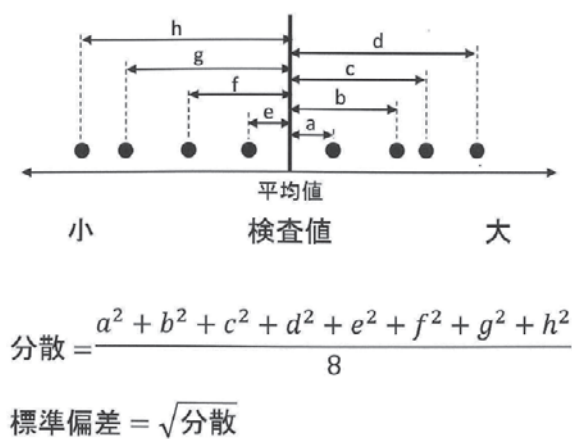


図2 分散と標準偏差の計算方法

は一致する。

仮説検定とP値

我々が行う医学研究の多くは、ある疑問を明らかにする手段として統計学的仮説検定を用いる。つまり、ある仮説を設定し、その仮説が成立する確率を統計学的検定により求め、その確率とあらかじめ定めた棄却水準とを比較することにより、設定した仮説を棄却するか否かを判定する。確率の英語Probabilityの頭文字Pをとって、この値はP値とも呼ばれる。一般的には、この棄却水準は5%に設定する。多くの研究では、ある事象について「差がある(差がつく)」ことを期待するが、仮説検定の際は「差がない」という仮説(帰無仮説)をまず設定する。この時点で、帰無仮説の逆の仮説である対立仮説(「差がある」という仮説)も自動的に設定される。そして統計学的検定により、帰無仮説が成立する確率が5%未満であれば、帰無仮説は棄却され対立仮説を採択する。この結果が、論文などでよく見かける「統計学的に有意な所見を認めた」といった記載となる。ここで注意しなければいけないのは、あくまでもこの検定は帰無仮説を棄却できるか否かを判定することが目的だということである。帰無仮説を棄却できなかったイコール帰無仮説を採択する、とはならない。したがって、棄却できなかった時の解釈は、「有意な差があるとはいえない」であり、決して「差がない」でない。この点を十分理解するための補足として、ある物を探している状況を想定する。もし見つければ、その物は「ある」と言えるが、見つからなかった場合はあくまでも「見つからなかった(あるとはいえない)」のであって、必ずしも「ない」とは言い切れない。なぜなら、単に探し方が不十分なかだけかもしれないし、もしくはもっと広い範囲を探せば見つかるか

もしれない。つまり、「ある」ことを証明できないイコール「ない」とは必ずしも言えないのである。繰り返しになるが、帰無仮説を棄却できなかった場合の解釈は、「統計学的に有意な差があるとはいえない」であることを十分に理解して欲しい。

上述のように、医学統計では帰無仮説棄却の水準を5%未満に設定することが圧倒的に多い。ところで、なぜその水準は5%なのだろうか。最初に5%を採用したのはR.A.Fisherという統計学者なのだが、その理由は「何となく」だったという記録がある¹。また、V.Cohnは、コイントスで5回連続表が出たら、それを見ていた人が「声」を上げた、と記録している¹。4回連続および5回連続で表が出現する確率は各々6.3%、3.1%なので、人間は5%より低い確率で出現する現象に出会うと、「本当に偶然なのか？」という違和感が生じるのかもしれない。いずれにせよ、今日まで慣例的に設定されている5%という水準は案外適当な線引きなのである。なお余談ではあるが、それがゆえに、過剰に有

