

臨床からの問い

—「統計学の哲学」序説—

出口 康 夫

1987年・初夏・ボストン

1987年の初夏のある日、一台の救急車が、アメリカ東海岸はボストンのロングウッドアベニューに駆け込んできた。あたりは、ハーバード大学医学部やその関連病院の建物が立ち並ぶ一角である。救急車は、ボストン小児病院 (Boston Children's Hospital)、通称「チルドレンズ」の中央棟の前で停まった。チルドレンズは、ハーバード大学医学部の付属病院の一つで、小児科の専門病院としては全米屈指の医療水準を誇る、ニューイングランドの中核的な先端医療施設の一つである¹⁾。救急車から搬出されたのは、前日生まれたばかりの新生児。呼吸困難に陥っており、下半身にチアノーゼが見られる。新生児肺高血圧症 (Persistent Pulmonary Hypertension in Newborn; ^{ピー・ピー・エッチ・エヌ} P P H N) の典型的な症状である。PPHNとは、新生児500人から700人に1例の割合で発生する症状で、通常は分娩とともに下がるはずの肺血管の血圧が下がらず、出産後も大部分の血液が肺へと流れ込まないままになっている状態をいう。全ての血液が肺を循環しない限り、肺呼吸は始まらない。つ

1) その周辺の様子も含めて、ボストン小児病院については、同病院のホームページ <http://www.childrenshospital.org/> を参照。また以下で取り上げるハーバード大学における ECMO 治験の詳細は、その治験を計画・実施した研究者たちによるつぎの諸論文に主として依拠している。O'Rourke, P.P., et al., Extracorporeal Membrane Oxygenation and Conventional Medical Therapy in Neonates with Persistent Pulmonary Hypertension of the Newborn: A Prospective Randomized Study, *Pediatrics*, 1989, vol.84, No.6, pp.957-963. Ware, J.H., Investigating Therapies of Potentially Great Benefit: ECMO, *Statistical Science*, 1989, vol.4, No. 4, pp.298-306. Ware, J.H., Rejoinder, *Statistical Science*, 1989, vol. 4, No. 4, pp.337-340. ちなみに、1987年初夏という時期は、O'Rourke et al., 1989, pp.960, 962. Ware, 1989, p.301 から割り出したものである。

まり PPHN は、胎児期の胎盤呼吸から肺呼吸への正常な移行が起こらない深刻な病気なのである²⁾。赤ん坊は、ただちに、チルドレンズの集中治療室へと運び込まれ

た。集中治療室での検査の結果を過去の患者のデータと照らし合わせたところ、この新生児の12時間以内の死亡率は約85%と出た³⁾。PPHN の中でも特に重篤な症状である。

PPHN には、人工呼吸器と血管拡張剤などの投薬を併用した治療が、医療の現場で試みられてきた。それに対して、血液を循環させるポンプとガス交換を行う高分子膜を備えた一種の人工心肺装置である体外式膜型人工肺 (Extra-corporeal Membrane Oxygenation; ECMO, 左図参照) を PPHN に用いる試みも1972年から

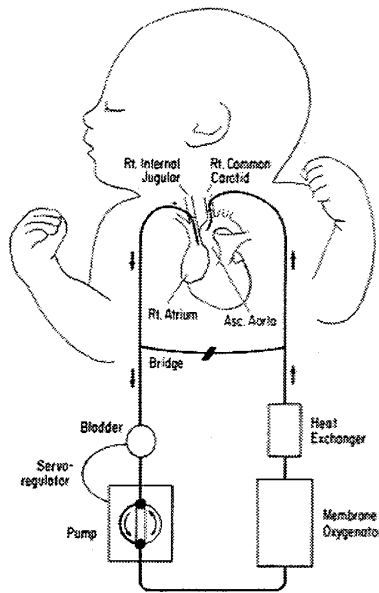


FIG. 1. Extracorporeal membrane oxygenation, shown schematically, is in essence a simplified heart-lung machine. Cannulas draw blood from the right atrium and return it to the aorta, thereby bypassing both heart and lungs.

ECMO 概念図 出典：Ware, 1989, p.299.

- 2) 胎盤呼吸から肺呼吸への正常な移行はつぎのようにして起こる。まず、産声とともに息が大きく吸い込まれ、新生児の肺が拡張し、その結果、全身の血圧より高かった肺血管の血圧が劇的に下がる。肺血圧の低下によって、多量の血液が一挙に肺に流れ込むようになり、その流れの急激な変化をうけて、左右の心房の間に開いていた穴が閉じる。その穴が閉じることで、肺を通らずに、右心房から左心房へと直接通じていた血流のバイパス通路が塞がれ、最終的に全血液が肺を循環するようになる。Cf. Cross, K.W., Cardio-respiratory Events at Birth, in Davis, J.A., and J. Dobbing, *Scientific Foundations of Pediatrics*, 1974, Heinemann Medical, London, pp.80-85. この一連のプロセスの初期の段階がうまく実現せず、肺血圧が高いままで、血流のバイパス通路が閉じない状態が PPHN である。PPHN はそれ単独で、ないしは他の症状との合併症として起こる。PPHN の主な合併症としては、横隔膜ヘルニア (congenital diaphragmatic hernia)、呼吸窮迫症候群 (respiratory distress syndrome)、胎便吸引症候群 (meconium aspiration syndrome)、敗血症 (sepsis) がある。Cf. Gomella, T. L., et al. eds., *Neonatology: Management, Procedure, On-Call Problems, Diseases and Drugs*, 3rd ed., 1994, Appleton & Lange, CN, p.290f.
- 3) ボストン小児病院と、ブリガム・アンド・ウィメンズ病院において、1982年と1983年の二年間にわたって蓄積されていた、PPHN の症状を示した34名の新生児のデータ (cf. O'Rourke et al., 1989, pp.958, 961f)。チルドレンズにおける ECMO の導入は1984年であるので、これらの新生児は皆、従来の人工呼吸器治療を施されていた。

始められ、1975年以降成功例を蓄積しつつあった⁴⁾。だが、ECMOは高価な上、複雑で大規模な装置であり、その運営には多数の熟達したスタッフを要する⁵⁾。また、ECMO治療は、新生児の頸動脈・頸静脈を切開して、そこから心臓までパイプを通すという、極めて侵襲的な治療法でもある⁶⁾。このような事情で、ECMOは限られた医療施設において、主に、通常の人工呼吸器治療ではおおむね80%以上の死亡率が予想される重篤な患者に対してのみ用いられてきた。しかし、そのような重症患者に対するECMOの効果は、人工呼吸器治療に比べて絶大であり、早くも1980年には、死亡率が30%以下に下がったという報告が出されている⁷⁾。またECMOの臨床例の全米データの収集も1984年から始まり、翌年から使用施設数・臨床数・全体としての死亡率が随時公表されるようになっていた。その公開データによれば、使用施設と臨床例は85年から87年にかけて、20施設で200例、25施設で500例、35施設で700例（いずれも概数）と増加の一途を辿っており、その間、生存率はほぼ80%の高率を維持していたことが分かる⁸⁾。

以上のような情勢の下、チルドレンズでも1984年以降ECMOによるPPHNの治療成功例を重ねていた。先の新生児をチルドレンズへ送り出した医師も、その成功例を知

4) 人工呼吸器治療、ECMO治療ともに、PPHNやその合併症の病因を直接取り除くことはしない。両者の役割は、新生児の肺機能を肩代わりすることで、肺機能が回復するまでの「時間かせぎ」をすることにある。ただ人工呼吸器治療による肺機能の「肩代わり」はあくまで部分的で、患者の肺は病因を抱えたまま吸気・呼気を強制的に繰り返させられている。また、人工呼吸器治療では、吹き込まれる空気の圧力で新生児の肺が傷ついてしまうという副作用がしばしば起こる。一方、ECMOによる心肺機能の「肩代わり」は完全で、患者の心肺は全く使用されない安静状態におかれ、肺が傷つくおそれもない。

5) アメリカにおける最近の例では、小児科医、外科医、ポンプをはじめとするECMO装置を担当する専門家一・二名、複数の正看護師(registered nurse)ないしは呼吸療法士(respiratory therapist)からなるチームが結成されている。Cf. Tulenko, D.R., An Update on ECMO, *Neonatal Network*, vol.23, no. 4, July/August 2004, pp.11-18.

6) 今日では、その他に、頸静脈のみを切開するタイプのECMOもあり、(心臓は除いて)肺に対するサポートのみが必要な場合に多用されている。Cf. Gomella et al.,1994, p.122f. Tulenko, 2004.

7) Cimma, R., et al., A Simple Objective System for Early Recognition of Overwhelming Neonatal Respiratory Distress, *Journal of Pediatric Surgery*, 1980, vol.15, pp. 581-585.

8) Bartlett, R., and C. Stolar, Extracorporeal Life Support: State-of-the-Art 1990, in Report of the Workshop on Diffusion of ECMO Technology, NIH Publication No.93-3399, January. 1993, <http://www.nichd.nih.gov/publications/pubs/ecmo/ecmo.htm>.

っていたからこそ、患者をわざわざ転院させたのである。そして既に触れたように、この新生児も ECMO 治療を施すに値する重症患者であることが確認された。しかも、その時、患者のいる集中治療室に設置されていた ECMO 装置は空いており、スタッフも待機していた。ECMO 治療を開始する条件はすべて整っていたのである。

—しかし、ここで一見、奇妙な事態が起こった。

ECMO はその新生児には装着されなかったのである。赤ん坊は、ECMO 治療が与えられないまま、別の集中治療室へと移され、そこで従来通りの人工呼吸器治療を施された。結局、治療の甲斐なく、その患者は約 8 時間後に、生後 2 日に満たない人生を閉じた。そして、この間、ECMO 治療が利用可能であったことは、両親にも伏せられたままであった。

このようなケースは実は始めてではなかった。過去 1 年半ほどの間に、チルドレンズにおいて、ECMO 治療を必要とする重症の PPHN の患者に対し、親の了解もとらないまま、利用可能な ECMO 治療があえて施されずに、結果として患者が死亡したケースが既に 3 例あった。先の新生児は 4 人目の死者だったのである。

—それでは、なぜ、このようなことが行われていたのか。

確かに、これらの患者が ECMO 治療を受けて助かった保障はない。しかし、その場合の死は、少なくとも事前データを見る限りは、利用可能な最善の治療法を施された上での結果であっただろう。ここで私が問題にしたいのは、「死亡」という結果ではなく、その死にいたるプロセスである。なぜ、4 人もの患者が、事前データに照らせば最善に思える、その場で利用可能であった治療を受けられないまま、死ななければならなかったのだろうか。

無作為臨床治験

実は、チルドレンズでは、ハーバード大学の研究者たちによって、前年 1986 年の 2 月 6 日から、重度の PPHN に対する、ECMO 療法と人工呼吸器療法の効果を比較する無作為臨床治験 (Randomized Controlled/Clinical Trail; ^{アール・シー・ディー} R C T) が実施されていたのである。先の新生児は、最初に運び込まれた集中治療室で ECMO を必要とする重い

症状であることが確認された時点で、治験に参加させることが研究者たちによって決定されていたのである。

RCTは、薬や外科手術といった様々な治療方法の効果や副作用を、実際に患者を治療しつつ調べることを目的としている。このような、治療を兼ねた一種の人体実験である臨床治験（clinical trial）の方法は、RCT以外にもいくつかある。その中でのRCTの特徴のいくつかをキーワード風に表現すると、「「同時対照群」を備え、「ランダム化」を実施する治験であること」となる。

新しく開発された治療法の効果や副作用を、従来からある標準的な治療法のそれと比べる比較実験において、新治療法を与えられる患者のグループを治療群（treatment group）、従来治療（もしそのような治療法がない場合は偽薬（プラセボ））を施されるグループを対照群（control group）という。対照群の中でも、治療群と比較することを目的に、同時並行的に設けられた対照群を同時対照群（concurrent control group）と呼ぶ。この同時対照群では、比較対象となっている治療法以外は、患者の症状の程度・全般的な治療環境など、治療の結果に影響を与えそうな要因が、できるだけ治療群と均質となるように配慮されている。一方、過去に、新治療との比較のためにではなく、通常の医療の一環として従来治療法を受けていた一群の患者を、後から振り返って改めて対照群と見なした場合、そのようなグループは既存対照群（historical control）と呼ばれる。当然のことながら、この既存対照群を、少なくとも同時対照群と同じ程度、治療群と均質なものとすることはできない。問題となっているチルドレンズにおける治験では、ECMO治療を施された患者のグループが治療群に、人工呼吸器治療を与えられた患者のグループが同時対照群に、それぞれ対応しているのである。

