

第6章

質の高い科学的証拠とその利用

柳本 武美

yanagmt@ism.ac.jp

総合研究大学院大学 統計科学専攻 教授、統計数理研究所

プロフィール

専門は生物統計学・理論統計学。生活環境中の化学物質の安全性と有効性の計量評価の研究に従事し、その中で多くの具体的なテーマも扱ってきた。近年は質の高い証拠の集積が重要であることを痛感し、証拠の獲得の枠組についての考察を進めている。

はじめに／この講義での問題設定とポイント

本日の講義の要旨として、下記の内容を案内に記した。

「無心な観測の積み重ねと鋭い洞察による科学研究は、実は誤り易い。質の高い証拠に基づいた厳密な推論が、より確かな科学的理論を構築する。質の高い科学的証拠は、明示的な仮定の下に周到な研究計画を立て、その上で訓練された多くの専門家の営為を通じて得られる。科学的証拠が社会の営みにより支えられるならば、研究の目的は社会のニーズと無縁ではあり得ない。前世紀に展開された科学的認識論の再検証を受けとめた上で、質の高い科学的証拠の獲得と利用の観点から、科学研究と社会との係わりを議論したい」

まず、私にはほとんど自明と思われるけれども、一般にはしばしば看過されている3つの言明を前提として話を進めていきたい。

- 人間は科学的真実そのものでなく、自らが認識した真実に基づいて行動する。
- 自らが事実を確信するに足る証拠と、他人に納得してもらう証拠とは

異なる。

- 司法と行政は社会の公平と利害の調整のために、認識した事実に基づいて判断する。

1. 「質の高い科学的証拠」の意味

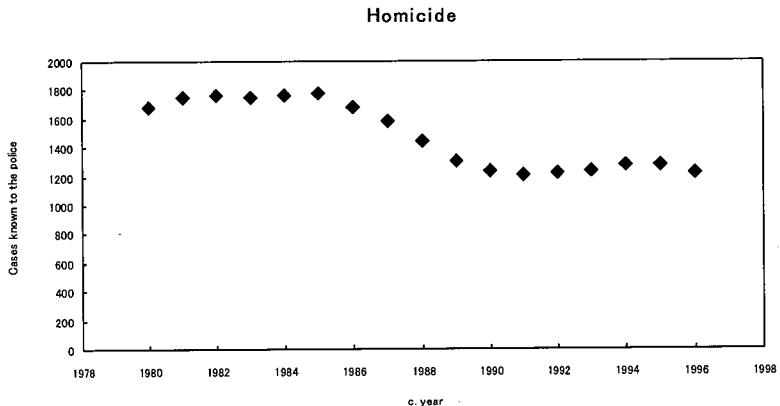
1.1. 統計データの陥りやすい誤解

まず、統計の基礎的な話から始めたい。【図表1】は殺人認知件数の推移を示しており、縦軸は認知件数である(通常、殺人認知件数は発生数であると考えられる)。しかし正確な件数を把握するのはきわめて困難である。そこで、警察の業務統計としての警察が認知した件数を数えているために認知件数と呼ばれる)。

日常的な感覚では、近年殺人は増えているという。しかし統計的にはほぼ一定しており、特に増えているわけではない。これも、統計的な知識を身につけていなければ陥りやすい誤解であろう。

科学の科学たる所以の1つは、場所と時を変えても同じ現象が期待できる

【図表1】 殺人認知件数の推移



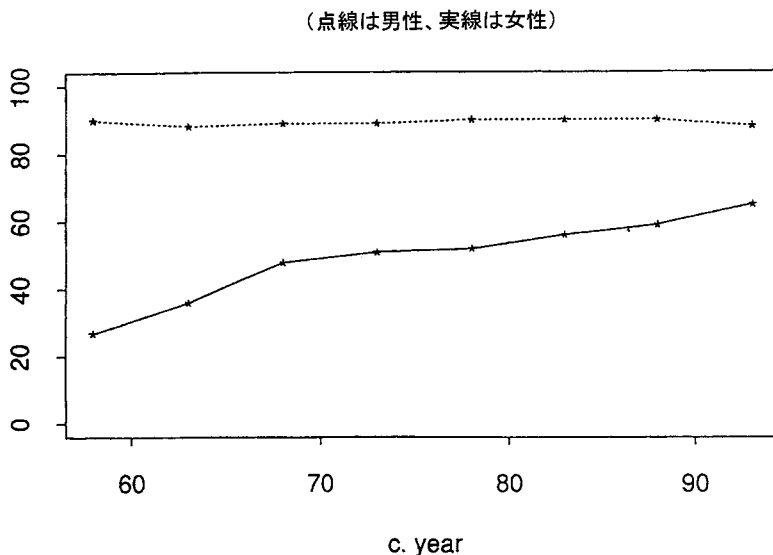
ところにある。そこで、1826年当時のフランスのデータを見てみよう(【図表2】参照)。この表を見ても、殺人件数はあまり変化していない。すなわち、たまたま最近の日本だけの現象ではないことが分かる。