意水準5% (P<0.05) にこだわる人のことを、一部の統計学者は「ピーチ病 (=P 値病!）」と呼んでいるらしい。

検定と推定

母集団と標本については上述したが、臨床研究では、母集団の特性に関する仮説が正しいかを標本から求める「検定」と、母集団の特性が収まる範囲を標本から求める「推定」という二つの考えのもとで統計解析を実施する(図3)。すなわち、検定とはその標本において帰無仮説が成立する確率を求め、有意水準0.05と比較するという「点」に基づいて統計学的有意性を判断する。一方、推定とは母集団の特性をある区間で推定することであり、検定で用いる有意水準0.05に対応した95%信頼区間をもって有意性を判断する。ここで、95%信頼区間とは、母集団から標本を取ってきて信頼区間を求めるという作業を100回実施したとき、95回はその中に母集団の値が含まれると考えられる区間のことである。両者は統計解析のアプローチ法の違いであり最終的な結果は同じとなる。すなわちその表現方法がP値で表されるか、95%信頼区間で表されるかの違いとなるわけだが、確率という「点」で表現される検定よりも、母集団の値を「区間」で表現できる推定の算出がより望ましいと考えられている。

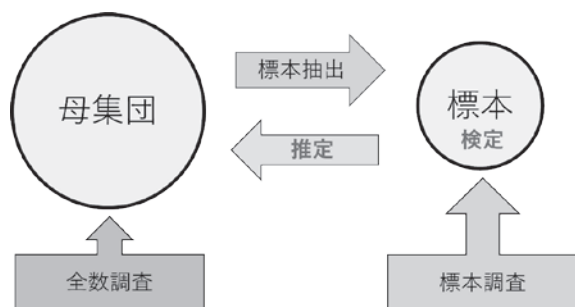


図3 検定と推定。検定とは標本を用いて母集団の特性に関する仮説を統計学的に検証すること、推定とはその標本が属する母集団の特性を統計学的に検証すること。

臨床研究における統計手法の実際

実際の臨床研究において利用される代表的な統計解析手法を表1に示す。誌面の関係で全ての解析方法について言及は出来ないが、以下のケースにおける統計解析手法の実際について述べる。

表1 臨床研究における代表的な統計解析手法

解析の目的	変数の型			
	2値変数	連続変数		生存時間
		正規分布	非正規分布	
データの記述	頻度集計/割合	平均±標準偏差	中央値(四分位範囲)	Kaplan-Meier
2群間の比較 (対応なし)	カイ二乗検定 Fisherの正確検定	t検定	Mann-Whitneyの U検定	Log-rank検定
2群間の比較 (対応あり)	McNemar検定	対応のあるt検定	Wilcoxonの符号 順位検定	
3群以上の比較	カイ二乗検定	分散分析 共分散分析	Kruskal-Wallis検定	多群Log-rank
関連性	ロジスティック回帰	相関分析(Pearsonの累積相関係数, Spearmanの順位相関係数) 線形回帰		Cox比例 ハザードモデル

表2 多重比較の統計解析手法

対照群と各群との比較
・ Dunnett 法
すべての2群同士の比較
・ Tukey-Kramer 法
・ Bonferroni 法
・ Steel-Dwass 法
・ Games-Howel 法
など
(Fisher's LSD 法は現在推奨されていない)

・独立した群間の2値変数の比較

2値変数の比較とは、具体的には割合の差の比較となる。通常は χ^2 検定を用いる。2群間の比較はもちろん3群間以上の比較も可能である。ただし、対象者数が少ない場合（正確には、期待値1未満のセルが1つでも存在する場合や期待値5未満のセルが全体の20%以上存在する場合）は χ^2 検定の正確性が低下するため、Fisherの正確検定を適用する。なお、Fisherの正確検定は計算上の複雑さから従来の統計ソフトでは2群間の検定しか利用できなかったが、近年のPC性能の向上から、3群間以上の検定も可能となりつつある。

・対応のある群間の2値変数の比較

対応のある群間とは、ある集団における一定時間の経過後や薬剤投与前後といった時間軸における2点、同一人物における右目と左目といった対となる器官、症例対照研究において対をなしている症例と対照、などが想定される。これら対応のある群間における割合の差の検定は、2群間であればMcNemar検定、3群間以上であればMcNemar検定の拡張版であるCochranのQ検定を用いる。