また、「ランダム化（randomization）」ないし「無作為化」とは、この文脈では、患者を治療群か同時対照群のどちらに振り分けるかの決定、言い換えると、患者に対して比較されている二つの治療法のどちらを施すかの決定を、「確率ないし偶然事象」（ここでは「確率（probability）」と「偶然（chance）」を同義語としておく）とされる事象の結果－例えば、コインを投げて「表」が出るか「裏」が出るか－に委ねることを意味する。ECMO治験に即して言えば、赤ん坊がECMO治療を受けるか、人工呼吸器治療を受けるかは、一種の「くじ引き」によって決定されていたのである。

患者の振り分けを何らかの確率事象の結果に委ねるとは、「患者の振り分け」という操作自体を一種の確率事象と化すことでもある。つまり、ここで言うランダム化、無作為化とは、「患者の振り分け」を確率（偶然）事象と化すことを意味しているのである。通常の医療では、患者にどの治療法を施すのかは、かなりの程度、個々人の患者の症状や選択肢に入っている治療法の長所・短所についての医師の個別的判断や、それに対する患者の同意に委ねられている。だが、患者の振り分けがランダム化されたRCTにおいては、ECMO 治験において人工呼吸器治療へとまわされ結果的には亡くなった4人の患者のケースも含め、個々人の患者の症状に応じた医師の判断によって治療が選択されることはないのである。

通常のRCTでは、比較される各々の治療法に、事前に定められた一定数の患者がほぼ同じ数だけ割り当てられる。しかし、チルドレンズにおけるECMO 治験では、先に紹介した全米データベースをも含めた事前情報が、ECMOが人工呼吸器治療よりも効果であることを示唆していた点を踏まえ、実験の早い段階でより劣った成績しか上げられなかった治療法へ割り当てられる患者数をなるべく少なくする、一種のアダプティブ・デザイン（adaptive design）が採用されていた。アダプティブ・デザインとは、実験の初期の結果に応じて、その後の実験のやり方を変えていく仕組みである。この治験では、患者が各々の治療法に振り分けられる確率を、それ以前にランダムに振り分けられた患者の治療結果に連動させて変えていくデザインが採用された⁹⁾。具体的には、どちらか片方の治療法で予め定められた一定数の死者が出た時点で、ランダムな振り分けを事実

9) アダプティブ・デザインの開発は、1940年代、ワルドによる逐次検定(sequential test)の提案にまでさかのぼる。ワルドの逐次検定は、この後で紹介するネイマン-ピアソン型仮説検定の拡張版と見なせるもので、全ての可能な標本統計量の集合の中に、帰無仮説を棄却する領域（棄却域）と受容する領域（受容域）に加えて、棄却も受容もせずにサンプルサイズを一つ増やして検定をやり直す第三の領域が設定される。実現値が第三の領域に入り続ける限り、サンプルサイズを増加させつつ検定は続けられ、最終的に、実現値が棄却域か受容域に入った時点で検定は終了する。即ち、逐次検定は、実験の初期の結果に応じてサンプルサイズが変わる仕組みを備えているのである。ちなみに、有意水準と検出力を算出する逐次検定においては、より少ないサンプルサイズで、一定の同じ有意水準・検出力を達成できる検定が「よりよい」検定であるとされる。Cf. Wald, A., *Sequential Analysis*, 1947, John Wiley & Sons, New York. また特に、ハーバード治験で用いられたような、ランダムな振り分けを初期の結果に応じて変えていくデザインは、データ依存型治療割り当て規則(data-dependent treatment assignment rule)と呼ばれる。

上中止し、その後は、患者を成績のよい治療法にのみ自動的に割り当てることが、予め決められていたのである。一般にアダプティブ・デザインの採用は、比較対象となる治療法の間に少なからぬ優劣の差があることが予想される場合に、劣った治療法を受ける患者の数をなるべく減らし、優れた治療法を受ける患者の数をできるだけ増やそうとする倫理的な配慮にもとづいている。ECMO 治験のケースでは、一般的に言って、定められた死者数の数が少なければ少ないほど、劣った治療法にまわされる患者の数は少なくてすむ。ハーバードの研究者たちは、事前情報に照らして、ECMO 治験においては、このような形で一定の倫理的配慮を加える必要があると判断したのである。

一方、この治験では、インフォームド・コンセントの取り方に関して、「ランダムイズド・コンセント (Randomized Consent; ^{アール・シー} RC) 法」と呼ばれる規則が採用され、その結果、一見すると倫理的配慮を欠いたようにも思える処置がとられた¹⁰⁾。具体的には、対照群にまわされ人工呼吸器をあてがわれた患者の親に対して、赤ん坊が治験の対象となっていること、ランダムに対照群に振り分けられたこと、そしてその治験を拒否することがいつでも可能なこと、さらには（既に触れたように）比較対象となる ECMO 治療の存在や、事前情報に照らした両治療法の成績の差といった事実が知らされなかったのである。通常の RCT では、まず患者に、実験の実施、ランダム化の介入、比較される二つの治療法についての事前情報などを説明し、治験参加の同意を得た上で—即ちインフォームド・コンセントを取った上で—ランダムな振り分けが行われる。それに対して、まず患者をランダムに振り分けた上で、治療群からのみインフォームド・コンセントを取り、対照群からはインフォームド・コンセントを取らないというのが RC 法のポイントである。このような RC 法は、対照群で用いられる治療法（ECMO 治験の場合は人工呼吸器治療）が、日常の医療の現場で定着している標準的なものである場合、かつその場合に限って許される方法として提案された¹¹⁾。

10) またの名を、提案者にちなんで「ツェレン法(Zelen's method)」とも言う。

11) RC 法に関しては以下を参照。Zelen, M., A New Design for Randomized Clinical Trials, *New England Journal of Medicine*, 1979, vol.300, pp.1242-1245. ちなみにツェレンは、ハーバード ECMO 治験の直接の実施者ではないが、その治験の統計デザインを設計したウェア(Ware)と議論を重ね、後で紹介する負の二項分布を用いたアダプティブ・デザインの考案においても大きな役割を果たしている。Cf. Ware, 1989, p.300.

そのようなケースでは、対照群の患者は、治験に参加しようが参加しまいが、いずれにせよ同じ標準的な治療を受けたはずであり、結局のところ、治験に参加するかどうかという選択肢は、その患者にとって実質的な違いをもたらさないことになる。実質的な違いを生まない選択肢に対しては、いちいちインフォームド・コンセントを取る必要がない。以上が、対照群の治療法が標準的である場合、そこに含まれる患者からインフォームド・コンセントを取らないRC法を用いてもよいとする理由である。ハーバードの研究者たちは、人工呼吸器治療をRC法の適用を許す程度に標準的なものと見なし、対照群からのインフォームド・コンセントを取らなくとも倫理的には問題は無いと判断したのである。いやそれどころか、この場合、RC法を用いることは、むしろ倫理的ですらあるとも彼らは考えていた。というのも、そうすることで、確かに新しいが未だ優れているとは言い切れないECMO治療に当たらなかったことを悔やむという、彼らに言わせれば「無用の悩みの種」を、対照群の親たちに与えずにすむからである¹²⁾。

さて、治験の結果は、事前の予想どおりであった。実験開始から約一年半の間に、ほぼ同程度に重症であると認定され、実験に組み入れられた患者のうち、ECMOに割り当てられた9人は、全員が治癒し副作用の兆候も見られなかったのに対し、人工呼吸器治療をあてがわれた10人のうち4人が死亡した。そして人工呼吸器治療において、冒頭に紹介した4人目の死者が出た段階で、予定どおりランダム化は中止され、その後は全ての患者に自動的にECMO治療が施された。その際、人工呼吸器治療における生存者6人という数字を受けて、ECMO療法でトータルとして28人の治癒例が出るか、4人の死亡例が出るまで治験が続けられることとなった。その後20人の患者がECMO治療を受け、そのうち1人を除いて全員が治癒し、ECMO治療でトータル28人目の生存者が出た1988年7月1日をもって治験は終了した¹³⁾。結果として、2年半の治験の期間において、ECMO治療においては29人中28人治癒1人死亡、生存率97%、一方の人工呼吸器治療においては10人中生存者6人死者4人、生存率60%というデータが得られた。人工呼吸器治療の生存率は、事前情報の20%以下よりは大幅に

12) O'Rourke et al., 1989, p.959.

13) Ware, 1989, p.301.

改善されたものの、ECMOの生存率との差はなお37%もあったのである。

統計実験としてのECMO治験

RCTは、他の多くの臨床治験と同様、統計学の方法論にのっとって実施される統計実験である。ECMO治験も例外ではない。ECMO治験で用いられた統計的技法は、主として古典方法論（classical methodology）における、ネイマン - ピアソン型仮説検定（Neyman-Pearsonian hypothesis test）であった¹⁴⁾。統計学は、科学における実験や観察、標本調査の方法論である。ハーバードの研究者たちも、古典統計学の規則に則って治験をデザインし、データを集め、仮説検定の技法を用いて、そのデータを加工し縮約・集約した上で、そこから一定の結論を導き出したのである。

例えば、患者を実験に参加させるかどうかを決定する明確な基準（entry criteria）を設定することも、統計学が指示する規則の一つである。そして、冒頭に紹介した新生児が最初に運び込まれた集中治療室で受けた検査も、この基準を満たしているかどうかを判定する役割を担っていたのである。また、患者のランダムな振り分け、即ちランダムマイゼーションを実施することも、古典統計学においては不可欠の作業として位置づけられている。データの収集プロセスをランダム化することで、はじめて、そのデータに対して正当に古典統計学の手法を適用することが可能となる。統計的方法論において、ランダムマイゼーションは様々な機能を果たしているが、その最大の機能は、古典統計学の技法の適用を正当化することにあるのである¹⁵⁾。

しかし、基準に合格し、ランダムに振り分けられた患者から得られたデータは、統計学の方法に則って集められたデータではあるものの、いまだ、統計的な分析が加え

14) 治験においては、ネイマン - ピアソン型仮説検定と共に「区間推定(interval estimation)」も用いられた。また、この治験の統計デザインを担当したウェアは、後に、その統計デザインを自ら検討し正当化するに際して、治験の結果に対してベイズ統計学の分析手法をも用いている。Cf. Ware, 1989, p.303.

15) 特に、臨床治験においてランダムマイゼーションが果たしている機能については、例えば以下を参照。Petrie, A., Why Randomization Is Essential and How to Do It, in Tygstrup, N., et al., eds., *The Randomized Clinical Trial and Therapeutic Decisions*, 1982, Marcel Dekker, N.Y., pp.105-116.