【図表2】 フランスにおける殺人件数の推移^[1]

方法	年別						
	1826年	1827年	1828年	1829年	1830年	1831年	
総数	241	234	227	231	205	266	
銃火器・短銃によるもの	56	64	60	61	57	88	
剣・短刀によるもの	15	7	8	7	12	30	
小刀によるもの	39	40	34	46	44	34	
棒・杖等によるもの	23	28	31	24	12	21	
石によるもの	20	20	21	21	11	9	
打物や刺物・刃物	35	40	42	45	46	49	
絞殺	2	5	2	2	2	4	
投殺・溺殺	6	16	6	1	4	3	
毒殺・撲殺	28	12	21	23	17	26	
焼殺	—	1	—	1	—	—	
不明	17	1	2	—	2	2	

しかし、一方では世の中が変化しないわけではない。【図表3】は統計数理研究所が定期的に行っている調査で、生まれ変わりに関する意識を調べたものである。男から男に生まれ変わりたいとする比率はほぼ一定で90%台を維持している。それに対して、女から女に生まれ変わりたいとする比率は増加傾向にある。このように世の中の現象には、変化しないものと変化するものがある。結局は調査しなければ分からない。日常的な感覚では、変化に対して、より敏感になる傾向があると思う。

【図表3】 生まれ変わりに関する意識の変化



1.2. 「質の高い証拠」と「質の低い証拠」

さて、第一の結論は「完全な証拠はない」ということである。ある命題を示すために完全な証拠が存在すれば、それ以上の証拠は不必要になる。しかし完全な証拠は存在しない。完全な証拠がなければ完全な命題もありえないことになる。自然科学や社会科学においてわれわれが信じている命題は、不完全な証拠で支えられているのである。

したがって、「完全な証拠」ではなく、「より質の高い証拠」と「より質の低い証拠」が存在すると考えるべきだろう。そして、人々のアイデアや努力を傾注して初めて、データの質がより高くなる。すなわち、完全なものはえられないが、努力により相対的によい証拠がえられる、ということだ。一方で、現実には、新聞報道や雑誌紹介記事などに、質の低い証拠に支えられたデータが流通している事実も認識しておく必要がある。新聞や雑誌を見ても気づくことだろう。

1.3. 質の高い情報はコスト高

質の高い情報を得ようとする、大きな労力とコストを必要とする。そういう意味では、現実の社会で必要な質のレベルという概念が当然要求される。それをそのまま論文に当てはめると、ある命題を論文化するに際しては、実用的な観点から言えば、ジャーナルなどに採択されるぎりぎりの質の高さが満たされていないなければならないことになる。さらにレベルの高いジャーナルに発表するためには、より大きな努力と労力を必要とする。

求められる証拠の質と強さは、領域によってさまざま考えられる(【図表4】参照)。数学や論理学では完璧な論理性が求められる。刑事裁判では「合理的な疑いを超えて立証されていること」が必要である。科学的知見では、ある程度の高い証拠が求められる。また薬の許認可や予防対策などの行政判断においては、それよりもう少し緩やかな、しかしある程度確実な証拠が求められる。民事裁判については、考え方により差があるが、ある程度の証拠が求められる場合が多い。意思決定については、企業は行政と異なり自己責任

【図表4】 求められる証拠の質と強さ

異なる分野で求められる証拠の強さ			
分野	証拠の強さ	コメント	判断と決定
i) 数学・論理学	完璧	(Beyond any doubt)	保留可能
ii) 刑事裁判	高度	Beyond a reasonable doubt	決定
iii) 科学的知見	高度	査読、学説	保留可能
iv) 行政判断			
a) (許認可)	確実	開示義務	(保留可能)
b) (予防対策)	確実、優位性		決定
v) 民事裁判	優位性、確実	Preponderance of evidence	決定
vi) 意志決定			
a) (企業)	不必要	商法等の制約	保留可能
b) (個人)	不必要	刑法等の制約	保留可能

