

## ■セミナー 臨床研究の枠組み

# 臨床研究における統計処理とその問題点

三宅由子\*

**要旨：**本稿では、統計的方法が臨床研究に適用される際の問題点について簡単に解説する。主な内容は、①統計的有意差検定の考え方の基礎（標本と母集団、帰無仮説の棄却、大数の法則）、②研究計画の立て方、③記述と仮説検証、④統計的検定による「有意差」と研究としての「意味ある差」の違いについて、である。 神経心理学 8:74~78

**Key Words：**臨床研究, 記述統計, 統計的仮説検定, 研究計画, 統計的有意差  
clinical research, descriptive statistics, testing statistical hypothesis, research planning, statistical significance

## I はじめに

臨床研究に用いられる統計処理について、誤用や乱用が指摘されることは少なくない。その多くは、統計的方法を誤った常識に基づいて適用しているためのように思われる。たとえば、統計学が研究の一番最後の段階に至ってから登場する場合がある。とにかくなにか数字を出して、あとで「星」——すなわち有意差——をつけるためだけに統計学を用いる。このような事態は、統計的仮説検定の考え方について誤解があるために生ずるように思う。この誤解を少なくするために、統計学を応用する上の基礎となる考え方について述べてみたい。基本的なことであるが、上記のような統計の用いられ方をみても、また筆者が受けた相談の経験からも、この点についての解説の必要性を感じることが多い。また「統計的仮説検定」のハウツーを求められる統計学の本では、あまり解説されていないという事情もあるように思う。

## II 症例研究—少数例と多数例

臨床研究のひとつの基盤は症例の個別研究で

あり、1例の症例はすべての出発点であるともいえる。そして1例の症例から把握された事実がどの症例にも見られるとすれば、それが法則性や普遍性に結び付く可能性を考えるのはよい。しかし多数例を集めれば群として扱ってよいとは必ずしも言えない。症例が偏って選ばれていると、みかけ上の法則性が現われる場合があるからである。

症例を群として扱い、統計的仮説検定を用いる際の前提として、無作為抽出標本というものがある。全ての症例を調べるわけにいかない場合、手に入るデータ——標本——からそれを推測する。その際に標本が「偏り」のないものであるという保証である。臨床データでは、厳密な意味でこのような標本を手に入れることはほとんどできない。しかしそれでも有意差検定が行われているのは、入手したデータをそのようなものと見なしているということである。つまり取り上げた疾患に関して、手に入る症例群を疾患全体の代表として取り上げても大きな問題はない場合である。ある疾患である病院を訪れた患者全員を標本とすれば、よほど特殊な事情がないかぎり、日本におけるその疾患の代表

1992年3月2日受理

The Utilization of Statistical Methods for Clinical Research

\*東京都精神医学総合研究所, Yuko Miyake: Tokyo Institute of Psychiatry

とみても、あまり不都合ではないだろう。しかし自分の証明したい仮説に合うものだけを取り出し、合理的な理由付けのないままそれを標本として扱うなら、それは無作為抽出標本とはみなしえない。

つまり、臨床データを統計的仮説検定を適用する標本として扱うには、最低限、結果を有意にするための作為が働いていないことが条件となる。こんなことは研究者の心得として当然のことと思われるかもしれないが、実際「この1例を抜いてみたらどうだ、星が付くだろう」などという会話を、一度は耳にしたという方は多いのではないだろうか。しかし証明すべき仮説を発見するまでの段階では仕方がないとしても、仮説を実証するには単なる「寄せ集め」データは適当とはいえないことが多い。発見すべき仮説を模索する、文字どおり手探りの段階では、どのようなデータの取り方をすると、どのようなデータが集まるかということだけに意味があり、有意差検定に大きな役割を担わせることはできない。この限界をはっきりわきまえておくべきであろう。手元のデータを、なるべく情報を失う事なく見やすい形にまとめていく方法に、探索的データ解析がある。度数分布の幹葉表示や箱ヒゲ図などはよく用いられるが、ここで詳しく説明している紙幅はないので、興味のあるかたは参考書(Hartwig, 1979)を参照していただきたい。

### III 記述的研究と仮説検証的研究

統計的研究には事実を数字で語るために統計学を用いる研究と仮説を検証するために統計学を用いる研究がある(三宅, 1991)。症例の観察を積み重ね、症例の集団を数字で記述することには大きな意義がある。症例の大多数が同じ特徴をもっているなら、それはその病気についてのなんらかの手掛かりを示している可能性がある。しかし症例(群)研究だけでは、疾患とその特徴の関連に意味があるとは言えない。たとえば、ある疾患の患者のうち90%が虫歯の治療経験をもっていたというだけで、虫歯の治療とその疾患の関連を言い切る人はいない。その