られる以前の「生のデータ (raw data)」に過ぎない。ECMO 治験における「ECMO 治療での生存者 28 人・死者 1 人・生存率 97%、人工呼吸器治療での生存者 6 人・死者 4 人・生存率 60%」という結果も、統計的分析にとっては、あくまで「素材」である。その素材を一定の仕方で加工し、さらに加工されたデータをもとに様々な推論を展開することにおいてこそ、統計的方法論の本領が発揮されるのである。ここではそのデータ加工と統計的推論のあらましを、ECMO 治験におけるネイマン-ピアソン型仮説検定を例に概観しておこう。

ネイマン-ピアソン型仮説検定は、帰無仮説 (null hypothesis) と対立仮説 (alternative hypothesis) という二つの仮説を設定することから始まる。多くの場合、対立仮説は研究者の予想を表しており、仮説検定は、帰無仮説を棄却し対立仮説を受容することを目指して行われる。ハーバードの ECMO 治験においては、「ECMO と人工呼吸器治療の生存率はともに 20% である」という帰無仮説と、「ECMO の生存率は 80% で、人工呼吸器のそれは 20% である」という対立仮説が設定された。後者は、治験開始時に出ていた全米データに基づいて設定されており、通例どおり研究者の事前予想を表現するものとなっている。

つぎに、生のデータを一定の数値として集約・縮約した検定統計量 (test statistics) が決定される。ECMO 治験における検定統計量は、人工呼吸器治療と ECMO 治療のそれぞれで 4 人の死者が発生した時点で、得られていた生存者の数である。いま治験を無限回繰り返したとして、得られる検定統計量の各々の値の出現頻度が、一定の確率分布関数として、それぞれ帰無仮説と対立仮説を前提した上で計算される。帰無仮説と対立仮説から数学的に導出される、この検定統計量の理論的な分布は標本分布 (sampling distribution) と呼ばれる。検定統計量は、その標本分布が帰無仮説と対立仮説から導くことができるようなものでなければならないのである。ちなみに ECMO 治験における標本分布は、「条件付き二次元負の二項分布」という確率関数である。

帰無仮説から導かれた標本分布をもとにして、帰無仮説の下で予想された検定統計量 (ECMO 治験の場合は、人工呼吸器と ECMO 治療での 4 人死亡時点での生存者数がともに同数で、死亡数の $1/4$ 、即ち 1 人となるという値) から、最も離れた検定統計量の値の領域において、極めて小さな出現頻度しか与えられていない領域が棄却域

(critical region) として選ばれる。この場合の小さな出現頻度は有意水準 (level of significance) と呼ばれるが、どのような値をとるかは、基本的に実験者の恣意的な選択に委ねられている。ECMO 治験の場合は、よく選ばれる有意水準の一つ、0.05 が用いられていた。この有意水準の下での棄却域は、例えば ECMO と人工呼吸器治療における生存者数の合計が 34 人となった場合、ECMO と人工呼吸器での生存者数が、それぞれ 28 人以上、6 人以下となる領域に相当する¹⁶⁾。

ここまで事前に決めておいて、実験が実際に行われ、検定統計量の実現値 (obtained value) が求められる。ただ、この実現値は、それ自身、確率事象とされたプロセスを介して得られたデータ—そのようなデータは確率標本 (random sample) と呼ばれる—を加工したものでなければならない。というのも、データが確率標本であるという前提の下で、先に言及した帰無仮説や対立仮説からの標本分布の導出が行われていたからである。ランダム化が古典統計学の方法論にとって不可欠であるのは、標本分布の導出という、その方法論の根幹をなす作業においてデータが確率標本であることが前提されているからなのである。

さて、確率標本から得られた実現値が棄却域に入った場合、「帰無仮説が極めて低い出現頻度しか与えなかった事態が現実には生じた」ことを理由として「帰無仮説が偽である」という結論が導かれる。その結果、帰無仮説が棄却され、対立仮説が採用されるのである。ECMO 治験において得られた実現値は、「人工呼吸器治療の生存者 6 人、ECMO 治療の生存者 28 人以上」というものであった。先に触れたように、この実現値は、両治療法の合計生存者数が 34 人である場合の 5 % 有意水準の棄却域 (これを 5 % 棄却域と呼ぶ) に入っている。ECMO 治験では、「ECMO と人工呼吸器治療の生存率はともに 20% である」という帰無仮説が退けられ、「ECMO の生存率は 80% で、人工呼吸器のそれは 20% である」という対立仮説が受容されたのである。対立仮説として表現された研究者の事前予想が、検定統計量の実現値によって「確証」されたとも言える。

しかし、稀にしか起こらないとされた結果が実際に起こったからといって、そのような予想を導いた仮説が誤っているとは必ずしも言えない。本当に稀な事象が現実

16) いま <ECMO 生存者数・人工呼吸器生存者数> と表現するとして、<28・6>、<29・5>、…、<33・1>、<34・0> という組み合わせからなる領域である。

起こった可能性は常に捨てきれないのである。つまり、実現値に対して低い出現頻度を与えたからといって帰無仮説を偽であるとする先の推論は、「本当は正しい帰無仮説を偽として退けてしまう」という誤り－これを「第一種の誤り」と呼ぶ－を犯している可能性があるのである。

一方、検定統計量の実現値が棄却域に入らなかった場合、今度は帰無仮説が真であるとして受け入れられる。つまり、帰無仮説が検定統計量の実現値に照らして確証されたことになるのである。しかし、ここにも先の第一種の誤りと似た誤り、即ち「本当は間違っている帰無仮説を真として受け入れてしまう」という誤りの可能性がひそんでいる。この誤りは「第二種の誤り」と呼ばれる。

一般に、必ずしも白か黒かの明確な判断がつかねるデータを前にして、一組の仮説のうちの一方を採用するという二者択一的な決定を行うさいには、この種の二つの誤りの可能性はつきものである。統計的仮説検定、なかんずくネイマン－ピアソン型の仮説検定の一つの重要な特徴、ないしは決定的な長所とは、この誤りの可能性を明示的に認めた上で、それを一定の確率値（即ち相対頻度値）として量的に表現できることにある。そして一旦、誤りの可能性が確率値として量的に表現されたならば、今度は、それらの誤りの確率値が低ければ低いほど、検定はより「よい」ものとなり、検定が与える対立仮説への確証の度合いはより「高い」ものと言えるようになる。統計的仮説検定の発明により、科学者は、検定の「よさ」や確証の「度合い」を測る「尺度」を手に入れたのである。後は、この尺度に照らして、さまざまな仮説検定の結果を比較したり、より「よい」検定、より「確証度」の高い検定の実施を目指していけばよいのである。

ところで、第一種の誤りの確率、即ちそのような誤りが起こる相対頻度は、上で触れた有意水準によって表現されている。その理由は以下である。そもそも有意水準とは、テストが無限回行われた場合、帰無仮説の棄却に結びつくような一定範囲のテスト結果に対して、帰無仮説自体が与えていた相対頻度であった。この相対頻度は、帰無仮説が正しいという前提の下で、帰無仮説の棄却が起こる相対頻度であると読み換えることができる。つまり、有意水準とは、テストを無限回繰り返すうちに、われわれが誤って、本当は正しい帰無仮説を退けてしまう相対頻度、即ち、第一種の誤りを犯してしまう確率を表しているのである。（より正確に言えば、有意水準は第一種の誤

りの確率の最大値に相当する。)

また、第二種の誤りの確率の指標としては、通常、検出力 (power) が用いられる。検出力とは、「帰無仮説が偽である場合に、首尾よくそれを退けることができる確率」であり、第二種の誤りの確率とはネガとポジの関係にある (正確に言えば「検出力 = $1 - \text{第二種の誤りの確率}$ 」となる)。即ち、検出力が高まれば高まるほど、その分だけ、第二種の誤りの確率は小さくなるのである。検出力は様々な要因によって決まる。具体的には、テストの興味の対象となっている真の値 (ECMO 治験のケースでは、二つの治療法の生存率)、帰無・対立仮説のタイプ (それらが特定の値を予測しているのか、それとも一定の範囲の予測にとどまっているのか)、標本分布の関数形 (この関数形はまた、サンプルサイズ - ECMO 治験のケースでは二つの治療法の合計患者数に相当 - によっても変わってくる)、検定のその他の細部のデザイン (有意水準や棄却域の選び方、片側検定か両側検定かの違い) などが、検出力に影響を与えているのである。もちろん、通常われわれはテスト対象の真の値など知らない。(知らないからこそ、わざわざ統計実験を行うのである。) したがって、検出力が真の値に依存している以上、検出力の正確な値を算出することは原理上できない。しかし、他の要因を色々と調整することによって、検出力をより大きく設定することは可能である。結局、第一種・第二種の誤りの確率がより低いという意味で、よりよい検定を目指すためには、有意水準をできるだけ小さく、検出力をできるだけ大きくするような仕方で検定をデザインすればよいのである。

しかし残念ながら、一般に、有意水準を下げれば下げるほど、検出力もまた下がってしまう。そこで次善の策として、有意水準を一定の小さな値 (例えば 0.05) に固定した上で、検出力をより大きくするという方策がとられる。一例をあげよう。一定の有意水準を決めたとしても、棄却域の取り方は実は様々である。しかし、ある特定の条件の下では、テストの対象となっている真の値がどのようなものであっても、常に、検出力を最大にするような棄却域が一つ存在することが、ネイマンとピアソンの基本定理によって証明されている¹⁷⁾。このような場合においては - 他のデザインは全て同じとして -

17) Neyman, J., and E.S. Pearson, On the Probability of the Most Efficient Tests of Statistical Hypothesis, *Philosophical Transaction of the Royal Society*, 1933, A231, pp.289-337.

数ある棄却域の中から、任意の真の値に対して常に検出力を最大にする棄却域を選んだ検定（「一様最強力検定」と呼ばれる）が、最もよい検定と言えるのである。

ところが、ECMO 治験のケースも含め、多くの場合には、一様最強力検定は存在しない。しかしそのような場合でも、大概の場合、サンプルサイズを増やせば増やすほど、（他の条件が同じならば）任意の真の値に対して、検出力がより大きくなることが知られている。ECMO 治験のサンプルサイズは両治療法での生存者の合計数であったが、この合計数はまた、ランダム化が中止される死者数によっても影響を受ける。つまり、ECMO 治験における（5%有意水準の下での）検出力は、（他の要因を固定すれば）合計生存者数と死者の数によって増減するのである。

これら二つの数のうち、合計生存者数は、治験が完了した時点ではじめて確定する数であり、治験のデザイン段階では、どうなるか未だ不明である。そこで、ハーバードの研究者たちは、合計生存者数を度外視し、死者数のみを考慮に入れて、（さらに、真の値は対立仮説の予測に一致するという前提の下で）あらかじめ検出力を算出した。具体的には、死者の数3人で55%、4人で77%、5人では84%という検出力が報告されている¹⁸⁾。

実は、ここで算出された、合計生存者数を度外視した検出力は、治験で実際に用いられた条件付きの標本分布ではなく、条件のつかない分布（即ち、単なる二次元負の二項分布）を前提として得られたものである。つまり、検出力の事前の算出においては、いわば標本分布の「すり替え」が行われているのである。われわれの計算によると、条件付き標本分布に対して求められた検出力は、合計生存者数と死者数に応じて、かなり複雑な増減の様相を呈した。（合計生存者数と死者数の片方を固定し、もう一方を増加させると、大局的に見れば、検出力は増加するが、その増加は決して単調増加ではない。）標本分布の条件を外し、合計生存者数を度外視することで、検出力を死者数の比較的単純な関数として描き、治験のデザインを容易にすること。これが「すり替え」の理由であったと思われる。ただいずれにせよ、検定で現実用いられた標本分布に対して求められた検出力でなければ、定義上、その検定の検出力とは言えない。上で並べられた数値は、死者の数の設定の仕方によって変わる検定の「よさ」を測る、統計学的には必ずしも正当とは言えない、インフォーマルな目安—いわば「擬似」検

18) O'Rourke et al., 1989, p.958f. Ware, 1989, p.301.