で行動するわけだから、それほど強い証拠は必要としない。しかし社会に対する影響の大きさから、商法等の制約を課している。個人の意思決定は他人に迷惑をかけない範囲であれば、根拠がなくても自由に行なうことができる。

このように、求められる証拠はいろいろなレベルで強度が異なり、またそれを支える証拠にもいろいろなレベルがある。

この考え方を、実際の事件に当てはめてみよう。1996年初夏に、堺市で0-157事件が発生した。原因として、ある業者が学校給食用に納入したカイワレ大根が疑われたが、それ以外にも、下記のようないろいろな仮説が立てられた。

- a) 小学校の学校給食が原因である。
- b) 学校給食の中でも中南地区で9、10日、北東地区で8日の献立が関与している。
- c) 上記 b) で共通した食材である特定の野菜が関与している。
- d) 上記 b) で共通した食材である牛乳が関与している。
- e) 特定のグループの犯罪（謀略説）である。

このうち、少なくとも e) 特定のグループの謀略説は、私が知るデータで見ると根拠はまったくなかった。d) の牛乳に関しては、統計学的には疑念が残るが、実際の生産現場とか納入業者が2社あったことから外された。もし c) が成立していれば b) が成立するし、b) が成立していれば a) が成立することになる。そこで結果として、c) までは確実であろうということになって、特定の野菜が原因と判定されたのである。

2. 「質の高い証拠」を求める方法論

2.1. 実証のための哲学の転換

次に、実証のための哲学や科学哲学の転換について触れよう。かつてわれわれが科学的推論の根拠としていたのは、ベーコン、ミル、マッハなどの考え方である。一方、デュエム、ラカトシュ、ポパーなどが、観察の理論的負荷

性、理論の全体性、反証主義などを主張した。

従来の実証の考え方についてなぞってみよう。科学を意識しはじめたのはおそらくギリシア時代からで、アリストテレスの自然学とキリスト教につながり、さらにガリレオ、デカルト、ニュートンらの新しい科学観、すなわち、自然をありのままに観察する考え方が重視されるようになった。なかでも統計学者に大きな影響を与えたのはミルの5つのカノンであり、一致法・差異法・一致差異併用法・共変法・剰余法などが生まれた。しかしこれらは、後になって考えれば、素朴でありすぎたのではないかと思われる。

科学哲学を代表する西田幾多郎は『善の研究』の中で、「毫も思慮分別を加へない、眞に経験其儘の状態をいふのである。例へば、色を見、音を聞く刹那、未だ之が外物の作用であるとか、我が之を感じて居るとかいふやうな考のないのみならず、此色、此音は何であるといふ判断すら加はらない前をいふのである」として、まったく何もない状態で純粹経験としての観察をすることにより、新しい発見があると提唱している。

今から考えれば現実離れしている。しかし私自身も以前はこの考え方をふまえて、予見を持たず純粹無垢な気持ちで自然を観察することが必要だと思いながらデータ解析などを行っていた。また議論する際には、互いに相手が純粹無垢でないことをお互いに批判しあっていたものだった。しかし完全に白紙であることは不可能である。

完全に白紙で観察することが不可能なのは下記の理由による。

- ① 観察を行なうためには、その前提としての理論が必要である (theory-ladenness)。何の先入観もなしに自然を観察できると考えるのは思い違いでしかない。
- ② 観察からある特定の理論を抜き出して、その是非を論じることはできない (holism)。先ほどのミルの5つのカノンのうち差異法について言えば、そのように努力することが必要ではあるが、本当に研究者ができると思えるのは思い違いである。
- ③ 1つの理論によって多くの観察が説明されること、あるいはある仮説が観察と適合することは、理論や仮説が正しいことを強く示唆するわけではない (falsification)。

こうした考え方は、すでにわが国の科学哲学では定着している。この関連の哲学者と著作には、板倉聖宣『科学の形成と論理』、内井惣七『科学哲学入門』、小林道夫『科学哲学』、都城秋穂『科学革命とは何か』、伊勢田哲治『疑似科学と科学の哲学』などがある。

