



—資料紹介—

## ランダム化比較試験の歴史 (3)

水谷民雄

### 3. 比較試験：1948年の分水嶺（全訳）

Richard Doll. Controlled trials: the 1948 watershed.

Brit Med J 1998;317:1217-1220.

<http://www.bmj.com/cgi/content/full/317/7167/1217>

#### 1946年以前の臨床試験

科学上の重要な進歩は出し抜けに起こることはめったになく、振り返ってみると、多年にわたって積み上げられてきたプロセスの集大成であることがわかる。このことは、1948年に始

めて報告された臨床試験への新しい手法の導入についても間違いなく当てはまる。この手法は、20世紀後半の臨床医学の進歩にとって重要な役割を果たした。私が医師資格を取得した1937年頃は、新しい治療法はほとんどいつもA教授、あるいはある指導的な教育病院の顧問医師の下での少数の(50例を上回ることはまれな)患者群における結果が、B教授(あるいは他の顧問医師)によって報告された結果、あるいは自らが以前に報告した結果に比べて良好であったという理由で導入されていた。このような状況下では臨床アウトカムのばらつき、偶然のばらつき、患者の選択における無意識(意識的な場合はもちろん)のばらつきなどが、得られた結果に一見重要と思われる違いをもたらした。その結果、多くの新治療法が互いに競り合う状況となっていた。消化性潰瘍の治療法はおそらく他の慢性疾患に比べてその利点を主張しやすかったものと思われる。そのため1948年に私が消化性潰瘍の研究を始めたとき、アルファベット各文字で始まる治療法をすぐにリストアップすることができた。標準的治療法自体も一つの教科書から別の教科書へと十分な評価抜きに書き写されていく傾向があった。

何人かの臨床家はこの状況を満足すべきものでないと感じ、同時的な対照群を設定することを要求した。このとき提案されたのは、通常、患者を一人おきに、比較される2つの方法の一方または他方で処置することであった。Fibigerはこの方法を1898年にジフテリアの治療に対する抗血清の臨床試験に用いたと記録されている<sup>1)</sup>。しかし、この方法は急速には広まらなかった。この方法は1930年代までには英国医学研究協議会(Medical Research Council)によって恒常的に用いられた。おそらくはその統計委員会の主宰者M Greenwoodの助言によるものであろう。それは1934年にはクループ性肺炎の血清療法を評価するために<sup>2)</sup>、また1944年にはかぜに対するパツリン(patulin)の試験のために<sup>3)</sup>用いられた。この方法は、1946年にWilsonらによって同じ患者グループにおいて2つの比較を可能にする要因分析へと拡張された。この方法は急性肝炎の治療において低脂肪食とジシステイン添加食の効果を試験するために4つのグループを必要とした<sup>4)</sup>。パツリンはかぜに対して効果がなく、低脂肪食は急性肝炎に対して効果がなかったが、ジシステインは急性肝炎の期間を数日短縮するよう見受けられた。

### ランダム化法の導入

交互割付法には一つの重要な欠陥があった。それは研究者に次の患者がどちらの処置を受けることになっているかが分かり、ある患者をその試験に参加させるのが適当かどうかを決める際にバイアスがかかる可能性がある(実際にもしばしばそうなった)ことである。

A, Bのラベルをつけた同じ形状の薬物を提供することによって患者への処置に関して研究者を目隠ししても、この難点は完全には解消できない。それは、研究者が速やかに一方(の処置)が他方に比べて優れていると感じ取り、次の患者の適格性を判断する際にバイアスがかかるようになるためである。

このようなバイアスを避けるための一つの方法は、患者を、どちらのグループがどちらの処置を受けるかわからない時点で、相似した2つのグループに分け、最終段階でコインを投げて、一方のグループに一方の処置を割り付け、他方のグループには別の処置を割り付けることであ

ろう。この方法は 1662 年に医化学者 van Helmont によって提案された。このとき彼は当時の大学研究者に対して、理論に基づく彼らの治療法と、経験に基づく自らの治療法を比較したいと次のように申し入れた。「病院などから熱病や胸膜炎にかかっている貧しい患者を 200~500 人選び、それを 2 群に分けたうえでくじを引き、一方のグループを私への割り当て、他方のグループをあなた方への割り当てとしよう。そしてどちらがより多くの葬式を出すことになるかを見ることにしよう」<sup>5)</sup>。残念ながらこの挑戦は受け入れられなかった。しかしこの手法は 260 年後に Amberson ら<sup>6)</sup>によって肺結核の治療におけるチオ硫酸金ナトリウムの効果を評価するために実行に移された。Amberson らは 24 人の患者を 12 人ずつに 2 分した。このとき両群の患者は 2 人一組としてそれぞれの条件をマッチさせた。その後コインを投げて、どの群がどの処置を受けるかを決めた。