出力とでも呼べるもの－に過ぎないのである。

とにかく、このインフォーマルな指標としての擬似検出力を前にして、治験のデザイン（この場合は、死者の数）が選ばれることになるのだが、ここで一つのジレンマが生じる。これまで述べてきたように、検定を「よりよい」ものとするためには、死者の数は大きければ大きい方がよい。その分、検定の擬似検出力は上がり、統計学が作り出した「尺度」に照らせば、実験結果の証拠としての「質」は向上するのである。しかし他方、先に指摘したように、劣った治療法を受ける患者の数をなるべく少なくするというアダプティブ・デザインの趣旨から言えば、死者の数は少なければ少ないほうがよい。ここにおいて、統計学、ひいては科学方法論からの要請と、倫理的な要請が、真っ向から衝突しているのである。

このジレンマを解く唯一正しい解答は存在しない。研究者は多かれ少なかれ恣意的な仕方で、ジレンマの二つ角の間のバランスを取らざるを得ないのである。ハーバードの研究者が着目したのは、死者の数を一人ずつ増やした場合の、擬似検出力が増える度合いであった。いま、死者の数を3人から4人に増やせば、擬似検定力が22%分アップするのに対し、4人から5人に増やしても、擬似検定力は7%分しかアップしない。擬似検出力22%分の増加は、許容される死者の数を1人増やすことを正当化するが、7%分の増加だけではそれは正当化されない。ハーバードの研究者たちの判断の根拠をあえて明示化すれば、こうなるであろう。ランダムな振り分けを中止する死者の数が3人や5人ではなく4人に設定された背後には、このような判断が働いていたのである。

以上で概観してきたように、統計的実験、なかんずくネイマン－ピアソン型の仮説検定においては、単にデータに基づいて一定の判定が下されるだけではなく、同時にその判定自体の質が、第一種・第二種の誤りの確率、言い換えると有意水準と検出力という一定の量的指標によって表現されることになる。ECMO 治験に即して言えば、「ECMOの生存率は80%で、人工呼吸器の生存率は20%である」という対立仮説が、「ECMOと人工呼吸器治療の生存率はともに20%である」という帰無仮説を差し置いて、有意水準5%・（擬似ではない本物の）検出力99.81%を以って確証された」という結論が得られたのである。

冒頭的新生児に話を戻そう。もしこの治験において死者の数が4人ではなく3人に

設定されていたなら、この赤ん坊には、人工呼吸器治療ではなく、ECMOが割り当てられることになっていた。この新生児は、つまるところ、仮説検定の結果の「よさ」を表す指標の一つである検出力－正確に言えば、統計学的には必ずしも正当とは見なされないインフォーマルな指標に過ぎない擬似検出力－を55%から77%へと22%分上げるため、そしてそのためだけに、事前データに照らせばより助かる見込みが高いと思われる治療法を受けることができなかったのである。

倫理的疑念

「なぜ、4人もの患者が、事前データから見て有望であると思えたECMO治療を受けられないまま死亡したのか」という先に立てた疑問は、「チルドレンズでは、ECMOと人工呼吸器の効果を比べる、統計実験の一種であるRCTが行われており、それらの患者はその治験の対象となっていた」という事実によって、さしあたっては答えられる。しかし、この答えは、ただちに、さらなる疑念をいくつも惹起するであろう。例えば、われわれは「それでは、このECMO治験は倫理的に言って、また方法論的な観点から正当なものだったのか」と問いたくなる。

そもそも人体実験は、様々な倫理的な問題をはらんでいる。それらの問題を踏まえ、古くは「ヒポクラテスの誓い」や、また20世紀の「ニュルンベルク綱領」や「ヘルシンキ宣言」など、人体実験一般に関する各種のガイドラインがこれまで提案されてきた。「同時対照群への患者のランダムな振り分け」を特徴とするRCTに対しても、これらの一般的な指針を特殊化した、いくつかの倫理的なガイドラインが設けられている。中でも、最も早く提案されたものの一つとして、後の様々なガイドラインに対する一つのモデルの役割を果たしたことで有名なのが、「RCTの父」と称される、イギリスの医療統計学者ブラッドフォード・ヒル（Austin Bradford Hill; 1897-1991）が提唱した一群の指針である¹⁹⁾。ヒルによれば、RCTにまつわる根本的な倫理的問題とは、同時対照群にランダムに割り当てられた患者に対して、有望だがその効果（や副作用）

19) Bradford Hill, A., and I.D. Hill, *Bradford Hill's Principle of Medical Statistics*, 12th ed., 1991, Edward Arnold, London, pp.211-215.

が未だ十分には確かめられていない「最新の治療」を用いないことの是非にある。言い換えると、最新の治療を用いないことが、どの場合には許され、どの場合には許されないかを見極めることが重要なのである²⁰⁾。

この「見極め」のための基準の一つとして挙げられるのは、標準的な治療と最新治療の（予想される）効果の落差である。例えば、その落差が、ごく普通の風邪が数日早く直るかどうかが程度の軽微なものであれば、同時対照群に対して最新治療を施さないことは倫理的には大した問題ではない、とヒルは言う。逆に、その落差が生死を分かつほど大きい場合、即ち、標準的な治療では助からない患者でも最新治療では救えるかもしれないというケースでは、同時対照群を設定することは、倫理的にほとんど不可能となる、というわけである²¹⁾。

これ以外にも、ヒルは、「医療平衡 (clinical equipoise)」と後に呼ばれることになった概念を用いた基準をも提案している。ここで言う医療平衡とは、比較される治療法的一方が優れているという「よい証拠 (good evidence)」が前もって得られておらず、どちらの治療法がより優れているかについて医者が知識 (knowledge) を持っていない、即ち無知 (ignorant) であるような状態のことを指す。つまり、「どちらの治療法の方がよいか」に関して、医者の知識が宙ぶらりんの平衡状態にあるような状況である。ヒルは、最新治療を施さないことは、そのような状況下でのみ倫理的でありうると思った。どちらか一方の治療法がより優れているか知りつつも、その治療法を用いないことは、目の前の患者に最善を尽くすという医師としての義務に反してしまうことになるからである²²⁾。この、「医療平衡が成り立った場合にのみRCTは倫理的となる」という考えは、早くは、ヒルも起草委員に加わった「イギリス医学研究評議会 (Medical Research Council)」の臨床治験の倫理に関する声明にも採用され²³⁾、その後、

20) Hill and Hill, 1991, p.211. Hill, A.B., *The Philosophy of the Clinical Trial*, The National Institute of Health Annual Lectures, 1953, Washington D.C., reprinted in Hill, A.B., *Statistical Methods in Clinical and Preventive Medicine*, 1962, E.& S. Livingstone, Edinburgh, p. 6.

21) Hill and Hill, 1991, p.211. Hill, 1953, p. 6 - 7.

22) この考えがヒル自身の著作で明文化されるのは、1966年以降である。Cf. Hill, A.B., *Principles of Medical Statistics*, 8th ed, 1966, (but not in earlier editions) Edward Arnold, London.

23) The Report of the Medical Research Council for 1962-63, reprinted in Hill and Hill, 1991, pp.314-319.

様々なバリエーションを加えながらも、今日にいたるまで、RCTが満たすべき倫理的条件の一つの指針として機能している。

ヒルはこのようにして、最新治療を施さないことが倫理的に許される基準をいくつか提示したが、一方で、これらの基準が満たされれば無条件にRCTを実施してもよい、という単純な立場はとっていない。例えば、医療平衡が満たされたとしても、妊婦や複数の病気を併発している患者、さらには老人や幼児を治験に参加させるかどうかには、なお一考が必要であるとの留保条件をもつけているのである²⁵⁾。

以上のようなヒルのガイドラインが修正の余地無く正しいかどうかは別として、それをあくまで暫定的な指針として採用したとしよう。すると、ECMO治験は倫理的に妥当だと言い切れるだろうか。例えば、ECMO治験においては、標準的な人工呼吸器治療と新しいECMO治療との間には、生死を分けるほどの重大な「落差」が予想されていたと言えないだろうか。また、治験が始まる時点でECMO優位の事前情報が蓄積されていたことを考えると、この場合に医療平衡が成り立っていたかどうかとも議論の余地があるように思える。さらに、たとえ医療平衡が満たされていたとしても、生まれたばかりの新生児を治験の対象とすることは、そもそも許されるのだろうか。これらの点に関して、ECMO治験が、少なくともヒルの指針に照らして見ても、倫理的に妥当なものだったかどうかは一控え目に言っても一疑問の余地が残るのである。

加えて、ECMO治験におけるRC法の適用が倫理的に許されるかどうかとも問題である。ヒルのガイドラインは、インフォームド・コンセントに対しては（他の条項にも増して）歯切れが悪い²⁶⁾。例えば、それは、合意を取り付けるべき患者の範囲、開示すべき情報

24) 医療平衡のさまざまなバージョンとその問題点については、Deguchi, Y., Are ECMO Trials Ethical? *Prospectus*, 2003, vol. 6, pp.124-150を参照。

25) Hill and Hill, 1991, p.213. 幼児や子供といったインフォームド・コンセントを取れない患者に対して、どのように治験を行うべきかについては、その後も議論が絶えない。Cf. Show, A., Dilemmas of "Informed Consent" in Children. *New England Journal of Medicine*, 1973, vol.17, pp.885-889. Levine, R.J., *Ethics and Regulation of Clinical Research*, 2nd ed., 1986, Yale U.P., New Haven, chap.10 'Children'. McLean, S.A., *A Patient's Right to Know: Information Disclosure, the Doctor and the Law*, 1989, Dartmouth P.C., England. chap.3 'Special Groups: children'.

26) 例えば、ヒルは、医療平衡が成り立っていた場合、インフォームド・コンセントが必要かどうかは、ケース・バイ・ケースであるとも読めるニュアンスに富んだ表現を用いている (Cf. Hill and Hill, 1991, p.213)。また彼は、全ての患者からインフォームド・コンセントをとるべき

の内容に関して、明確な指針を与えていないのである。これらの点については、ヒル以降、様々な角度から議論がなされた。ランダム化比較試験の実施を患者に開示すべきかも、議論が分かれた点である²⁷⁾。ただ、70年代半ばまでには、その点も開示すべきであるという意見が、少なくともアメリカでは大勢を占めるにいたっていた²⁸⁾。そのような状況下で1979年にRC法が提案され、賛否両論、新たな議論を巻き起こしたのである²⁹⁾。ECMO 治験は、このような論争の渦中であって、あえてRC法を実地で用いた最初のRCTの一つだったのである³⁰⁾。そもそも、RC法は妥当か。またそれをECMO 治験で用いることは適切か。「要らぬ不安の種を取り除く」というハーバードの研究者の言い分は真っ当か。RC法を巡る論争状況を視野に入れば、これらの点もまた一控えめに言って議論が分かれるところであろう。

以上の諸点を考え合わせれば、ECMO 治験の倫理性が問われるのも、ある意味自然であろう。私としても、「この治験は、かの4人の患者からECMO 治験を受ける機会を倫理的に不当な仕方奪い、結果として、4人を、いわばむざむざと死なせてしまったのではないか」という疑念がぬぐい去れないのである。実際、治験を実施したハーバードの研究者たち自身も、一方では実験の正当性を主張しつつも、他方では、倫理的問題に対して、「議論や批判が多々あるであろうことを予測していた」と、ジャーナリストのインタビューに対して答えているのである³¹⁾。

だとも言い切っていないのである。

27) この点に関する賛否両論については、Levine, R.J., and K. Lebacqz, Some Ethical Considerations in Clinical Trials, *Clinical Pharmacology and Therapeutics*, 1979, vol.25, pp.728-742 を参照。

28) 例えば、ランダム化比較試験の開示に反対していたチャルマーズ (T. Chalmers) も、その頃から自説を事実上撤回し、RCTにさいしてランダム化比較試験の実施をも患者に開示するようになっていた (cf. Levine, 1986, p.195)。ちなみにチャルマーズは、イギリスで開発されたRCTをアメリカに導入し、その普及につとめた、いわばアメリカにおけるRCTの「伝道師」とも呼べる人物である。

29) Cf. Levine, 1986, pp.194-199.

30) ハーバードのECMO 治験に先立つRC法の採用例の一つは、この後紹介するECMOについてのミシガン治験である。ちなみにRC法の開発者ツェレンは、反論を受けて、ランダム化比較試験をした後、対照群からもインフォームド・コンセントを取る、一種の妥協案を提案するにいたっている。Cf. Zelen, M., Randomized Consent Designs for Clinical Trials: An Update, *Statistics in Medicine*, 1990, vol.9, pp.645-56.

31) Knox, R.A., A Harvard Study on Newborns Draws Fire: Doctors Faulted for Limiting Life-saving Treatment, *The Boston Globe*, August 7, 1989.