2.2. 質の高い証拠を得るための努力

次に、質の高い証拠を得るための努力が払われている具体例として、臨床試験を取り上げよう。標準的な臨床試験に見られる特徴として、第一に比較を行うこと、第二に対峙する仮説を否定していることがある。すなわち、対峙する命題を否定することによって、自らの仮説の正しさを検証する。第三に、試験の方法、評価法を予め明示する。つまり、データを集めてから解析法を工夫するのではなく、ある仮説に基づいてデータを集めて、収集法に則った解析をする。ここでいう標準的な方法は、日本のみならず世界的にも完全に受け入れられている。最近では、EBM、すなわち医療は信頼できる証拠に基づいて行なわれるべきだとの考え方が主流になっている。

このように、臨床試験では上記の方法をとらなければ質の高い証拠は得られないと考えられているが、多くの科学者には違和感、抵抗感、反感があるようだ。そのことを理解するために、素朴な方法として従来から行なわれていた方法を考えてみる。たとえば、「有能で経験豊かな医師が、典型的な患者に、患者の特性を考慮して最善の治療を施す。そして患者の予後を注意深く観察する。こうした様々な治療経験を積み重ねるうちに、良い治療法が見出される」。この一連の過程において、ある命題を否定する記述はどこにもない。またデータをとるとき、事前にもどのように解析するかも考えず、データをとってから解析方法を考えている。このようにして得られた証拠の質の高さは、それに従事した人(医師)の能力、熱意に依存するしかない。これは自分自身は納得できても、他人に説明するのは非常に難しいロジックである。いかに真面目に取り組んでいると力説しても、その真面目さがどのように担保されているかを判断する根拠に乏しい。

次に、ややモダンだが素朴な方法は次のように示される。すなわち、「臨床

検査を十分に行って診断を確定させるとともに、DNA検査などで患者の特性を把握する。診療記録は詳細に記録して巨大なデータベースを作成する。データマイニング、機械学習、推論エンジンなど最新の情報科学技術を駆使するとともに、スーパーコンピュータを使用して、「データベースから患者優先のテーラーメイド治療法を見つける」。これは現代の生物系の研究者が特に強調している方法論である。しかし、ここにも命題の否定はどこにもない。またあらかじめ、どのような解析をするかにも触れていない。この方法では、ある基準以上の高い質の証拠を得ることはできない。

では現代的な方法とはどのようなものか。それはすなわち、「予め定めた治療の方法と対象疾患に基づいて計画書を作り、計画書に基づいて治療群と対照群の2群に分ける。そして予め定めた方法で治療の効果を判定し、予め定めた方法で解析する。その結果、治療群での効果が対照群と異なるときに効果があると見なす」という考え方である。このように2群に分けて考える考え方の原型は、1930年代に農事試験、品質管理、社会調査などから生まれた。

標準的な臨床試験の初期例としてストレプトマイシンの試験がある。ストレプトマイシンが結核に効くという発見は、医学史上、エポックメイキングな出来事であった。【図表5】は、結核に対して画期的な薬ができたということで行なわれた試験結果である。

【図表5】 BMRC の SM 治療試験 6ヵ月後のX線変化^[2]

(British Medical Research Council, Streptomycin treatment of pulmonary tuberculosis. Br Med J 1948)

	全例 (%)	改善		不変	悪化		死亡
		著明	中軽度		著明	中軽度	
SM 群	55 (100)	28 (51)	10 (18)	2 (4)	5 (9)	6 (11)	4 (7)
対照群	52 (100)	4 (8)	13 (23)	3 (6)	12 (23)	6 (11)	14 (27)

こうした統計データを見慣れていない人には、薬の効き目が大したことはないと思うかもしれない。しかし、これは医師を驚かせたデータであった。今にして思えば、こんな臨床実験を実施してよかったのかというほど、非常にクリアなデータがとれたわけである。

ストレプトマイシン以前には、セファラチンなどさまざまな薬が繰り返し試されていた。多くはしばらく効果がありそうだとおもわれたものの、すぐに効果がなくなるという繰り返しであった。そのためにストレプトマイシンも当初はそれほど期待されていなかった。そうした中で臨床試験が行なわれて、その効果が確認された。