Armitage<sup>5)</sup>が指摘しているように、この手法は、患者の症例を正確にマッチングさせることが事実上不可能であるだけでなく、付随するランダム・エラーを評価する方法がないという難点がある。この 2 つの難点は、個々の患者をランダム化することによって、一定の量的限度内で、克服することができる。この手法は 1926 年 Fisher<sup>7)</sup>によって報告された農業実験において用いられた（土地の区画がランダムに割り振られた）。B Hill は、1937 年に出版した医統計学の原理に関する一連の論文で、この手法を臨床医学で用いることが望ましいと認めている<sup>8)</sup>。しかし当初、彼は個々の患者をランダム化することを推奨せず、比較すべき 2 つの処置を一人おきの患者に割り振ることを提案した。それは、1990 年に彼自身が述べているように、処置をランダム化することに言及すると、医師たちがびっくりして同時的対照群を用いなくなると考えたためである<sup>9)</sup>。1946 年になって、Hill は期が熟したと考え、個々の患者をランダムに割り振ることを推奨し、この提案は医学研究者の間に速やかに受け入れられていった<sup>10,11)</sup>。彼がこの手法を提案したのは、これによってランダム・エラーの評価が可能になるといった理由からではなく（Fisher は主にこの理由からこの方法を提唱した）、この方法によって選択バイアスが回避できるという実際的な理由からであった。処置が個々の患者にランダムに割り振られるようになった最初の臨床試験は、肺結核のためのストレプトマイシンの試験<sup>11)</sup>ではなく、百日咳ワクチンの効果を調べるために計画された試験<sup>10)</sup>であった。ストレプトマイシンの試験は D'Arcy Hart と Daniel によって計画され 1946 年 9 月、百日咳ワクチンの試験より数か月遅れてスタートした。しかしこの結果は、百日咳ワクチン試験のそれより 3 年早く、1948 年に報告された。このことにより、ストレプトマイシンの臨床試験は、確かに報告されたのは最初であったが、最初のランダム化比較試験であるという事実と不相応な評判を得ることとなった。これらの試験では、被験者への処置を、効果判定者に対して盲検化すること、そして実行可能な場合は被験者へも、できる限り盲検化することへの努力が払われた。いずれの場合にも倫理的な考察が大きな役割を演じた。

### 倫理的な考察

当時は意見を求めることのできる倫理委員会のようなものは存在せず、医学研究協議会その他の信頼できる機関によって制定された倫理規範もなかった。ヒポクラテスの誓いは医療倫理

を始めて明示したものであり、新たに資格を得た医師はこれの宣誓を求められた。Hill は医師ではなかったが、医療というものの性格を深く理解していたので、彼の王立内科医科大学 (Royal College of Physicians) における医療倫理と比較臨床試験に関する講義は敬意と賞賛とともに受け入れられた<sup>11)</sup>。

ストレプトマイシンの試験で最初に直面したのは、動物実験で有効であり、いくつかの報告では臨床試験でも有望な結果を得ている薬を一部の患者には与えないことが倫理的かどうかという問題であった。しかし、英国にはごく少量の薬しかなく、外国から買うこともできなかった。限られた量の薬を以前なら間違いなく致命的であった粟粒結核と結核性髄膜炎の患者の治療に用いることは合意された。残ったストレプトマイシンでは他の重症結核患者のごく一部を治療するのが精一杯であった。医学研究協議会・ストレプトマイシン臨床試験委員会は「もしも治療効果の迅速で効果的な解明が期待できる厳密に管理された臨床試験を計画する機会を逃してしまうとしたら、それは非倫理的なことである」との合意に達した<sup>12)</sup>。したがって、「一部の患者には薬を与えないことが倫理的に正当化できるのか」という問いに対する回答はためらいなく「イエス」ということになった。

委員会が考慮した他の 2 つの問題は、「関係する医師は試験計画を変更することができるのか」「対照患者に外見的に類似したプラセボを投与する必要があるのか」ということであった。合意されたのは、医師は常に「患者の健康を回復するうえで必要であると信じるすべてのことをなすべきである」ということであった<sup>12)</sup>。これは、もしいずれかの患者に人工気胸 (ストレプトマイシン導入以前の唯一の肺結核の特異的治療法) が有益であると思われたら、それがストレプトマイシン群と対照群のバランスを壊すかどうかにかかわらず、この治療を行うべきであることを意味する。プラセボは 4 か月間にわたり 1 日 4 回の筋肉注射を必要とするので、患者のためを考えてそれを使用しないことにされた。プラセボを使用しなくても治療に対する反応は客観的に評価することが可能である (結核のような重篤な病気には心理的なファクターはほとんど影響しないと考えられる) ので、「研究の正確さを期すために常識を窓から投げ捨てる必要はなかった」<sup>12)</sup>。