方法論的疑念

ECMO 治験についてわれわれが抱く懸念は、倫理的なものに限らない。「RCT、ひいては科学的探究の一環として、適切な方法が採用されていたかどうか」という方法論の観点からの疑念もいくつか指摘できるのである。その最大のものは、「そもそもこの ECMO 治験は必要ないし十分なものだったのか」という疑問であろう。

ハーバードの治験が行われた時点で蓄積されていた事前情報を、もう一度ここで確認しておこう。ECMO は PPHN に対して、1972 年から臨床例を重ねており、その詳細な結果も専門雑誌に報告されていた。また 1984 年からは組織的なデータベースの収集が始められ、治験開始時点で、全米 30 前後の施設において実施された、数百個の臨床例が集約され、ECMO が、生存率 80% という、人工呼吸器治療をはるかに引き離す好成績を挙げていたことが公表されていた。またこのデータベースに登録されている実施施設数、臨床例数は、いずれも増加の一途を辿っており、治験を行わずとも、毎年、数百例を超える新たなデータが追加されていく体制が整っていたのである。

さらに、実は、ハーバードの研究者によるチルドレンズにおける ECMO 治験の以前にも、ミシガン大学のグループが、1982 年から 1984 年にかけて、人工呼吸器を装着された同時対照群と ECMO を与えられた治療群を比較する初めての RCT を行い、その結果を 1985 年に公表していた。チルドレンズでの治験は、ECMO と人工呼吸器の効果を比べる二番目の RCT だったのである。ちなみにミシガン実験を主導したバートレット (Robert Bartlett: 現ミシガン大教授) は、60 年代から ECMO の開発に携わり、先に触れた 1975 年における PPHN 患者の ECMO 治療の最初の成功例を導き、その後、ミシガン大学に拠点を移して、ECMO の臨床例を蓄積するとともに、専門家向けのセミナーを相次いで開催したり、例の全米データベースをも構築するなど、PPHN に対する ECMO 治療の開発と普及に中心的な役割を果たした人物である³²⁾。もちろん、バートレットのチームも、治験を行う時点で ECMO 優位の事前情報を得て

32) ちなみに、バートレットの ECMO 開発の功績に対して、2003 年度にアメリカ外科協会 (American College of Surgeons) のジェイコブソン発明賞 (Jacobson Innovation Award) という権威ある賞が贈られている。

おりーというより、彼のチームこそが、その事前情報の主な発信元であったわけだが、それにもとづいて、治験においてECMO治療を受けたほとんどの患者は助かるが、逆に人工呼吸器治療を受けたほとんどの患者は助からないかもしれないとの予想を表明していた³³⁾。このような予想の下で、ミシガンのチームは、ハーバード治験とは異なった、患者数がより隔たった治療群と対照群をもたらす可能性が高いアダプティブ・デザインを採用したのである³⁴⁾。結果は、「10人の患者がランダムに振り分けられ、ECMOにあてがわれた9人全員が治癒、人工呼吸治療にあてがわれた1人が死亡」というものであった。この生のデータにもとづいて、ミシガンの研究者たちも、ネイマン-ピアソン型の仮説検定を行った。その検定の結果は、必ずしも標準的な仕方では報告されているわけではないが（例えば、帰無仮説や有意水準は明示されていない）、それを標準的な仕方では読み直すことは可能である。即ち、ここで用いられた帰無仮説は「二つの治療法の生存率は同じ」というものであるのに対して、例えば「どちらか一方の生存率は80%以上で、両者の生存率の差は40%以上にのぼる（つまり他方の生存率は60%以下である）」という仮説が対立仮説として置かれているのである。この対立仮説を真であると仮定し、実際に得られたサンプルサイズをもとに検出力を

33) Bartlett R., et al. Extracorporeal Circulation in Neonatal Respiratory Failure: A Prospective Randomized Study, *Pediatrics*, vol.76 No.4, October 1985, p.480. Bartlett, R., and R. Cornell, Comment, *Statistical Science*, 1991, vol. 6, No. 1, p.63.

34) ミシガン・チームが採用したアダプティブ・デザインは、ツェレンが発明した「勝ち馬に乗る規則」の改良版であり、「ランダム化された勝ち馬に乗る(randomized play the winner)規則」と呼ばれているものである (cf. Zelen, M., Play the Winner Rule and the Controlled Clinical Trail, 1969, *Journal of American Statistical Association*, 1969, vol.73, pp.840-843. Wei, L.J., and S. Durham, The Randomized Play-the-Winner Rule in Medical Trials, *Journal of American Statistical Association*, 1978, vol.73, pp.840-843)。この規則は、「壺の中から球を手探りで取り出す」という確率論でよく用いられる「壺モデル(urn model)」を用いてつぎのように説明することができる。まず壺の中に赤球と白球を一つずつ入れておき、手探りで一つの球を取り出す。取り出された球が赤球ならば患者はECMOに、白球ならば人工呼吸器にあてがわれる。この状態で、最初の患者がECMOまたは人工呼吸器に振り分けられる確率はともに $1/2$ である。いま赤球が取り出され、最初の患者がECMOに振り分けられたとする。その患者が助かれば、取り出された赤球に加えてもう一つの赤球を壺に入れ、もし助からなかったとしたら、赤球を戻す一方で白球を壺に加える。すると二番目の患者がECMOと人工呼吸器に振り分けられる確率は、最初の患者が生存した場合には $2/3$ と $1/3$ に、死亡した場合には $1/3$ と $2/3$ に変わる。そして、このような振り分け確率の変更を続けていって、最終的にどちらかの生存者数が予め決められた数（この場合は10人）に到達した時点で治験は終わるのである。

計算すると95%という高い値が得られる³⁵⁾。他方、有意水準は通常用いられる値をはるかに超えた50%に達している³⁶⁾。

つまり、ミシガン・チームは、劣った治療法へまわされる患者の数を極力おさえ、両治療法の効果の差が実際に大きかった場合には、それを見落とす確率を最小限にとどめる一方で、これらの代償として、両者の差が無かった場合には、その事実を高確率で見誤ってしまうという欠点を甘受する道を選んだのである。このような選択の背後には、ECMOの方が人工呼吸器よりもはるかに優れているという、ECMO開発者としての確固たる信念があったであろうことは想像に難くない。しかし結果として、ミシガン治験は、患者を1人しか含まないという極端に小さな対照群しか持たず、有意水準という指標に照らしても決して質が高いとは言えない検定となってしまった。それでも、検出力といういま一つの指標に照らせば十分満足すべき質を備えた検定によって、統計学的に正当な仕方で、「ECMOの方がはるかに優れている」という仮説が確証されたことも事実である³⁷⁾。つまり、ミシガン治験に対する評価は様々でありうるにせよ、ハーバード治験の開始時には、全米データベースなどの情報以外に、ECMO優位を示すRCTの結果も、事前情報として得られていたのである。ECMOと人工呼吸器の優劣を判定するために、このような事前情報に加えて、さらにもう一つのRCTの結果を付け加えることが、果たして必要だったのだろうか。

他方、もしハーバード治験が必要だったとしても、それはECMOと人工呼吸器の優

35) 「一方の生存率は90%で、他方は10%」という、より大きな差を想定する対立仮説の検出力は実に100%となる(cf. Bartlett et al., 1985, p.485)。

36) 以上の標準的な結果への読み替えは、一年後にハーバード実験を計画・実施することになる研究者たちによるものである(cf. Ware, J.H., and M.F. Epstein, Extracorporeal Circulation in Neonatal Respiratory Failure: A Prospective Randomized Study, *Pediatrics*, 1985, vol.76, pp.849-851)。

37) ミシガンの研究者たちは、治験終了後も8ヵ月間にわたってさらに同様のデータを取りつづけた。その結果はECMO治療を受けた8人は全員生存、人工呼吸器治療を受けた2人は全員死亡というものであった。これらのデータと治験期間中のデータを加えた結果に照らして、「人工呼吸器治療の生存率が78.5%以下だとして、人工呼吸器よりもECMOの方が優れている」という帰無仮説が1%有意水準で棄却されるという分析をも、ミシガン・チームは下している(cf. Bartlett et al., 1985, p.485)。しかし、治験終了後には患者のランダムな振り分けが行われていない以上、このような検定結果も統計学的には正当なものとは言えない、インフォーマルなものにすぎない。

劣を判定するために十分なものであったのだろうか。この問いに答えるためには、その後のECMOの普及や、PPHNの治療法一般の変遷を踏まえ、いわば結果論的に考えてみるのも一つの手であろう。

ECMOは現在では、重度のPPHNに対する有効な治療法として広く認められ、北米・西ヨーロッパ・日本など少なからぬ国や地域の小児医療センターにおいて用いられるに至っている³⁸⁾。またそれに伴い、ECMOデータベースも、2004年7月の時点で、呼吸器疾病を患った新生児に限っても19,061人、(心臓病を患った新生児や、小児、成人への適用例を含め) トータルでは28,985人の症例を蓄積するにいたっている³⁹⁾。しかし一方で、ECMOより侵襲的でなく、高価でもない代替治療法も相次いで開発された。例えば、改良型の人工呼吸器、界面活性剤、従来より効果的な血管弛緩剤である一酸化窒素といった手段を用いた治療法⁴⁰⁾が、ちょうどハーバード治験と踵を接するように、1990年ごろから臨床に応用され出し、ECMOの役割はこれらの代替治療が効かなかった場合の「最後の手段」としてのそれに限定されるようになった⁴¹⁾。その結

38) 最近のある論文では、深刻な呼吸障害を持った新生児に対するECMOの使用は「十分に確立されている」と書かれている。Cf. Skarsgard, E., et al., Venovenous Extracorporeal Membrane Oxygenation in Neonatal Respiratory Failure: Does Routine, Cephalad Jugular Drainage Improve Outcome?, *Journal of Pediatric Surgery*, 2004, vol.39, no. 5, p.672.

39) Extracorporeal Life Support Organization, ELSO Registry Report, International Summary, Ann Arbor, Michigan: ELSO; July 2004. Retrieved September 17, 2004 from <http://neonatal.peds.washington.edu/NICU-WEB/ecmo2.stm>

40) 改良型の人工呼吸器とは、速いテンポで空気を送り込む高振動数呼吸器のこと。このタイプの呼吸器は、新生児の肺を傷つけることが少ない。また界面活性剤の投与は、肺胞表面上の界面活性物質の不足から起こる呼吸窮迫症候群に対して効果的であり、胎便吸引症候群に対しても胎便を洗い流すという役割を果たす。一酸化窒素は大気汚染源となる有害物質であるが、微量なそれは、人間の体内における様々な生理現象を司っており、筋肉を弛緩させ血圧を下げる働きも持つ。気管から入った一酸化窒素は、肺にすばやく吸収されるので、肺血管の筋肉だけを弛緩させ、全身の筋肉には影響を及ぼさない。即ち、狙った部位だけに効果を集中させることができる「魔法の弾丸」である。Cf. Wash-Sukys, M., et al., 'Neonatal ECMO: Iron lung of the 1990s?', *The Journal of Pediatrics*, 1994, vol.124, No. 3, pp.427-430. Tulenko, 2004.

41) 90年代半ば以降のECMOを設置した医療機関数の減少傾向については、つぎのような医療経済的な要因を指摘する向きもある。一時は、高価で大規模な医療機器を導入することが、患者獲得の有効な戦略であるとして、各中核病院はECMOのような装置を争って設置した。しかし90年代初頭の「ヘルスケア・リフォーム」の流れの中で、医療の費用対効果を重視する傾向が強まるにつれ、医療現場における「軍拡競争」は終わり、ECMOセンターの数も減った、というわけである。例えば、Wash-Sukys, et al., 1994. このような医療経済的な視点は、「ECMO

果、ECMOの年間当りの臨床数、ECMOセンター数とも、それぞれ1992年と1993年をピークに頭打ち、ないし漸減傾向に転じることになる。しかし「最後の手段」としてのECMOの性能は相変わらず高かった。ECMO治療の生存率は、PPHNの多様な合併症に応じて様々であるものの、全体として見れば、90年代半ばにおいて80%台から90%台、2004年時点でも50%台から90%台の高率を維持していたのである⁴²⁾。(90年代半ば以降の見かけ上の成績悪化は、代替治療の普及によって、より重篤な患者のみがECMOにまわされるようになったため、というのが大方の見解である⁴³⁾。)

同様の傾向はチルドレンズにおいても見られる。チルドレンズにおいても、改良型人工呼吸器と一酸化窒素が、それぞれ1991年と1992年から用いられるようになり、それに伴い、ECMOの臨床例は減少した。それでもECMOは、PPHN治療の信頼できる最後の切り札として用いられ続けているのである^{44,45)}。

このように、ECMOがPPHN治療の最終兵器として定着するに至った一連の流れの中で注目すべきは、ハーバード治験以降、同種のRCTが2件実施されていたという点であろう。それらの中でも重要なのは、1993年から1995年にかけてイギリスで行われた(現在のところ)最後の治験である。この治験はハーバード治験をも含めた、それ

の普及において、ハーバードRCTの結果をも含め、臨床治験が与えた「証拠」が、どれだけの役割を果たしたのか」を考える際にも有効であろう。

42) Roy, B.J., et al., The Changing Demographics of Neonatal Extracorporeal Membrane Oxygenation Patients Reported to the Extracorporeal Life Support Organization (ESLO) Registry, *Pediatrics*, 2000, vol.106, No.6, pp.1334-1338. Extracorporeal Life Support Organization, ELSO Registry Report, International Summary, Ann Arbor, Michigan: ELSO; July 2004. Retrieved September 17, 2004 from <http://neonatal.peds.washington.edu/NICU-WEB/ecmo2.stm>.