ストレプトマイシンのような薬なら、臨床試験は不要と主張する人がいる。しかしそうではない。日本における治験法の先駆者である砂原茂一は『臨床医学研究序説』の中で、他の薬が開発されていたために耐性菌が出現しなかったが、もしストレプトマイシンだけが孤立して発見されていたら、短期間に薬剤耐性化していただろうと述べている。

ただし、ストレプトマイシンは結核治療のターニングポイントの薬であったが、改めて考えてみると、効く時期、効かない時期があり、それほど安定性はない。つまり結核の中にも、ストレプトマイシン耐性菌と感受菌があり、常に効果があるというわけではない。同様のことは他の例でもある。

一方、当時、抗ヒスタミン療法が風邪に有効であると理解されていた。しかし臨床試験の結果は、【図表 6】の通りである。風邪のための抗ヒスタミン

【図表 6】抗ヒスタミン療法の臨床試験の結果^[3]

	全例	治癒	改善	不変・悪化
Treated	579 (100)	51 (8.8)	307 (53.0)	221 (38.2)
Placebo	577 (100)	49 (8.5)	296 (51.3)	232 (40.2)

Hill (1950), Patulin Clinical Trial Committee (1944)
Double-blind study

療法に効果があるとは見てとれない。

2.3. 質の高い証拠を得るための技法

ここで、質の高い証拠を得るための考え方、手続きについて改めて説明しておこう。

1) 厳密な比較を行う

治療を施した群だけではなく、これと比較する対照群を設けて厳密に比較する。よくデータと仮説を比べるというような言い方をしている人がいるが、それはまったく不可能ではないにせよ、きわめて困難である。なぜかと言えば、データと仮説のようにまったく属性の異なるものを比較しようとするからである。当然ながら、仮説と仮説を比べるほうが楽だ。そのことは、1つだけの群で推論する不十分さを物語っている。1つだけの仮説では、データと仮説を比べるしかない。それに対して、2つの仮説をもっていれば、データを通じて2つの仮説を比較検証でき推論が容易になる。

2つの群を比較する際には、検査の方法、判断基準などできるだけ両者を同じように扱うことが重要である。たとえば、公平な治療効果の判定のための1つの工夫が双盲法（二重盲検法）である。また割り付けで最も重要なのは無作為割り付けである。ところで実薬群の人が薬を飲まなかったり、対照群の人がなんらかの理由で薬を飲んでしまったりなど、約束が守られていない場合がある。こうした場合でも、形式的に割り付けた方の群の患者として解析する手法がとられる。一見奇妙な気もするが、厳密な比較のためには割り付けを優先させる。

2) 仮説を否定（棄却）する

たとえば、薬が効いているかどうかを検証する場合、「薬が効いている」ことを立証するのではなく、「薬が効いていない」という仮説を否定することによって薬の有効性を立証しようというものだ。

ここでは、否定と肯定の非対称性に注意する必要がある。ある仮説が否定されるとき、その仮説が誤っているか、あるいはその試験の実施に伴う仮定

が誤っているかのどちらかである。しかしある仮説が否定されないときは、非常に多くの原因が考えられるため、その仮説が正しいとする証拠としては弱い。たとえば「適切に症状を確かめた患者に対して上手に使うと良く効く薬」という言い方をすれば、ほとんどの薬に当てはまる。医師が下手だったとか、患者が症状について本当のことを言わなかった、安静を守らなかったなど、さまざまな言い訳が可能であり、こういう命題を否定するのは困難だ。

だから、このような命題を挙げるのであれば、どのようにして症状を確かめるのかを明記する必要がある。また、上手に使うという基準を設定しておく必要もある。そうしないと、言い訳がたくさんあり、すべてが肯定できることになってしまう。そのことをポパーは、「多くの観察と矛盾しない理論は空虚である」と言明して人々を驚かせた。しかし改めて考えてみると、ごく当たり前のことを主張しているわけだ。

3) 仮説と手続きを事前に設定する

仮説と手続きを事前に設定することが必要である。臨床試験の計画に際しては、当該臨床試験以前の試験の結果から多くの知識があり、治療法、治療効果の判定方法、患者の選択規準なども予めプロトコールに記載しておき、予め検証する仮説と試験の方法を設定しておく必要がある。ところがほとんどの研究者は、データを入手してから新しい知識を発見したり、自己の仮説に適合させることに熱心である。