委員会はインフォームド・コンセントを得ることの必要性について検討しなかった。これに優越する論点は患者の幸福であった。Hill は、インフォームド・コンセントを得る際に患者の抱えているリスクについて怖くなるような説明をする必要があるのに、形式的な同意を得るよう強制することを強く批判した。「医師は有効性と安全性が提示されている新薬を使用する前にいつでも患者の同意を求めらるだろうか。もしこの答えがノーだとしたら、医師がその薬の効果を以前のオーソドックスな治療法のそれと比較し得る方法で試験しようと決めるとき、インフォームド・コンセントの必要性を生み出すのはどの手順なのか」<sup>12)</sup>。

今日なら彼の問いかけに対する答えは異なったものになるかもしれない。しかし、彼が関心を抱いた「臨床試験の倫理基準と通常の医療における倫理基準は別のものであってはならない」という原理は今なお妥当なものであり、われわれが患者に接する際の主要な決定要因である<sup>13)</sup>。Hill の意見によれば、患者の利益に反するときにインフォームド・コンセントを求めることは非倫理的であり、今日一般に同意されているように適当な独立した委員会の承認を得

ることを条件に、インフォームド・コンセントを要求しないようにすべきである。

百日咳予防の臨床試験については状況が違っていた。6～18 か月の子どもの両親が子どもを試験に参加させてほしいとの依頼を受け、研究について説明したパンフレットをもらった。これには予防接種の半数は百日咳には無効であるが、抗カタル作用はあるだろうとの情報が記されていた。同意書が提出されるまではどの子どもも試験に加えられなかった。この条件は、研究報告書の謝辞でも「多くの親たちが、自分の子どもが必ずしも百日咳ワクチンの投与を受けけるわけではないことを十分承知したうえで、この試験に参加することに同意した」と強調されている<sup>9)</sup>。

### 臨床の場での抵抗

ランダム化試験法が新薬承認機関へ提出される試験報告の必須要件となるまで、その普及は緩慢であり、抵抗を免れなかった。最も一般的な反対意見は T Lewis<sup>14)</sup>の「試験的治療の統計学的手法（彼はランダム化試験法をこのように呼んだ）」に対する批判の線にそったものであった。彼はロンドンのユニバーシティー・カレッジ病院（University College Hospital）臨床研究部門の主任教授であり、1930年代の英国における臨床研究の第一人者であった。彼は1945年、ランダム化比較試験法が導入される以前に、亡くなったが、彼の批判がどのようなものになったかは想像がつく。Lewisは「急性疾患の試験的治療を行うときは、できるだけ相似した2群の患者を、一方の群には治療薬を投与し、他方にはこれを与えない点を除けば、厳密に同じ方法で、また同時に処置すべきである」と考えていた。しかし彼は「試験的治療の統計学的手法（ランダム化試験法）は一時しのぎの便法に過ぎず、そこからは直接的な進歩は何も生まれないことを認識する必要がある。なぜなら症例を集団として扱う研究では、（治療による）利益を受ける症例とそうでない症例を区分することができず、ある症例においてその治療が有効かどうかを判断する基準を決めることができないから」とも述べている<sup>14)</sup>。

Lewisの反対論は、ランダム化法が導入されたのち数年間は何回となく繰り返して唱えられた。これに対して Hill は、「治療に反応する患者とそうでない患者を区別する基準が明らかになれば、われわれはそれを臨床試験に組み入れるだろう」（AB Hill, 私信）と答えていた。Hillの反論に対して特に問題となるような反応はなく、ランダム化法は、もし臨床医が実用上十分明確に患者の層（訳注：同様の特性をもつ患者のサブグループ）を限定することができるなら、その層の範囲内での妥当性が認められるようになった。

### 近年における発展

初期のランダム化比較試験は、その規模が小さすぎて穏和な効果を検出する可能性が少ないという正当な理由で批判を受ける余地があった。小規模な臨床試験は効果が大きいときは有効であるが、これはまれなことである。また、1950～60年代に A Jones や著者ら<sup>15)</sup>が行った消化性潰瘍の治療に関する一連の臨床試験のように、効果は穏和でもアウトカムが定量的に測定できる場合（これらの試験では一定期限内の潰瘍サイズの減少率を記録した）にも、ランダム化試験は有効である。しかし、アウトカムが定量的というより定性的な場合は、患者の数が十