43) Roy et al., 2000.

44) Wilson, J.M., et al., ECMO in Evolution: the Impact of Changing Patient Demographics and Alternative Therapies on ECMO, *Journal of Pediatric Surgery*, 1996, vol.31, pp.1116-1122.

45) 今日では、さらに、呼吸窮迫症候群の発症を防ぎ、PPHNの症状をやわらげる副腎皮質ステロイドの出生前投与、横隔膜ヘルニアに対する出生直後の外科手術、超音波画像解析や羊水穿刺などによる出生前診断の普及によって、ECMOの「出番」は一層減りつつある。しかし、相変わらずECMOはこれらの治療法が効果がなかった場合の「最後の切り札」であり続けている。Cf. Tulenko, 2004. なお、ECMOについての最新の情報は、例えば、コクラン・ライブラリーにおいても見出すことができる。The Cochrane Library, Issue 3, 2004, <http://www.cochrane.org//cochrane/revabstr/ab001340.htm>.

以前の治験と比べてはるかに規模の大きいものであり、55の医療機関において、185人の患者を対象として実施された。また、この治験においてアダプティブ・デザインは一切採用されず、結果としてほぼ同数の患者がランダムに治療群と対照群に振り分けられた。ただ、ここでも、ECMOの優位は実験の初期段階から明白となり、最終的には、実験をモニターしていた検討委員会の勧告によって、予定の患者数に達する以前に治験は中断された。しかし中断された時点で、人工呼吸器に比べてECMOの方が優れているという結論が、統計的に導き出されたのである⁴⁶⁾。

もちろん、これらの「その後」の治験が本当に必要であったかどうかとも検討を要する事柄である。しかし結果として、ハーバードの治験は、それに引き続いた同種の治験を阻止することができなかったのである。このことは、その治験の結果が、同様の比較実験を「打ち止め」とするだけの決定的なものではなかったのではないか、そしてそのような不十分な結果をもたらしたという点で、その治験のデザインに問題があったのではないか、という疑念を生む⁴⁷⁾。ECMOの普及と、「最後の手段」としての定着にさいしてRCTの結果が何ほどかの影響をたとえ与えたとしても、その中でハーバード治験が一定の役割を果たしえたかどうかは－これまた控えめに言って－疑問の余地が残るのである。ECMO治療の歴史を概観した論文で、ミシガン治験について、それは統計的テストとしては妥当であったが、医学界全体を説得することには失敗した、と評したものがある⁴⁸⁾。それと同じことが、ハーバード治験についてもあてはまりうるのである⁴⁹⁾。

46) UK Collaborative ECMO Trial Group, UK Collaborative Randomised Trial of Neonatal Extracorporeal Membrane Oxygenation, *Lancet*, 1996, Vol.348, July 13, pp.75-82.

47) 「同種の実験を打ち止めにできなかったことは、治験のデザインに問題があったことを示唆している」という考えは、ミシガン・チームによっても表明されている。すると当然、彼らとしては、自らの治験の方法論的な正当性を訴える限りは、その治験の後に実施された同種の実験であるハーバード治験を、不必要で、従って倫理にもとるものと見なさざるをえないのである。Cf. Bartlett and Cornell, 1991, p.64.

48) Soll, R.F., Neonatal Extracorporeal Membrane Oxygenation - a Bridging Technique, *Lancet*, 1996, vol.348, no.9020, 13 July. 正確には、「ECMOは統計的テストには合格したが、世界の医学界を納得させることには失敗した」という表現である。もちろん、当のミシガン・チームは、自らの治験の結果、多くの医者や医療機関がECMOを使うようになったことを強調している(Cf. Bartlett and Cornell, 1991, p.64.)。

49) しかし、少なくともチルドレンズでは、治験の結果を受けて、改良前の人工呼吸器治療でも治らなかった全ての患者に対して、ECMO治療が施されることとなった。Cf. Wilson, J.M., et al., 1996.

ハーバード治験に対する方法論的な疑念は、以上のものに止まらない。その治験のデザインの細部の妥当性もまた問題になりうるのである。例えば、アダプティブ・デザインもその一つである。先に指摘したように、ハーバードの研究者たちは、ECMO優位の事前情報を踏まえ、倫理的な懸念から、一定のアダプティブ・デザインを採用せざるを得ないと判断した。しかし、他方、彼らが先行するミシガン治験の結果に不満を抱き、新たな治験の必要性を感じたのは、ミシガン治験の対照群が1人の患者しか含んでいなかった点にあった⁵⁰⁾。彼らに言わせれば、ミシガン治験で採用されたアダプティブ・デザインは、治験の初期の結果に過敏に反応し過ぎ、結果として十分な対照群を与えることができず、ひいては（高い検出力を得る代償として）極めて高い有意水準をもたらさざるをえなかった点に問題があった。このような反省を踏まえ、ハーバードの研究者たちは、初期の結果により鈍感にしか反応せず、より大きな対照群を与えてくれるであろうアダプティブ・デザインを採用したのである。しかし、彼らもまた、後にイギリス大規模治験を実施した研究者から、自分たちがミシガン治験に下したのと同じ批判を受けることになる。イギリスの研究者たちは、ミシガン治験に加えてハーバード治験の結果をも、アダプティブ・デザインの採用によって十分な数の対照群を確保できなかったという点で「説得力に欠ける (inconclusive)」と批判したのである⁵¹⁾。このように、アダプティブ・デザインを採用すべきかどうか、採用すべきとして、どのようなデザインを採用すべきか、という観点からも、ハーバード治験には、検討の余地があるのである⁵²⁾。

50) Ware and Epstein, 1985, p.850f. O'Rourke, et al., 1989. p.958.

51) UK Collaborative ECMO Trial Group, 1996, p.75.

52) ECMO 治験以前に、データ依存型治療割り当て規則が、実際の実験に用いられた例はほとんどなく、ミシガン治験とハーバード治験が、この種のアダプティブ・デザインを実際の治験で用いた、それぞれ最初と二番目のケースではないかと言われている(cf. Armitage, P., The Search for Optimality in Clinical Trials, *International Statistical Review*, 1985, vol.53, pp.15-24. Lin, D.Y., and L.J. Wei, Comment, *Statistical Science*, 1989, vol. 4, No. 4, p.324.)。このような事情で、そこで考案・採用された特定のアダプティブ・デザインの是非のみならず、そもそも、場合によっては極端に大きさが異なった治療群と対照群をもたらしかねない、データ依存型治療割り当て規則一般の採用の是非が議論の対象となったのである。またアダプティブ・デザインのように、たとえ統計的方法として妥当であったとしても、めったに使われない「奇妙な」方法は、必ずしも統計学に明るいとは言えない医学者・医者には「胡散臭く」映り、ひいては、そのような方法によって得られた治験の結果は医学界を「説得する」力を欠くという指摘もあ

また、ハーバードの研究者たちが治験の途中でランダムマイゼーションを事実上、打ち切ったことも、方法論的には問題となる点である。先に述べたように、ランダムマイゼーションは、ハーバードRCTで用いられたネイマン-ピアソン型仮説検定や区間推定といった古典統計学の方法論の適用にとって不可欠の作業である。その不可欠の作業を中断し、その後に得られた結果をも含め、古典統計学の手法を適することは果たして妥当なのであろうか。

さらに、死者の数の設定が、統計学的には正当な指標とは言えない擬似検出力にもとづいて決定されていたこともまた、ハーバード治験の問題点の一つであると言える。先に触れたように、この擬似検出力は、ハーバード治験の本来の検出力とは、厳密に言って、何の関係もない数値に過ぎない。また「死者の数を増やせば増やすほど数値が上がり、その上がり方は鈍くなる」という死者の数の設定の根拠となった擬似検出力の性質は、本物の検出力においては見出せない。死者の数の設定は、患者の生死を分ける、倫理的にも極めてデリケートな事柄である。だが倫理を度外視しても問題は残る。ハーバードの研究者たちがミシガン治験を評した際に指摘したように、有意水準や検出力によって実験のデザイン、特にサンプルサイズ（ないしは死者の数のように、それに影響を与える要因）を決定することは、たとえアダプティブ・デザインを採用したとしても、できるだけ確保されるべき、ネイマン-ピアソン型仮説検定の基本的な手続きである⁵³⁾。それだけに、その決定が、一見、統計学的に妥当な手続きを踏んでいるかのように装われつつ、その実、（厳しく言えば）ナンセンスとも言える仕

る(cf. Ware, J., and M.F. Epstein, 1985, p.849, Miké, V., et al. Neonatal Extracorporeal Membrane Oxygenation (ECMO): Clinical Trails and the Ethics of Evidence, *Journal of Medical Ethics*, 1993, vol.19, p.213.)。このような見方からすれば、本文で言及した「統計的テストとしては妥当だが医学界全体を説得することに失敗した」というミシガン治験に対する評や、イギリスの大規模治験の研究者によるアダプティブ・デザインに対する批判、さらには、コクラン・ライブラリー(The Cochrane Library)における「アダプティブ・デザインを採用したことで、その実験結果を解釈することが困難になっている」というハーバード治験に対する否定的な評価などは、「アダプティブ・デザイン・アレルギー」とでも呼べる、少なくとも統計学的には合理的ではない反応であることになる。Cf. The Cochrane Library, Issue 3, 2004, <http://www.cochrane.org/cochrane/revabstr/ab001340.htm>.<http://www.cochrane.org/cochrane/revabstr/ab001340.htm>.

53) Ware and Epstein, 1985, p.851.

方で下されているのは、方法論の観点から大いに問題なのである。

分かれた評価

以上、ハーバードのECMO治験にまつわる倫理的・方法論的な疑念をいくつか挙げたが、実際、ハーバードのRCTは、その研究者の予想通り（ないしは、彼らも認めたように、予想以上の⁵⁴⁾）その倫理的・方法論的妥当性を巡る論争を巻き起こしたのである。

まず、ハーバード治験の結果が専門雑誌に発表される4ヶ月前、1989年8月7日の地元『ボストン・グローブ』紙に、「ハーバードのRCTが、専門家の間で議論を巻き起こしている」という記事が出た。この記事の3日後には、アメリカ統計学会の年次総会でハーバード治験に関する異例の公開討論会が開かれた。この公開討論会には、『ボストン・グローブ』紙の記事でインタビューされていた専門家も含め、11人の医療統計学者が、それぞれの立場からの議論を展開し、その結果は、専門雑誌に40ページ以上に亘る論文集として掲載された⁵⁵⁾。また、その後ほどなくして、このハーバードのECMO治験をケース・スタディとして取り上げ、臨床治験の倫理と方法論を論じた論文も2本出ている⁵⁶⁾。さらに、これら直後の反応だけでなく、ECMOが普及していくその後の過程においても、ハーバード治験に対しては繰り返し、様々な角度から検討が加えられてきた。先に紹介したイギリスの大規模治験の論文における、批判的な言及もその一つである。

これらの場に出された意見を、ここでいちいち取り上げることはしない。とりあえず、先に挙げておいた懸念をも含め、色々な議論が多様な立場から表明されただけ言っておこう。また私自身の考えについては、本論での議論を踏まえた上で、後で改

54) Knox, R.A., 1989.

55) Ware, J.H., et al. Investigating Therapies of Potentially Great Benefit: ECMO, *Statistical Science*, 1989, vol. 4, no. 4, pp.298-340.

56) Lantos, J., and J. Frader, Extracorporeal Membrane Oxygenation and the Ethics of Clinical Research in Pediatrics, *The New England Journal of Medicine*, 1990, vol.323, No. 6, p.409-413. Miké, et al., 1993, pp.212-218.