さて、これまで臨床試験を中心に述べてきたが、何も特別な例ではない。そこで、臨床試験と刑事裁判の類似性を見てみよう(【図表7】参照)。刑事裁判では、公共の名のもとに個人に刑事罰を課すわけだから、そのような行為をするためには強い合理的な証拠がなければならない。それは国に委任された検察が行なうものである。刑法の場合は、命題をあらかじめ設定しておかなければならないし、刑法にないものは刑事罰を課すことができない。また刑法の解釈は厳密に行なわれるべきである。したがって刑法制定以前の行為は犯罪とは見なされず、遡及的加罰は禁止されている。すなわち、ある行為が刑法で犯罪と明記されて初めて犯罪となるのである。また命題をあらかじめ記述しておく必要もあり、いったん起訴状に書かれた内容は変えること

【図表7】 臨床試験と刑事裁判の類似

	臨床試験	刑事裁判
命題の設定	仮説を事前に明示	罪の法定主義 遡及的加罰の禁止
命題の記述	プロトコール	刑法・訴追状
取り扱いの公平さ	対照群の設定 無作為化	弁護士 主尋問と反対尋問
判断の方法	統計解析法の明示	刑事訴追法

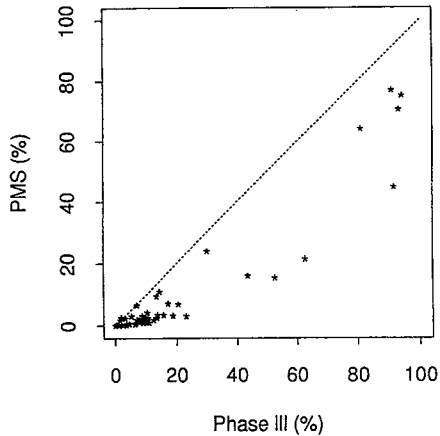
はできない。このように刑事裁判の世界では、高い質の証拠が求められている。

では、ビジネスの世界はどうか。セブン-イレブン創始者である鈴木敏文は、POSデータ・システムの導入者でもある。彼は、心理学と統計学を駆使してビジネスを行なっていると主張している。そして、POSデータを集めて売れ筋商品を調べるようなことをしていけないという。そうではなく、別に仮説を立てておいて、仮説を利用して発注量を決めて、仮説を検証するためにPOSデータを活用すべきだと指摘する。そうでなければ、POSデータを誤った方向に解釈してしまうとしている。

先に解析の方法をあらかじめ決めておかなければならないと指摘したが、その考え方は、たとえば高エネルギー研究所のBファクトリーでのデータ解析にも採用されている。扱っているデータは物理学的データであるから、人々の価値判断や感情的要素は比較的入りにくいと考えられている。しかし、データのチェックはオープン前に行い、データを固定したオープン後はそのチェックが正しいものとして解析する。データ変更しないことが大切である。常識的に判断すれば、解析した後でおかしいと思ったデータは修正すればいいと考えるかもしれないが、その分得られる証拠の質は低くなる。どちらかという物物理系の方は、データの最適変換のような技法を好む傾向があるので、Bファクトリーの研究者の考え方は、私には非常に刺激的であった。