疾患にかからない人でも、同じくらいの治療経験があると想像できるからである。この例はあまりに単純だが、「虫歯治療」を他の一見意味ありげなもの置き換えた場合、考え違いをする可能性がでてくる。つまり、症例を集めそれを数字で記述することは、さまざまな仮説を浮かび上がらせる。しかし浮かび上がった仮説を実証するためにはそれなりの手続きが必要である。例えばその疾患ではない人の虫歯治療経験を調べるなど、症例群において観察された事実の意味があることの証拠をつかまえられるように、研究を計画する必要がある。言い換えれば、仮説を発見するための統計的研究と、仮説を検証するための統計的研究があるということになる。仮説検証を目指した臨床研究は特に慎重に計画されるべきであろう。

### IV 臨床研究の計画——動機から評価まで

治療の責任を負いながら臨床研究をするには、さまざまな困難が伴う。それだけにデータが無駄にならないよう、また所期の目的が達せられるようにしなければならない。そのために、臨床観察から得られた研究への動機を言語化し、大目標と個別の目標を明確化し、目的達成のための戦略を考え、それに必要な資源を検討し、結果をどのように公表し評価するかまでを、簡単なものでよいから計画書として書いてみることをお勧めする。それも頭の中で組み立てるだけでなく、実際に文字として書いてみるのが重要である。その詳細は他の文献(三宅, 1992)に譲るが、ぜひ一度は検討していただきたい。一見面倒にみえるかもしれないが、計画書によって得られるものは大きい。まず実際の作業をする上での無駄が少なくなる。次に協力してくれる患者さんやスタッフに説明しやすくなる。スタッフの間で調査について共通の認識をもつことは、データの信頼性を増すうえで非常に重要なことである。また近年、患者のプライバシーにかかわる問題を扱ったり、直接患者に協力を求める調査研究では、治療と同様に「説明したうえでの同意 informed consent」が要求されるようになってきている。現に外国

の専門誌では、論文を発表する際の必要事項として、投稿規定に記載されていることがある。最後に結果を論文にまとめるときにも、計画書は大変に役立つ。

この作業の目的は、まず言語化による明確化である。なにをやりたいと思ひ、そのためになにが使える、その結果なにがえられるか。これを実際に調査に動く前に机上でたどってみることであり、思ひ浮かんだ仮説を現実と照らし合わせて検討することでもある。動機から大目標へ、大目標から個別の目標へ、現実化していくほどに困難さが見えてきて、実際にできなくなってしまうという危惧をもたれるかもしれない。また現実化するほどに、当初の想像が当たらなかったり、できないことが分ったりして、つまらなく見えるかもしれない。しかし逆にあらかじめ予想される困難なら、それを回避するか乗り越えるかの手を考えることができる。思ひ付きの面白さだけに振り回されて、とんでもない結論を導く危険も減るだろう。また個々の研究はそれほど興味を引かないとしても、積み重ねの結果に見通しがあれば、こつこつと継続していくことも楽しみがある。そして当初想定した目標が達成不可能という結論に達したら、大目標の下でもういちど達成可能な現実的目標を考慮してみることもできるだろう。

疫学研究におけるバイアス（偏り）の問題（Wynder, 1990）は特に近年しばしば取り上げられるようになったが、その中で重要なもののひとつとして、研究者の意識したあるいは無意識の願望が、結果に反映する可能性が論じられている。もちろん完全にではないが、研究計画を立てることによって、バイアスのかかる危険はかなり回避することができる。なにをやりたいと思ひて研究を始めたか、そのためにどのような方法を用いてどのような対象を調べるかを検討すると、言いたい結論を整合性をもって導くにはどうすればよいかが見えてくるからである。

## V 統計的仮説検定の考え方

実際の統計的仮説検定のやりかたの基礎につ

いては成書（柳井, 1977；高木, 1991）に譲り、ここでは検定の際に現われやすい誤解（三宅, 1992）とそれに基づく検定結果の読みちがいについて述べよう。誤解の多くは統計的仮説検定についての次のふたつの原則に対するものであり、そこから生ずる誤りであるように思われる。①統計的仮説検定は「帰無仮説の棄却」によって有意差ありとする。つまり二重否定の論理であること。②絶対値として同じ差であっても、標本数が多いほど有意差と判断されやすくなる（大数の法則）こと。

まず、帰無仮説が棄却されたときには有意差ありとするのは正しい。しかし棄却されなかった場合に「差なし」と断言するのは間違いであり、「差の有無については何とも言えない」というのが正解である。差がないことを証明する統計的手段はない。帰無仮説が棄却されなかったのは、本当に差がないのか、あるいはこの程度の標本数では差が検出できないのかのいずれかである。逆に有意差があるからといって、それが直ちに「意味のある」データであるとは限らない。つまり検定の結果有意差が出たとしても、それは必ずしも差の絶対値の大きさを保証しているわけではない。たとえば非常に大量のデータによってごくわずかの差を検出したとしても、それがどれほどの「意味」をもっているかについては、有意差検定の守備範囲の外なのである。

また、帰無仮説が棄却されなかったとき、症例数が多くなると有意になると言われ、症例を追加して有意差を得た経験をもつかたは多いのではないだろうか。これがとにかく症例を集めさえすればよいという誤解・曲解のもとであり、それがまた統計への不信感を生む原因にもなる。大数の法則とは、標本数が多いほどわずかな差を検出できるという原則である。したがって、特に記述的な研究の段階では、統計的に有意であるかないかは、実際その数字を取り上げる意味があるか否かについて論じる材料のごくごく一部なのである。