・独立した群間の連続変数の比較

連続変数が正規分布に従うと見なせる場合は、2群間の平均値の比較であればt検定、3群間以上における平均値の比較であれば分散分析を用いる。正確には2群間が等分散であるか否かにより使用するt検定の種類が異なってくるが、後者の場合に用いるt検定（Welchのt検定）は等分散の場合でも適用可能なので、近年、統計ソフトではWelchのt検定を標準設定として計算されるようになりつつある。

3群間以上の検定で注意が必要なのは多重検定の問題である。分散分析における帰無仮説は「群間の平均値に差はない」であり、この帰無仮説が検定の結果棄却されれば、結論は「群間の平均値に有意差がある」

となる。ここで、「群間の平均値に有意差がある」とはいつでもどの群とどの群に有意差があるかまでは明らかにできない。ここで次の手段として思いつくのは、各群をまたいだ全ての2群の組み合わせパターン（例えば全体が3群であれば3通りの2群間の組み合わせ）においてt検定を行うこと（多重検定）であろう。しかし、以下の理由からこの方法は統計学的に御法度とされている。3群間の比較で考えたとき、まず、分散分析で帰無仮説が棄却できない（この場合は3群間に差があるとはいえない）確率は95%である。一方、1回あたりのt検定で帰無仮説が棄却できない（両群間に差があるとはいえない）確率も95%である。したがって、t検定によって3群間のいずれにも差があるとはいえないことを示すには3回繰り返すt検定の全てにおいて帰無仮説が棄却できないことを示さねばならないが、その確率は0.95の3乗で0.86（86%）となる。ということは、3群間のいずれかに差があることが示される確率は $1-0.86=0.14$ （14%）となり、本来の水準である5%より大幅に甘くなる。これが多重検定を御法度とする理由である。この対策として、3群以上の多群間の比較を行うための様々な多重比較の統計手法が提案されており（表2）、その多くが分散分析を実施せずとも用いることが可能である。これらの詳細は専門書に譲るが、ここではBonferroni法について簡単に触れる。これは帰無仮説棄却の水準となるP値を調整する方法で、多重検定を繰り返した回数で本来の帰無仮説棄却水準を除し、その値を新たな棄却水準として用いるというものである。3回の多重検定が必要であれば、0.05を3で除した0.0167（1.67%）をP値の棄却水準として、検定から導かれたP値を評価することになる。この方法は、検定の回数が多くなると有意水準が厳しくなりすぎるといった短所があるものの、多群間の割合の差の検定（ χ^2 検定、CochranのQ検定、等）にも応用が可能で、理解が容易であるため、知っておいた方が良いであろう。

連続変数が正規分布に従うと見なせない場合はノンパラメトリック検定を用いる。2群間の比較であればMann-WhitneyのU検定、3群間以上の比較であればKruskal-Wallis検定を用いる。3群以上の各群から組み合わせた2群間を検定するノンパラメトリック多重比較の方法は存在しないため、その場合は各群の組み合わせにおいてMann-WhitneyのU検定を行い、Bonferroni法を用いてP値の棄却水準を調整する方法を適用する。