めて述べたいと思う。ただ、ここで注目すべきは、全体として見れば、専門家の間で賛否両論が分かれたという事実である。ハーバードのECMO治験に対しては、いくつかの点で明らかに倫理に反するという見解もあった一方で⁵⁷⁾、それは「細心の注意を払って計画された、配慮の行き届いた科学的実験であった」⁵⁸⁾という評価もなされたのである⁵⁹⁾。

加えて、ハーバード治験は決して孤立したケースではなかった。先に見たように、ECMO優位の事前情報の中で、結果として利用可能なECMO治療を受けずに患者が亡くなった治験は、これ以外にも3件あったのである。さらに、これらのECMO治験を別としても、人の命にかかわる病気に対して、ある治療法が有効であることを示す事前情報はあったものの、それを裏打ちするRCTの結果がない（ないしは十分に説得的なRCTの結果がない）という状況下で、あえて同時対照群を備えたRCTが実施されたケースはいくつもある。例えば、ミシガン治験やハーバード治験と同時期に実施された、神経管閉鎖障害の予防に対する葉酸の効果を調べるためのRCTも、そのような事

57) 例えば、ロイヤルは、赤ん坊は一人たりとも人工呼吸器治療にランダムに割り当てられるべきではなかったとし、また、この治験におけるRC法の採用を「重大な過ち(a grave error)」と表現している。Cf. Royall, R., Comment, *Statistical Science*, 1989, vol. 4, No. 4, p.318.

58) Begg, C., Comment, *Statistical Science*, 1989, vol. 4, No. 4, p.320.

59) 例えば、ECMO治験におけるRC法の採用に関しても、様々な議論があった。批判の一つの焦点は、人工呼吸器治療が、治験以外でも自動的にあてがわれていたという意味で「標準的」な治療と言えるかどうかである。例えば、ロイヤルは、ECMOが実際に臨床に応用され、それを見越した上での外部からの転院が相次ぐような中核的な医療機関においては、人工呼吸器治療を「標準的」な治療法とみなすことはできず、従って、人工呼吸器治療を受けた患者の親からインフォームド・コンセントを取らないというデザインは正当化できないという趣旨の批判を行った。Cf. Royall, R.M., *Ethics and Statistics in Randomized Clinical Trials*, *Statistical Science*, 1991, vol. 6, No. 1, p.60. この批判は、ミシガンECMO治験におけるRC法の適用に対するものだが、同様の議論は、そのままハーバード治験にも当てはまる。また、人工呼吸器治療法の細部は各医療機関によってまちまちであり、かつ、つねに部分的に更新されつづけていた。従って、そのうちの特定の方法を固定して、インフォームド・コンセントが不要な「標準的」な治療法と見なすことはできない、という批判もなされた。Cf. Lantos and Frader, 1990, pp.400, 411. ところが他方で、RC法が倫理的な配慮にもとづいて採用された例として、ハーバードのECMO治験を挙げ、対照群の親から無用の心配の種を除くというハーバードの研究者の考えも、いささかの批判的なニュアンスをも込めずに紹介している例も見られるのである。Cf. Torgerson, D., and M. Roland, *Understanding Controlled Trials: What Is Zelen's Design?*, *British Medical Journal*, 1998, 21 February, vol.316, p.606.

例に当たるのである⁶⁰⁾。

60) 神経管閉鎖障害 (Neural Tube Defect: NTD) とは、脳や脊椎に起こる先天性の異常の一種。通常は、胎児の発達初期に神経板の両端が癒合して管状の組織、即ち神経管 (neural tube) が形成され、それがその後、脳や脊椎などの中枢神経系へと成長する。この神経板の癒合が不完全なまま両端が閉じず、神経管が正常に形成されないことで、結果として脊椎が左右に分かれたり、脳が形成されなかったりする症状を一括して神経管閉鎖障害 (NTD) と呼ぶ。この先天異常の発生の割合を下げるためには、ビタミンBの一種である葉酸 (folate) や、葉酸を含むビタミン類の妊娠初期における摂取が有効であることを示す実験や観察の結果が、80年代前半までにいくつか得られていた。例えば、おのおの500人前後の妊婦を、葉酸を含むビタミン類を摂るグループと摂らないグループに分け、それぞれ0.7%と4.7%というNTDの発症率を得た、大規模治験の結果もそうである。(Cf. Smithells, R.W., et al., Apparent Prevention of Neural Tube Defects by Vitamin Supplementation, *Archives of Diseases in Childhood*, 1981, vol.56, pp.911-918. Smithells, R.W., et al., Vitamin Supplementation and Neural Tube Defects, *Lancet*, December 19/26 1981, p.1425. Smithells, R.W., et al., Further Experience of Vitamin Supplementation for Prevention of Neural Tube Defect Recurrences, *Lancet*, May 7 1983, pp.1027-1031.) しかし、この実験では、ランダムなグループ分けは行われなかった。つまり、この大規模治験はRCTではなかったのである。一方、同時期に比較的少数の妊婦を対象としたRCTの結果も報告されていた。このRCTにおいては、葉酸を摂った44人の妊婦から0件、採らなかった65人から6件のNTDの発症が確認され、(対立仮説を置かない) フィッシャー型の仮説検定で、「葉酸は効果なし」とする帰無仮説が4%の有意水準で棄却された。(正確に言えば、ランダムに葉酸摂取グループに振り分けられた妊婦から2件のNTD発症が得られたが、これらは葉酸摂取の指示に従わなかったケースとして、葉酸不摂取グループに算入されている。) (Cf. Laurence, K.M., et al., Double-blind Randomized Controlled Trial of Folate Treatment Before Conception to Prevent Recurrence of Neural Tube Defects, *British Medical Journal*, 1981, vol.282, pp.1509-1511.) これらの結果を受けて、専門家の意見は分かれた。ある者は、葉酸を含むビタミン類の有効性を示す「証拠」は十分に得られたとし、同時対照群を備えた治験のこれ以上の実施は不必要だと主張した。(Cf. Edwards, J.H., Vitamin Supplementation and Neural Tube Defects, *Lancet*, January 30 1982, pp.275-276.) この主張は一部ジャーナリズムにも支持された。(Cf. *Yorkshire Post*, Doctors Hit at Spina Bifida Trail, 7 July 1982, *The Guardian*, Wasted Years, Damaged Lives, 13 December 1982.) 他方、これらの結果は証拠として不十分だと見なす研究者もいた。例えば、先の大規模治験は、ランダム化を欠いている点で、RCTは患者の数が少なすぎるという点で、それぞれ問題があるとする指摘がなされたのである。(Cf. Wald, N.J., and P. Polani, Neural Tube Defects and Vitamins: the Need for a Randomized Clinical Trial, *British Journal of Obstetrics and Gynaecology*, 1984, vol.91, pp.516-523.) そして、このように専門家の間でコンセンサスが得られていない以上、RCTの実施は倫理的に許されるとして(cf. Wald and Polani, 1984, p.520f)、イギリスの医学研究評議会はRCTが必要との結論を出し、1983年7月から1991年4月にかけてイギリスを中心とする7カ国の施設で大規模なRCTを実施した。このRCTでは、過去にNTDを負った胎児を妊娠した経験を持つ1817人の妊婦たちが、葉酸を与えられる治療群と与えられない対照群にランダムに振り分けられた。(詳しく言えば、葉酸のみ、葉酸とその他のビタミン類、その他のビタミン類のみ、空のカプセル、

また、ハーバード治験そのものも、倫理的・方法論的な側面をも含め、二つの (Institutional Research Board) の承認を経て実施されたものである⁶¹⁾。ここで言う IRB とは、治験の妥当性を事前に審査する、研究機関ごとに設けられた、専門家からなる第三者機関である。今日、欧米の主要な医学雑誌では、IRB の承認を経てなされた治験の報告であることが、論文掲載の条件とされるケースが増えつつあるが、全てがそうであるわけではない⁶²⁾。実際、ハーバード治験の結果が掲載された雑誌 (アメリカ小児科学会の機関誌である『小児科学 (Pediatrics)』) においても、当時、IRB の承認は要求されていなかった⁶³⁾。その意味で、ハーバードの治験は、制度面から見て、むしろ必要以上に模範的な実験であったと言えるのである。

さらに、先に触れた『ボストン・グローブ』紙の記事以降、マスコミが一批判的であれ肯定的であれ—この治験を取り上げることはなかった。また、治験の対象となり、人工呼吸器治療で亡くなった4人の患者の家族が、医師や病院を相手取って、刑事や民事の訴訟を起こすといった動きもなかった。もちろん、この治験を行ったハーバー

がそれぞれ与えられる4つのグループが設定された。) その結果、治療群と対照群における NTD の発生率は、それぞれ1.0%、3.5%となり、前者の後者に対する比について求められた95%信頼区間推定値は、0.12から0.71までと算出された。このことは、同じ妊婦たちに対するランダム振り分けを無限回繰り返した場合、その95%の結果において、治療群の発生率は、対照群の発生率の12%から71%の範囲に収まることを意味する。また、同じ生のデータに対して、逐次検定も行われ、「両発生率は等しい」とする帰無仮説が、有意水準5%、検出力75%をもって棄却された。この棄却結果を受け、イギリスのECMO大規模治験と同様、検討委員会の勧告によって治験は中止された。(Cf. MRC Vitamin Study Research Group, Prevention of Neural Tube Defects: Results of the Medical Research Council Vitamin Study, *Lancet*, 1991, vol.338, pp.131-137.)

61) チルドレンズのIRBと、治験のもう一つの舞台となった、ブリガム・アンド・ウィメンズ病院のIRB。Cf. O'Rourke et al., 1987, p.959, Ware, 1989, p.305. ちなみに、ブリガム・アンド・ウィメンズ病院もまたハーバード大学医学部の関連病院で、全米トップクラスの治療・研究水準を誇っている。

62) ハーバード治験の約10年後になされた調査によると、102の主要な英文雑誌のうち、IRBないしは倫理委員会の承認を条件として課しているのは約半数に留まっており、4分の1は倫理的ガイドラインすら設けていない。Cf. Amdur, R.J., and C. Biddle, Institutional Review Board Approval and Publication of Human Research Results, *Journal of American Medical Association*, 1997, vol.277, pp.909-914.

63) Manuscript Preparation: Instructions for Authors, *Pediatrics*, 1989, p.A5. ちなみに同誌は、今日でも、IRBや倫理委員会の承認を論文掲載の条件に課していない。Instructions for Authors, *Pediatrics*, December 2004, vol.114, no.6, p.A5.

ドの研究者たちは、その後、何の支障もなく研究活動が続けている⁶⁴⁾。

つまり、ここにはマッド・サイエンティストも悪徳医者も登場しないのである。

ハーバードのECMO治験は、明らかに、例えばナチスや日本軍の731部隊による人体実験やタスキギー梅毒実験のような、その責任者が罪に問われてしかるべき「邪悪」な人体実験ではなかった⁶⁵⁾。その実施は完全に合法的であり、制度面から見てもなんらの瑕疵もなかったのである。その倫理的・方法論的妥当性を巡っては論議が巻き起こったが、その議論も、通常の「学問的」な議論の枠内に概ね収まるものであった。確かに、ハーバードのECMO治験は、その倫理的・方法論的な妥当性に関して異論が続出したという意味で、『ボストン・グローブ』紙に載った、ある専門家の言を借りれば「極端な (extreme)」ケースであるとは言える。しかし、それは現代アメリカにおける科学的活動、医療活動の一般的なスタンダードに照らしても、常軌を逸した「異常 (abnormal)」な治験であるとは言えないのである。

しかし、だからこそ問題の根は深いと言える。異常なケース、病理的なケースであれば、問題はある意味、局所的である。我々は、それに係った特定の人物をその他大勢の人たちから切り分けた上でその責任を問い、そのようなケースを許した制度の問題点—そのような問題点があるとして—を同定し、その部分を適切に再設計すればよいのである。その場合、大部分を占める「正常 (normal)」なケースは、「異常さ」を際立たせる「背景」ないし「地」としてのみ機能し、それに対して必ずしも批判の目は向けられない。

64) 例えば、ハーバード治験の論文の第一著者は、2004年末現在、ハーバード大学医学部やその関連病院の外郭団体の治験部門の責任者 (Director) であり、最終著者は、同学部小児科の準教授である。またその治験の統計デザインを担当した医療統計学者は、現在、ハーバード・スクール・オブ・パブリック・ヘルスの学部長 (Dean) を務めている。

65) ナチスや731部隊による人体実験については説明の要はないだろう。タスキギー梅毒実験とは、アラバマ州タスキギーで、梅毒の（積極的な治療を行わない場合の）自然経過 (the natural course of disease) を観察するために1932年から40年以上にわたって行われた人体実験である。この実験においては、「無料の特別治療を施す」という広告の下に集められた約400人の梅毒患者（全て貧しい黒人男性であった）に対して、梅毒に対するペニシリンの効果が広く認められた40年代以降も、実験開始当時用いられていた重金属投与治療のみを与え続け、1972年に内部告発によって事実が公になるまで、科学的には貴重な梅毒の自然経過についてのデータが集められつづけたのである。Cf. *Twenty Years After: The Legacy of the Tuskegee Syphilis Study*, *Hastings Center Report*, November-December 1992, pp.29-40.