分大きくないと、穏和ではあるが重要な効果がしばしば見落とされるであろう。例えば、心筋梗塞、脳卒中、がんなどの治療に関する初期の臨床試験の多くは、今日の基準からすると十分な規模ではなく、結果的に有効性が認められないという誤った結論を導いた。Hill らは何よりも原理が受け入れられることに関心があり、何千人という患者の臨床試験を求められたら自己破滅していたであろう。また、自分の理論を進んで冷厳な科学的吟味にさらそうとする医師、ましてや外科医はきわめてまれであった。当初から多施設による臨床試験が組織され（ストレプトマイシンの臨床試験には 6 施設が参加）、副腎皮質ホルモンやコルチゾンの潰瘍性大腸炎に対する臨床試験には数施設が参加し、また医学研究協議会による各種の白血病に関する臨床試験<sup>16)</sup>にはさらに多くの施設が参加した。このようにランダム化比較試験が通常的手法として受け入れられるまでには長い年月を要した。こうしてようやく画期的な心筋梗塞の治療についての国際的研究 (ISIS) を組織することができるようになった。この研究は数百の施設とランダムに割り付けられた数万人の患者の参加を得て、ありふれた疾病の治療における穏和な改善効果を明らかにした<sup>17)</sup>。

仮に Hill がいなくても、ランダム化法は遅かれ早かれ、そして米国ではおそらく Rustein によって、導入されたことであろう。Rustein は Hill と共同してアングロ・アメリカンでの急性リュウマチ熱治療における副腎皮質ホルモン、コルチゾン、アスピリンの臨床試験を計画した<sup>18)</sup>。しかし、医療の繊細さと医療倫理に対する Hill の理解と、デザインの簡潔さとプレゼンテーションの明快さに対する Hill の配慮がなかったとしたら、ランダム化比較試験法の普及はもっと遅れたであろう。現代の論文執筆者はこのことを忘れないで欲しい。

## 引用文献

- 1) Dodds E. C., Goldberg L., Lawson W., Robinson R. : Estrogenic activity of certain synthetic compounds. *Nature* 141, 247-248 (1938).
- 1) Fibiger J. Om serumbehandling af Difteri. *Hospitalstidende* 1898; 6: 309-325.
- 2) Medical Research Council Therapeutic Trials Committee. The serum treatment of lobar pneumonia. *BMJ* 1934; i: 241-245.
- 3) Medical Research Council Patulin Trials Committee. Clinical trial of patulin in the common cold. *Lancet* 1944; ii: 373-374.
- 4) Wilson C, Pollock MR, Harris AD. Diet in the treatment of infectious hepatitis. *Lancet* 1946; i: 881-883.
- 5) Armitage P. The role of randomisation in clinical trials. *Stat Med* 1982; i: 345-352.
- 6) Amberson Jr JB, McMahon BL, Pinner M. A clinical trial of sanocrysin in pulmonary tuberculosis. *Amer Rev Tuberc* 1931; 24: 401-435.
- 7) Fisher RA. The arrangement of field experiments. *Journal of the Ministry of Agriculture* 1926; 33: 503-513.

- 8) Hill AB. Principles of medical statistics. London: Lancet , 1937.
- 9) Hill AB. Memories of the British streptomycin trial in tuberculosis. Control Clin Trials 1990; 11: 77-79.
- 10) Medical Research Council Whooping-Cough Immunization Committee. The prevention of whooping-cough by vaccination. BMJ 1951; i: 1463-1471.
- 11) Medical Research Council Streptomycin in Tuberculosis Trials Committee. Streptomycin treatment for pulmonary tuberculosis. BMJ 1948; ii: 769-782.
- 12) Hill AB. Medical ethics and controlled trials. BMJ 1963; i: 1043.
- 13) Collins R, Doll R, Peto R. Ethics of clinical trials. In: CJ Williams ed. Introducing new treatments for cancer: practical, ethical, and legal problems. Chichester: John Wiley , 1992.
- 14) Lewis T. Clinical science illustrated by personal experiences. London: Shaw and Sons, 1934:178-183.
- 15) Doll R. Medical treatment of gastric ulcer. Scott Med J 1964; 9: 183-196.
- 16) Medical Research Council Working Party on Leukaemia in Childhood. Improvement in treatment for children with acute lymphoblastic leukaemia. Lancet 1986; i: 408-411.
- 17) Second International Study of Infarct Survival Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. Lancet 1988; ii: 349-360.
- 18) Rheumatic Fever Working Party of the Medical Research Council of Great Britain, Subcommittee of Principal Investigators of the American Council on Rheumatic Fever and Congenital Heart Disease, American Heart Association. Treatment of acute rheumatic fever in children: a cooperative clinical trial of ACTH, cortisone, and aspirin. BMJ 1955; i: 555-574.

[完]