さて、これらの検定はどの程度のデータ数があれば適用できるのだろうか。パラメトリック検定の場合

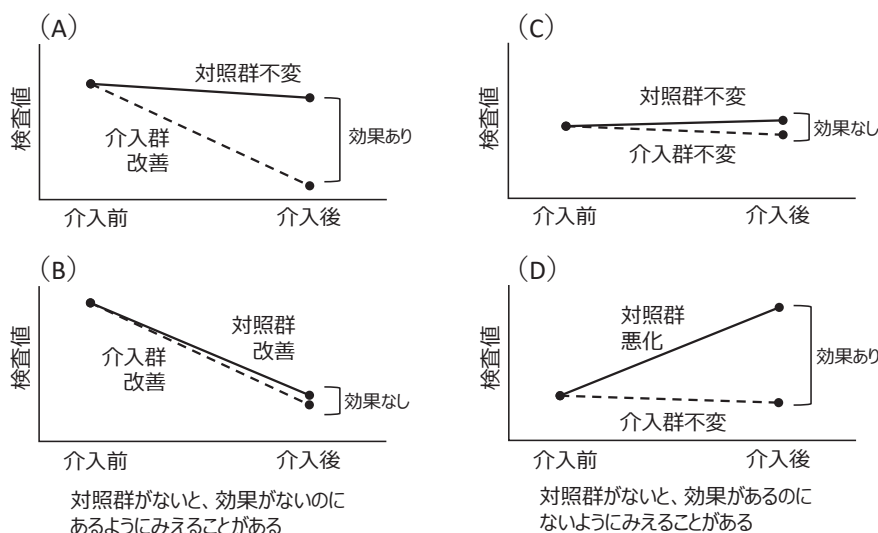


図4 介入群および対照群における介入前後の検査値の変動

は、基本的に正規性があると判断できる（推測できる）だけのデータ数（最低で各群 15 程度以上）が必要と考えられており、それ以下の場合にはノンパラメトリック検定を用いることになる²。しかし、ノンパラメトリック検定は解析対象となるデータ数が各群 7 未満だと極端に検出性能が低下し、各群 6 未満ではそもそも有意性の判定が不可能となる³。この点からも、少なくとも各群最低 7 例以上、可能であれば 15 例以上を収集する研究デザインとすることが、統計解析の視点からは望ましいといえる。

・対応のある群間の連続変数の比較

連続変数が正規分布に従うと見なせる場合は、2 群間の平均値の比較であれば対応のある t 検定、3 群間以上における平均値の比較であれば対応のある分散分析を用いる。この時、上述した対応のある群間の 2 値変数の比較にも当てはまるが、結果の解釈において注意が必要となることがある。例えば、対応のある 2 群間の連続変数の平均値を比較する研究として、薬剤投与など介入前後における検査値を比較する研究が考えられる。しかし、検定の結果、介入前と比べ介入後の検査値が有意に改善したからといって、その介入が有効であると結論付けることはできない。すなわち、**図 4A, B** の点線のように、たとえ介入群で介入後に検査値が改善したとしても、**図 4B** のように対照群でも同様に検査値が改善していた場合、介入の影響で検査値が改善したとはいえない（時間の経過などその他の要因で変化したと解釈するのが妥当である）。**図 4C, D** も同様であり、介入前後で検査値の改善がなくても、**図 4D** のように対照群で数値の悪化があれば、介入の

効果（悪化を予防する効果）を認めたことになる。介入しないことが倫理的に問題となる場合はこの限りではないが、介入の効果を示すためには原則として対照群を設定し、群内の前後比較ではなく、各群におけるベースラインからの変化率や変化量について群間比較を行うなどの研究デザインが求められる。

・2 変数の関連性

関連性を明らかにしたい 2 変数に、説明する側にあたる変数と説明される側にあたる変数の位置付けを特に想定しない場合は、相関係数を算出する。変数同士の線形関係を評価する場合はパラメトリック検定である Pearson の累積相関係数を用い、変数同士の順位に基づいた相関関係を評価する場合はノンパラメトリック検定である Spearman の順位相関係数を用いる。実際には相関係数 (Pearson の r, Spearman の ρ (ロー)) と P 値を算出し有意性を判断する。相関係数の算出に変数の形式上の制限は特にないが、通常は連続変数または離散変数の関連性を判断する場合に用いる。一方、研究仮説に基づき、関連性を明らかにしたい 2 変数を説明する側（説明変数または独立変数）と説明される側（目的変数または従属変数）に各々位置付ける場合は回帰分析の適用となる。目的変数が連続変数の場合は線形回帰分析、2 値変数の場合はロジスティック回帰分析を用いる。線形回帰分析は回帰係数および標準化回帰係数と P 値を算出、ロジスティック回帰分析は回帰係数またはオッズ比および P 値を算出する。可能な限り区間推定すなわち 95% 信頼区間も併記する。ロジスティック回帰式は、元々は確率（割合）を導くための式 (**図 5**) である。この式を、(疾病など対象とな