またきちんとしたデータをとれば、いろいろなことが見えてくる。たとえ

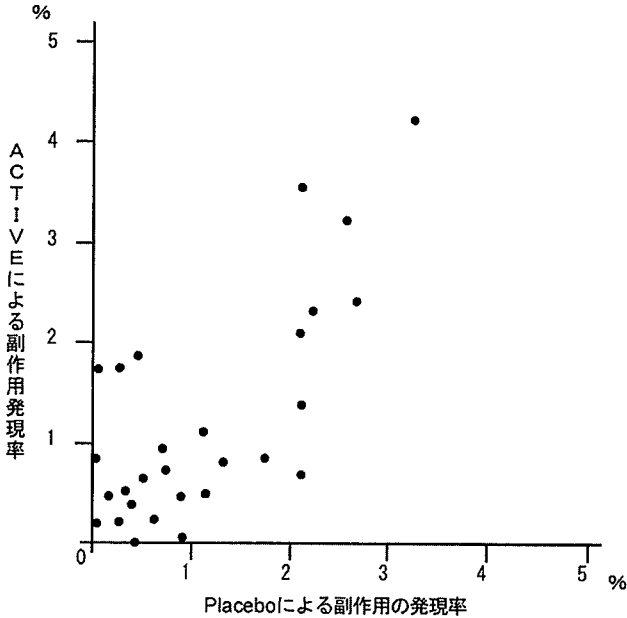
【図表8】 第3相試験とPMSにおける副作用の出現率



ば、【図表8】を見よう。横軸のフェーズ3とは、臨床試験の段階（フェーズ）を指しており、縦軸のPMS (Post Marketing Severances)とは、薬を売り出した後の調査結果である。図中の1つの点が1つの薬剤を表している。第3相の副作用の出現率（報告率）のほうが、PMSでの出現率が高い。これは常識的に考えても納得できる。それは臨床試験の時のほうが、より厳しく観察すると考えられるからである。しかもこの傾向は、副作用の程度が小さいときにより顕著に現れている。

次に【図表9】を例にしよう。図では、実薬と擬似薬（プラセボ）による副作用の出現率の差を比べている。実薬と擬似薬とも出現率はかなり比例していることが観察される。これは狭義に考えれば、薬理的には奇妙なデータである。もちろん擬似薬の作り方に失敗したという理由は考えられるが、たぶんそうではないだろう。そうではなく、もし実薬が本当に副作用が出るかもしれない薬であったとしたら、それを看過して副作用を報告しないことを臨床医がおそれたという仮説のほうが、はるかに信憑性がある。このようなデータが得られているのは、よく管理されて臨床試験が行なわれているからだと思える。

【図表9】 実薬と擬似薬による副作用の出現率比較^[4]



2.4. 専門家の役割の重要性

さて、「社会と科学」の関係においては、専門家の役割の位置づけが大きな意味をもつ。実際には、専門家の重要性が高まってきているのに、これまでわれわれは専門家の重要性についての認識が不十分だったのではないだろうか。たとえば、食品の安全性評価の専門家は毒性学の専門家であると考えてしまいがちだが、食品についての知識やリスク評価法の知識を持つ人が専門家と考えるべきである。そういう意味では、実施疫学者が重要な専門家と見なせよう。0-157 事件の場合も、アメリカには多くの食中毒専門家がいたが、日本では食中毒の専門家はいなかった。したがって実際に担当したのは食中毒専門家ではなく、輸入食品の管理をしていた獣医や、DNA解析をしてい

た細菌学者などであった。しかし本当は疫学者がもつとも近い専門家であったと思う。細菌を探すより、実際に目の前にいる患者の症状を優先するべきであった。

他の例で言えば、臨床試験の場合、生物統計や治療評価学の専門家が参加する必要があるし、また各病院内治験センターの専門担当者も必要である。実際、専門家のレベルが高くなることによって、証拠の質も高まる方向に向かうことが期待される。また教育試験についても、従来のように1人の教師が担当するのではなく、試験問題を作る人、結果を解析する人、試験問題を編集する人など、さまざまな専門家が集まることなしには、良い試験を構築できない。

国際的な取り決めのための会議などの場でよく指摘されるのは、日本には周辺領域の専門家はいるが、本当の専門家はいないということだ。諸外国と比較して、日本人は語学が苦手だと言われる。しかし、より困ったことは、高度な専門性に大きな問題があり、日本が交渉事で弱いという事態を引き起こしていることだ。その結果、国益を損ねている。その大きな理由は、大学、研究所などのキャリアパスの問題にあると思う。

2.5. 質の高い証拠の必要性と限界

最後に、質の高い証拠の必要性と、またその一方でその限界についてまとめておこう。

証拠の過剰品質には十分な配慮が必要である。流通業の基本は「安く仕入れて、高く売る」ことにある。同様に、必要最小限の質の証拠が、現実の社会では歓迎される。しかしながら、もしも提出する知見が重要であるならば、その知見は質の高い証拠によって検証される必要がある。実際に質の高い証拠が求められる分野は、すでに見てきたように、治療法の評価、科学上重要な言明、司法(特に刑事裁判)、行政判断など数多い。

最後に、特殊な領域として、研究論文に求められる証拠の質について述べておこう。もちろんジャーナルの質により、有名ジャーナルに掲載されれば高く評価されるが、それほどレベルの高くないジャーナルに投稿して掲載さ

れても評価は高くない。そのとき、そのときの勝負ということになれば、目の前の勝負に勝つために、論文として採択される最低限度の証拠を提示しなければならない。また、たぶんどの方野でもそうだと思うが、証拠の質を高く見せる技法も必要である。レフェリーをすればよく分かるが、中身より結果の見栄えを評価する傾向が強いことは否めない。