多重比較の問題も近ごろよく言われるようになった。3群以上の群について例えば平均値の

差の検定をする場合、一元配置分散分析が適用される。この方法では全体として群間のばらつきと群内のばらつきが比べられ、前者が相対的に大きい場合、有意になる。しかしそれだけではどの群とどの群の間に差があるのかは分からない。そこで2群間の比較をするが、これをただの2群の差の検定で行うと、組み合わせの数は、3群なら3、4群なら6、5群なら10になり、それだけ検定が繰り返される。こうなると有意水準が全体として大きくなり、たとえば5%を越えるという事態が起こり得る。そこで、全体として有意であった場合、その内部での2群の比較をする方法が考察されている。すでにコンピュータの統計パッケージには組み込まれているものが多い。

最後に最も誤解の多い相関関係の検定について述べる。相関係数の検定は、無相関の検定である。したがって、0でなければどんなわずかな相関であっても有意になる。しかし有意の相関と聞くとそれがかなり強い相関であると誤解されているのではないかと思われる場合が多々ある。標本数が百での相関係数0.2、あるいは千例の0.1は、無相関の検定をすれば有意の相関ありになるが、散布図を描いてもあまり関連があるようにはみえない。一方10例で0.6という相関係数は有意ではない。しかしもし症例数の多い研究で同じ数値が得られたとしたら、かなり強い相関である。つまり有意にならないからといって、0.6という相関係数まで否定することはない。つまり、有意であるからといって「強い」相関関係があるとは限らず、有意でないから相関がないとは言えない。それを知らずに相関係数とその検定の数値だけに目を奪われると、検定結果が有意というだけで、ごく弱い相関を検出しただけなのにあたかも完全に相関するかのよう誤解をする危険がある。

## VI おわりに

統計的仮説検定における帰無仮説の棄却という考え方、大数の法則、有意差と意味のある差の違いについて理解されると、統計に対する誤解のかなりの部分が解消するのではないだろう

か。また記述的な統計の大切さ、姿・形をとらえることが「数字で語る」というデータ解析の始まりだということも銘記しておきたい。実際に目の前にあるデータが、どういう意味をもつ「標本」なのかを検討し、データの姿・形をよくみておくと、有意差検定にどれほどの意味をもたせることができるか、その標本が代表している母集団はなにか、という検討を常に意識することができる。そうしていれば、結果の読み過ぎ、無理な推論、不必要な検定などを避けることが可能だと思う。

暗黙のうちに都合の悪いデータを削除し、結論にとって都合のよい症例をかき集めるという作業がされている場合があるだろうと思う。それは研究者の良心と姿勢の問題ととらえられがちだが、一方では学会や学術雑誌に発表する都合からの影響も受ける。もちろん、統計的仮説検定の方法は、観察されたデータについて、一定の前提の下に、合理的な推論の手順を与える。主観だけにたよった推論は排除されなければならない。しかし、それをあまりに重くみすぎると、統計的仮説検定さえしてあれば、あとはどうでもよいということになる危険性がある。統計的有意差検定が正しく用いられることによって、臨床研究はより豊かな成果を手にすることと思う。

最後に、専門家との交流をお勧めしたい。生兵法は怪我のもと。最近いろいろな統計的方法が発展してきたので、その応用範囲や前提について、一般論が難しくなっている。もちろん統計の専門家の側に問題がないとはいえないこともあるが、逆に今の日本では臨床研究のわかる生物統計学者が育つ基盤が弱いとも言える。その強化のためにも、タテワリの専門領域の枠を超え、互いにその分野の専門家として歩み寄ることによって、より豊かな研究成果が得られる可能性は非常に大きいと思う。互いに信頼をもって分からない点を明確化しながら共同研究をすることができれば、臨床研究における生産性は飛躍的に増大すると思う。疫学、公衆衛生、衛生、医療情報などの部門に点在する専門家の活用を考えていただきたい。本音で共同

研究ができるような人的資源をさがすことも、臨床研究の大切なポイントなのかもしれない。

#### 文 献

- 1) Hartwig F, Dearing BE : Exploratory data analysis. SAGE Publications, London, 1979  
(柳井晴夫・高木広文訳：探索的データ解析の方法。朝倉書店，東京，1981)
- 2) 三宅由子：臨床研究における記述的方法と仮説検証の方法——精神障害の亜型分類に関する北村と笠原らの討論への参加——。精神神経学雑誌 93；690-691, 1991
- 3) 三宅由子：臨床データのまとめかた——研究計画から論文作成まで——。杏林書院，東京，1992
- 4) 高木廣文・三宅由子：よくわかる医療・看護のための統計学入門。メディカ出版，大阪，1991
- 5) Wynder EL, Higgins IT, Harris RE : The wish bias. Journal of Clinical Epidemiology 43 ; 619-621, 1990
- 6) 柳井晴夫，豊川裕之：医学・保健学の例題による統計学。現代数学社，京都，1977