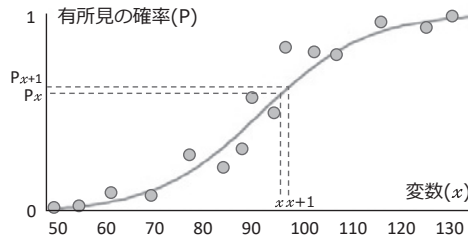
ロジスティック回帰式
= 「確率 (割合)」 を求める回帰式

$$p = \frac{1}{1 + e^{-(\beta x + z)}}$$

β: 回帰係数
z: 切片

↓ 変形

$$\frac{p}{1 - p} = e^{\beta x + z}$$

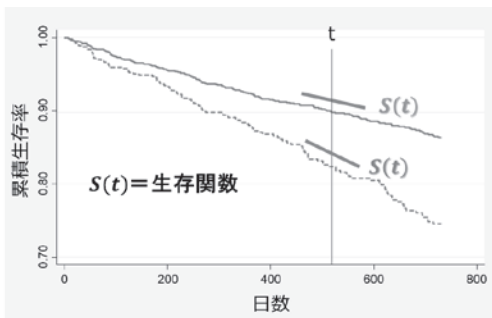


$$\text{オッズ} = \frac{\text{「有所見あり」の確率}}{\text{「有所見なし」の確率}} = \frac{P}{1 - P}$$

x 1単位増加毎の有所見に対するオッズ比: x のオッズに対する x+1 のオッズ

$$\begin{aligned} \text{オッズ比} &= \frac{e^{\beta(x+1)+z}}{e^{\beta x + z}} && \text{オッズ比} = \frac{x+1 \text{のオッズ}}{x \text{のオッズ}} = \frac{P_{x+1}/(1 - P_{x+1})}{P_x/(1 - P_x)} \\ &= e^{\beta(x+1) - \beta x} && x \text{の1単位増加毎の有所見に対するオッズ比は} \\ &= e^{\beta} && \mathbf{Exp(\beta)} \end{aligned}$$

図5 ロジスティック回帰分析におけるオッズ比の算出



ハザード: ある時間 t における、その瞬間(t+Δt, Δt→0)の死亡率
↓
時間 t における、生存関数の傾き $(-\frac{d}{dt} \ln(S(t)))$
↓
∴ **ハザード比 ≡ 「生存関数の傾き」の比**

図6 Cox 比例ハザードモデル

る事象が) ありの確率 (割合)/なしの確率 (割合) すなわちオッズを求める式に変形, さらに異なる説明変数 (x) の値を本式に代入し, オッズ比を求める.

説明変数を回帰式に複数含める多変量ロジスティック回帰分析の際, モデルに含めることができる説明変数の数は, 明らかな統計学的な根拠はないものの, アウトカム (目的変数) 発症数と非発症数の少ない方の数10~20に対して説明変数1つが目安と考えられている.

・生存時間分析

研究対象者の予後の追跡調査には生存時間分析を用

いる. 生存時間分析とは, イベントが起きるまでの時間とイベントとの間の関係に焦点を当てる分析方法である. 生存時間分析は「生存」という言葉が用いられているが, 必ずしも生存/死亡情報のみならず, その研究で規定した疾病発症などのイベントも扱うことができる. 従って, 「生存率」は「イベント回避率」や「無事故率」などと表現されることもある. 累積生存率は通常 Kaplan-Meier 累積生存曲線から算出, また生存曲線の群間差については Log-Rank 検定または一般化 Wilcoxon 検定にて判定を行う. 通常は Log-Rank 検定が使用されることが多いが, 一般化 Wilcoxon 検定は追跡初期に群間差が広がる場合に検出しやすいため, その研究内容によって両者を使い分ける必要がある. 一方, イベント発症 (イベント回避) の要因を分析し, その関連性を明らかにする解析方法として Cox 比例ハザードモデルが広く用いられている. Cox 比例ハザードモデルはロジスティック回帰の要素を盛り込んだ分析方法であり, イベント発症に関連する因子 (説明変数) の相対危険度 (ハザード比) を求めることができる. ここで, ハザードとは任意の時間 t における, その瞬間の死亡率すなわち生存関数 (正確には \ln (生存関数)) の傾きである. よって, ハザード比とは生存関数の傾きの比と表現することができる (図6). ここからも分かるように, Cox 比例ハザードモデルは2つの生存関数の傾きが任意の時間 t において等しく比例 (どの時間 t でも同じ比を取る: 比例ハザード性が成立) することを仮定したモデルであるため, 同モデルを用いる際は比例ハザード性の成立を確認する必要がある.