研究者として生きようとする場合、夢だけを追うと生きていけないのも現実である。しかし一方で、夢を追わなくては生きる価値がないのではなからうか。

<注釈>

特に断わらない引用は、演者による作成あるいは公的資料により演者が作成した。

- [1] 小杉 肇 (1969). 統計学史通論 恒生社厚生閣
- [2] 砂原茂一 (1987). 臨床医学研究序説 医学書院
- [3] Hill, A. B. (1950). Clinical trial of anti-histamic drugs in prevention and treatment of the common cold, *Br. Med. J.*, 24, 425-429.
- [4] 清水直容, 藤田利治 (1989). 副作用研究—現状と展望—臨床評価 17 巻

<質疑応答>

—— 昔の科学は観察主体、現在の科学は実験、もしくは理論に基づく観察主体だと思ふ。観察の場合、新しい存在の発見のときなど、対峙する仮説は存在しないのではないか。昔の観察においては、素朴な科学論が成立したが、それが現在の科学観の中に持ち込まれたとき、科学論と実際の科学との間に齟齬が生じたと思ふ。科学論の話をするときには、観察系、実験系の区別をする必要があるのではないか。

柳本 観察系の科学があるという認識がないが、観察の蓄積という意味ではあ

るかもしれない。そのときには明確な仮説はないだろう。それまでの理論の上での解釈に基づく観察であって、それより新しいものは理解しがたいだろう。たとえば新しい鉱石が発見されたとして、固さ、組成などの点でそれまで自分たちが理解している以外の特性が存在すれば、そのことによって新しさを理解する。本当に理解を超えるものがあつたら、ある種の特性を書くだけで、説明しにくいものになるだろう。

—— 新しいものの発見は常に説明しにくい、それもやはり科学の1つではないか。

柳本 それほど意外性のある発見とはどういうものか。それまでの自分たちの知識のフレームで理解できないほど、大きな発見はないのではないか。

—— ことの大小が重要性をもつとは思えない。たとえば、カミナリイカは実験室では孵化可能だが、実際に海底で採取された卵には砂がまぶされていた。それはまったく予見できなかったことなので、仮説も存在しない。

柳本 砂をかけるか、かけないかについて考えていなかったというのは、かけないと考えていたのではなくて、何も考えていなかった、というのが事実だろう。

—— 見た事実は存在するだろう。

柳本 見た事実がすべてではない。認識した事実のうち、関心のあることを説明したわけで、すべてを説明したわけではない。

—— 0-157、狂牛病、鳥インフルエンザなどの国際的な問題に対しては仮説の設定自体が難しいし、空振りの可能性もある。

柳本 結果を何に使うのかによる。科学的な認識をするというレベルで考えたら、とうていそういう資料は得られないから、その中で生きていくしかない。しかし、もしもっと質の低い証拠でアクションがとれるとすれば、それでよいはずだ。あらかじめ計画していなくても、そのように見なし

で解析し、そのズレについては、その分証拠の質が低いと考えればよい。現実に裁判の世界でも、刑事裁判と民事裁判があり、違うものとして扱っている。他の分野でも、そのようなかたちで、目的に応じて必要なエビデンスの質を変えている。だから関係者がどのように判断するか、だ。その例として、現在、質の低いエビデンスで人々が行動を起こしているのはテロだ。その是非はともかく、非常に質の低い証拠であっても、より強いアクションを起こしている。今日の話の眼目は、人々は何を目的として証拠を集め、アクションを起こすかということだ。ジャーナルも評価されるレベルがまったく違う。そのとき、そのときに応じて対応しているので、そのときいろいろな齟齬が出てくるのはしかたない。

- レフェリーする立場として、自分の考えとは違うが、統計的な手法を使って正しく書かれている。しかし、社会的な意味もよく分からないし、それによってジャーナルの評価が高まるとも思えない。そういう場合、レフェリーはどうあるべきか。

柳本 それは、エディターの仕事だと思うが、証拠の信頼性は結果ではなく、手続きにあることは確実だ。どのように作成したかによって品質が決まる。しかしそれをジャーナルがどのように扱うかについては、エディターが全権をもつというスタイルだから、彼らの器量による。