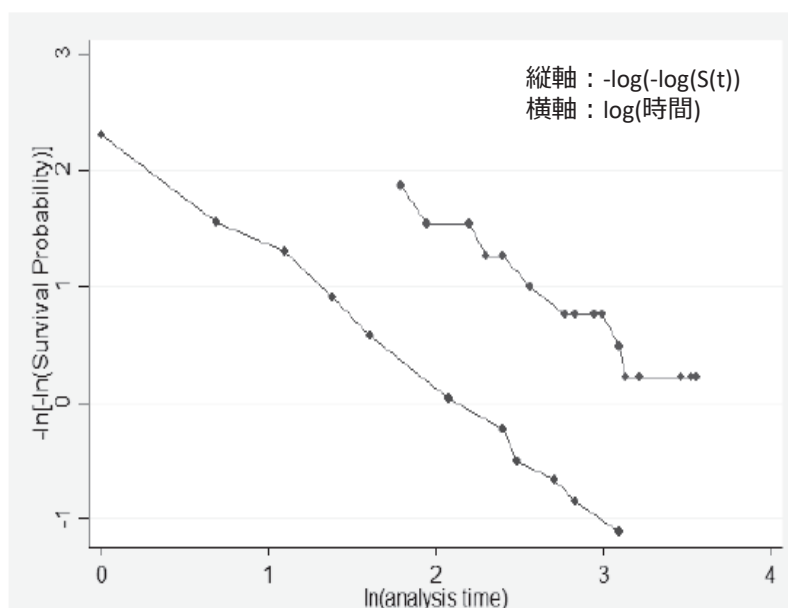


図7 比例ハザード性の確認 (log-log プロット)

ある。比例ハザード性の確認は、視覚的な方法としては縦軸に $-\log(-\log(\text{生存関数}))$ 、横軸に $\log(\text{時間})$ をとったグラフを描写し、両直線が概ね平行であれば成立していると考えられる(図7)。検定による確認は、Schoenfeld 残差の算出などで行うことができる。もし比例ハザード性が認められなかった場合は、Weibull 回帰など他の生存時間分析を用いて解析することになる。

多変量 Cox 比例ハザードモデルにおいてモデルに含めることができる説明変数の数は、ロジスティック回帰分析の場合と同様に、イベント発生数と非発生数の少ない方の数10~20に対して説明変数1つが目安と考えられている。

・診断精度の分析

疾病診断を目的とした新たなバイオマーカー等のモダリティ開発の際には、その診断精度の評価が必須である。基本的な診断精度の評価指標として、感度、特異度、偽陰性、偽陽性、陽性尤度比、陰性尤度比、陽性反応的中度、陰性反応的中度があげられる(図8)。感度・特異度はトレードオフの関係にあり、検査の目的および疾病の性質に基づき至適カットオフ値を設定する。すなわち、疾病の見逃しを少なくすること(=陰性なら疾病なし)に重点を置く場合は感度の上昇を優先したカットオフ値を、疾病の誤診断を少なくすること(=陽性なら疾病あり)に重点をおく場合は特異度の上昇を優先したカットオフ値を定める。感度と特異度からその診断モダリティの性能を評価する場合

は、縦軸に感度、横軸に1-特異度をプロットした受信者動作特性(Receiver Operating Characteristic; ROC) 曲線を描出し、その曲線下面積(AUC)を評価する。明確な基準はないものの、AUCが0.8~0.85以上を認めれば臨床に耐えうる性能を有する検査と判断できるであろう。尤度比は感度・特異度から求める。陽性尤度比はその検査が陽性であることによって疾病診断の確率(正確にはオッズ)がどの程度上昇するかの指標であり、陰性尤度比はその検査が陰性であることによって疾病診断のオッズがどの程度低下するかの指標である。陽性尤度比が高く陰性尤度比が低い検査ほど疾病診断に優れた検査と言える。陽性反応的中度と陰性反応的中度は目的とする疾病の有病率に依存する。陽性反応的中度は検査陽性の場合に疾病を有する確率に相当し、1-陰性反応的中度は検査陰性の場合に疾病を有する確率に相当する。

臨床の現場における疾病の診断は、数種類の検査を組み合わせられて行われることが多い。また、新たな診断モダリティを開発する場合、既存の検査に追加で検査を行うことによる診断精度の改善能を示すことが求められる。新たな検査の追加による診断精度の改善を示す解析方法の一つとして、ロジスティック回帰分析を用い、従来の検査モデルおよび従来の検査に新たな検査を追加したモデルの各々における疾病保有確率を計算し、それらの確率を用いて描出したROC 曲線におけるAUCを比較する手法がある。すなわち、従来検査のモデルのAUCと比較し新たな検査を追加したモデルのAUCに有意な増加を認めれば、新たな検査の

		疾病		計
		あり	なし	
検査	陽性	a	b	a+b
	陰性	c	d	c+d
計		a+c	b+d	a+b+c+d

$$\begin{aligned} \text{感度} &= \text{疾病あり群における検査陽性の確率} = \frac{a}{a+c} \\ \text{特異度} &= \text{疾病なし群における検査陰性の確率} = \frac{d}{b+d} \\ \text{偽陽性} &= \text{疾病なし群における検査陽性の確率} = 1 - \text{特異度} \\ \text{偽陰性} &= \text{疾病あり群における検査陰性の確率} = 1 - \text{感度} \\ \text{陽性尤度比} &= \frac{\text{疾病を有する人が検査陽性となる確率}}{\text{疾病を有さない人が検査陽性となる確率}} = \frac{\text{感度}}{\text{偽陽性率}} \\ \text{陰性尤度比} &= \frac{\text{疾病を有する人が検査陰性となる確率}}{\text{疾病を有さない人が検査陰性となる確率}} = \frac{\text{偽陰性率}}{\text{特異度}} \\ \text{陽性反応的中度} &= \text{検査陽性群における疾病ありの確率} = \frac{a}{a+b} \\ \text{陰性反応的中度} &= \text{検査陰性群における疾病なしの確率} = \frac{d}{c+d} \end{aligned}$$

図8 診断精度の各種評価指標

追加は統計学的には有用であろうと判断できる。しかしながら、AUCは識別能に関する指標であり、これだけで新たな検査が臨床的に使えるか？を判断するには不十分である。すなわち、従来モデルの疾病保有確率に対する新モデルの疾病保有確率がより正確な方向、つまり疾病を有すると診断された者における新モデルの疾病保有確率は上昇へ、疾病を有すると診断されなかった者における新モデルの疾病保有確率は低下へ向かうか、といった再分類の評価軸も取り入れる必要がある。その統計手法の一つとして、純再分類改善度 (Net Reclassification Improvement: NRI) と統合識別改善度 (Integrated Discrimination Improvement: IDI) がある。もともとNRIは、リスク分けされた対象者に対して検査を追加することによってより正確なリスクカテゴリに再分類される割合を計算する指標であったが、その後カテゴリレスNRIとしてリスクカテゴリを考慮しない方法へと応用されるようになった。本稿ではその汎用性に鑑みカテゴリレスNRIをNRIとして紹介する。NRIは、疾病を有すると診断がなされた人数に対して、新たな検査を追加したことによって疾病保有確率が上昇した人数から低下した人数を引いた人数が占める割合と、疾病を有すると診断されなかった人数に対して、新たな検査を追加したことによって疾病保有確率が低下した人数から上昇した人数

を引いた人数が占める割合の和で表される(図9)。IDIは、新たな検査を追加したことによる、疾病を有すると診断された者の疾病保有確率の変化(上昇)値と、疾病を有さなかった者の疾病保有確率の変化(低下)値の個人レベルにおける総和である。NRIとIDIは誤分類の重みが考慮されないという欠点を批判されることもあるが、現在でも比較的広く受け入れられる指標となっている。

交絡因子の調整

交絡因子とは、「要因」と「アウトカム」の関係に影響を与える第3の因子のことである。交絡因子の条件として、1) アウトカムと要因の両方に影響し、2) 要因からは影響されない、ことが必要となる。一例としてライターの所持と肺癌を取り上げ説明する。ライターの所持と肺癌の発症率との関連を調査したところ、ライター所持者で肺癌の発症率が有意に高かったという結果が認められた。しかしながらその機序としては、ライターを所持している人は喫煙者が多く、喫煙が肺癌の発症率上昇に関与していることは明白である(決してライター所持自体が肺癌リスクを高める訳ではない)。すなわち、この例での「喫煙」は、ライター所持と肺癌の関連における交絡因子になる。これは典型的な例ではあるが、現在でも学会発表

• NRI

		疾病		計
		あり	なし	
追加検査による 疾病保有確率	上昇	a	b	a+b
	低下	c	d	c+d
計		a+c	b+d	

$$NRI = (a - c)/(a + c) + (d - b)/(b + d)$$

• IDI

	疾病	
	あり	なし
追加検査による 疾病保有確率の変化値の合計	a	b*

*マイナス方向への変化を正とする

$$IDI = a + b$$

図9 純再分類改善度 (NRI) と統合識別改善度 (IDI)

や一部の公表論文において、交絡の影響と思われる結果を報告した研究をしばしば目にする。交絡の問題をクリアするための対策としては、研究段階における対象者の限定、マッチング、あるいはデータ解析の段階における層別化、多変量解析、傾向スコアを用いた分析、などが考えられる。

終わりに

本稿では、医学統計のいろはとして臨床研究を実施する際に知っておきたい指標、統計手法、およびこれに関連する考え方を記した。しかしながら、臨床研究で最も大切なことは、その研究結果が「臨床的意義」を有するかどうかである。統計学的有意性の判定は臨床的意義を保証するものではない。統計学的な有意性は、ある現象（研究結果）がどのくらいの確率で出現しうるのかを統計モデルに基づいて算出し、人為的に定められた有意判定水準を満たすか否かによって判断されているに過ぎないことを、本稿を読んだ方であれば理解できるであろう。その研究結果に臨床的意義があるかどうかは、様々な医学知識を総動員して個別に判断されるべきものである。一方で、冒頭で記したように、医学研究は科学である以上、その結果に対して客観的な判定がなされる必要がある。したがって、少なくとも現在の医学研究に統計解析は必須である。大

切なのは、医学研究の成果を発信する人、その成果を受け取る人の双方が正しい医学統計知識を身につけ、その結果を正しく解釈することである。本稿によって読者の多くが医学統計に興味を持ち、ひいては読者の統計解析リテラシーの向上に貢献できれば幸いである。

Conflict of Interest：開示すべき利益相反はなし。

文 献

1. 折笠秀樹：P値論争の歴史. 薬理と治療 2018; 46: 1273-1279.
2. 浜田知久馬：解説論文 ノンパラメトリック検定の考え方. アプライドセラピューティクス 2015; 6: 63-69.
3. 永田 靖：多重比較法の実際. 応用統計学 1998; 27: 93-108.

(受付：2022年5月8日)

(受理：2022年5月9日)

日本医科大学医学会雑誌は、本論文に対して、クリエイティブ・コモンズ表示 4.0 国際 (CC BY NC ND) ライセンス (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) を採用した。ライセンス採用後も、すべての論文の著作権については、日本医科大学医学会が保持するものとする。ライセンスが付与された論文については、非営利目的で、元の論文のクレジットを表示することを条件に、すべての者が、ダウンロード、二次使用、複製、再印刷、頒布を行うことができる。