しかし、「4人の新生児が、利用可能であり、かつ有望と思えた治療法を受けられな
いまま死亡した」という冒頭の事例は、極端なケースとは言え、現在のスタンダード
では「正常」な科学的活動・医療行為とされるRCTにおいて発生しているのである。
このことは、正常な科学的活動・医療行為を支えているものの考え方の中に、このよ
うな形での4人の死を準備するものが含まれていた可能性を示している。冒頭に立て
た問いにこだわり続ける限り、我々はもはや、現代科学や医学の制度や概念的枠組み
を、「自明」で「当たり前のもの」とみなし続けることができない。ハーバードの
ECMO治療による4人の死者は、我々に、現代の科学・医療の実践において正常とさ
れている事柄、自明視されている前提に対する、根本的な問い直しを求めているので
ある。その意味で、正常の枠内に辛うじて収まる「極端なケース」は、その「正常さ」
に潜む問題点の存在を－いささかグロテスクな仕方ではあるが－我々に告げるのであ
る。『ボストン・グローブ』紙のインタビューに応じた、くだんの研究者の言を再び借
りれば、ハーバードのECMO治療は「極端なケース」であるからこそ、「問題を劇的に
する」という役割を果たしうるのである⁶⁶⁾。

医療の科学化

「4人の死」を準備した、現代の科学・医療の実践において自明とされている事柄を
見定めるために、ここでは、ハーバードの研究者たちをしてECMO治療の実施へと駆
り立てた動機は何であったのかを、改めて確認することから始めたい。

当時の状況をもう一度振り返っておこう。1975年、ECMO治療によるPPHN患者の
最初の治療成功例を出したパートレットは、1980年以降、ミシガン大学で着々と
ECMOの臨床例、成功例を重ねるとともに、それをデータベース化して公開すること
も始めていた。また彼はPPHNに対するECMOと人工呼吸器の効果を比較する最初の
RCTを行い、患者数1という極端に小さなグループではあったが、ともかく同時対照
群を備えた比較実験で、ECMOの優位性を示す統計的結果を得ていた。一方、チルド
レンズでも1984年以降、ECMOによる成功例を重ねていた。このような中でチルドレ

66) Knox, 1989におけるロイヤル (R. Royal) の発言。

ンズにおけるハーバードの研究者たちは、ECMO優位の「予想」を持つにいたっていた。言い換えると、彼らは、ECMOと人工呼吸器治療の間には優劣の差が存在すると考えていたのであり、そう考えたからこそ、彼らの治験においてアダプティブ・デザインを採用したのである。しかし、彼らはこの「予想」を、「よい証拠」にもとづいた「知識」であるとは見なしていなかったと言える。「ECMOが人工呼吸器より優れた効果を持つ」という「予想」が「知識」と言えない以上、この場合における医療平衡は破られていないことになる。その意味で、ハーバードの研究者にすれば、彼らのRCTは、ブラッドフォード・ヒルの標準的なガイドラインに照らして、倫理的に妥当であると言えるものだったのである。

このような状況下で、ハーバードの研究者たちは、自分たち、そしておそらくは医学界の少なからぬ人たちが抱いていたECMO優位の「予想」を「よい証拠」によって裏付け、「知識」と化すことを目指して、さらなる治験の計画を立てたのである。このことはまた、彼らが、すでに自分たちも採用し、医学界でも標準化されつつあった、「重度のPPHN患者にはECMO治療を施す」という治療選択を、単なる予想ではなく、「証拠」にもとづいたものにしようと目論んでいたことを意味する。

彼らのこのような動機の背後には、「治療の選択、ひいては医療行為一般は、証拠にもとづかなければならない」という考えが横たわっていることは容易に見て取れる。このような考えは、なにも彼ら特有のものではない。大きく歴史を振り返ってみると、このような考えは、近代以降、繰り返し唱えられてきた、「医療・医学の科学化」のプログラムの一環をなす、医学界に広く見られる考えなのである。例えば、19世紀フランスの生理学者で、医学の実験科学化を唱えたクロード・ベルナール（Claude Bernard; 1813~78）もまた、その著書『実験医学序説』（*Introduction à l'étude de la médecine expérimentale*, 1865）で、新しい治療法の効果は、動物実験、人体実験において科学的に実証されねばならないと明確に述べている（3.4.3）。ちなみに「医療・医学の科学化」と言うとき—その科学化に賛成する側であっても反対する側であっても—科学化される以前の医学、ないし科学化されざる医学とは、一言で言えば、^{わざ}技・アート（art）である。「アートからサイエンスへ」というスローガンが繰り返し唱えられたからといって、それが全面的に実現するとは限らない。特に、生理学や病

理学、病因学といった、医学の他の分野に比べても、治療法の開発やその効果・副作用の研究を、実際に患者に対する臨床的応用と結びつけて研究する治療学 (therapeutics) ないし臨床医学 (clinical medicine) の科学化は遅れ、そのプロセスは現在でも進行中であると言える。例えば、80年代末にカナダとイギリスで始まり、瞬く間に他国にも広まっていった、「証拠に基づいた医学 (Evidence Based Medicine)」の運動、いわゆる^{イービーエム}EBM運動もまた、そのような臨床医学や医療の科学化のプロセスの最新バージョンと見なせるのである。ちなみに、EBM運動とは、日常の医療行為を、有名医の「権威」や個々の医者「の」に頼って行うのではなく、「科学的証拠」にもとづかしめるべきであるとする運動であり、そのために、各種の医学情報をオンライン化して開業医に提供することを一つの柱としている⁶⁷⁾。

証拠、その資格と基準

治療効果の判定や治療選択を「証拠」にもとづかせようとする、医学なканずく臨床医学の科学化の運動の中で、常に問題となってきたのが、「どのようなデータを「証拠」と見なしてよいか」というデータの証拠としての「資格」の問題と、「どのようなデータがより信頼できる「質」のよいデータなのか」という証拠の「質」の問題であった。EBM運動においても、RCTの結果をトップにランクした「証拠の質ないし強さ (strength) のヒエラルキー」なるものが提唱され、医療を単に証拠にもとづけるだけではなく、「より質のよい証拠」、「ヒエラルキーにおいてより高い地位を占める証拠」にもとづけるべきであると主張されている⁶⁸⁾。

このような観点は、ハーバードの研究者をECMO治験に向かわしめた動機において

67) EBM運動については、例えば以下の文献を参照。Sackett, D., et al, Evidence Based Medicine: What It Is and What It Isn't, *British Medical Journal*, 1996, vol.312, pp.71-72.

Mayer, D., Essential Evidence-Based Medicine, 2004, Cambridge U.P., Cambridge England.

68) 例えば、U.S. Preventive Services Task Force, *Guide to Clinical Preventive Services*, 2nd and 3rd ed., 2000-2002, International Medical Publishing, McLean VA, p.862. ここで示されているヒエラルキーとは、質の高いものから、単独の大規模なRCTの結果、複数の比較的小規模のRCTからのメタアナリシスの結果、コーホート研究、ケースーコントロール研究、事例研究の順となっている。

も確認することができる。先に、彼らが、事前情報から ECMO 優位の予想を立てた上で、その予想の「証拠立て」を意図したことを指摘した。その際、問題とされたのは、単なるデータの「数」ないし「量」ではなく、データの証拠としての「資格」や「質」なのである。なぜか。ハーバード RCT によって得られる ECMO の臨床例は、治験のデザインからして、せいぜい 30~40 例止まりである。これは、それまで蓄積され、今後さらに加速度的に増加するであろうと当時予想された ECMO データベースの臨床例数からすれば、大した数ではない。データの数ないし量という観点からのみ見れば、全米データベースが構築されている以上、わざわざ治験を行うことにさほどの意味はないのである。このことは、ハーバードの研究者が、データの量の多寡を、必ずしもそのまま、そのデータの証拠としての資格の有無や質の高低に結びつくものと見なさず、データの量を増やすためと言うより、質の高い証拠を得るために、あえて治験を計画したことを意味する。データの「量」より、その証拠としての「資格」ないし「質」が、彼らの関心事であったのである。

このことはまた、彼らが、ECMO 優位を示す、それまで蓄積されていた全米データやミシガン治験の結果といった事前情報を、「証拠」ないしは「質の高い証拠」とは見なしていなかったことをも意味する。他方、彼らは、自分たちが行う RCTこそは、これまで手に入らなかった「証拠」ないしは「質の高い証拠」を与えることができると考えていたのである。

このような態度の背後にあるのは、「あるデータが科学的証拠と見なせるかどうかの基準」や、「証拠の質のよしあしを決定する基準」である。ハーバードの研究者は、彼らなりの基準を用いて、ECMO データベースの情報や治験開始時に蓄積されていた事前情報を、少なくとも質の高い、よい証拠であるとは見なさなかったのである。ちなみに、彼らの基準の重要なファクターは、十分な数の同時対照群を備えており、結果として低い有意水準と高い検出力を兼ね備えているかどうかであった。ECMO 全米データベースを用いて ECMO と人工呼吸器の成績を比べる場合、比較対象となる人工呼吸器治療の臨床例は既存対照群とならざるを得ない。また、先にも触れたように、ミシガン治験は、十分な数の同時対照群を備えているとは言えず、また十分に小さな有意水準を持つものとも言えないというのがハーバード研究者の判断であった。

これら証拠の資格や質についての基準－それをここでは一括して「証拠の基準 (criteria of evidence)」と呼ぶことにしよう－の背後には、科学における実験や観察で得られたデータに、一定の質を備えた証拠としての資格を与える方法論が控えているといえる。このような方法論は、証拠理論 (theory of evidence) と呼ぶことができるだろう。そして異なった証拠の理論を採用すれば、異なった証拠の基準が用いられることになり、結果として、同じデータが与えられたとしても、そこから「証拠」ないし「質の高い証拠」が得られるかどうかの判定は異なりうるのである。例えば、先に触れた、ハーバードのECMO治験を巡る議論においても、各研究者の見解を異ならせた要因の一つは、各人が採用している証拠の基準、ひいては証拠の理論の違いであったと言えよう。

既に触れたように、ハーバードの治験を巡る議論に参加した研究者の多くは医療統計学者であった。先に、科学における実験・観察・標本抽出の方法論として紹介した統計学はまた、一定の証拠の基準を備え、それに照らして、実験・観察データを一定の質を持った証拠たらしめる証拠の理論でもある。例えば、先に概観した、古典統計学におけるネイマン－ピアソン型仮説検定は、ランダムイゼーションを中核とする一定の手続きに則って得られた「生のデータ」に対して、帰無・対立仮説を反証ないし確証する「証拠」としての資格を与え、同時にその証拠の「質」を、有意水準や検出力といった量的指標として明示する理論的枠組みであるとも言えるのである。

しかし、統計学は決して一枚岩の方法論なのではない。統計的方法論とは、より細かく見れば、複数の方法論からなる緩やかな集合体なのである。「統計的方法論」という単語は、単数名詞ではなく、複数ないし集合名詞であるとも言えよう。そして、それらの方法論のあいだには、互いの妥当性や優劣を巡って論争が繰り返されてきた。ハーバードの研究者たちが依拠した古典方法論も、－「古典」という名が示すように－、数ある方法論の中で最も標準的なものと目されてはいるが、あくまで複数の方法論のうちの一つに過ぎない。医学に話を限っても、それ以外にも、ベイズ方法論や、臨床疫学の方法論といった複数の統計的方法論が現実にも用いられているのである。

このような情勢の下、例えばベイズ方法論を採用する研究者は、ベイズ的な証拠の基準にもとづいて、ハーバード治験開始時にはECMO優位を示す「証拠」ないしは「よい証拠」が蓄積されていたと主張し、ハーバード治験は不要であったと論じたのである。また別の論者は、古典方法論を奉じ、その証拠の基準に照らして、治験開始時には少なくとも「よい証拠」がなかったと見なし、ハーバード治験の必要性を擁護したのである。

また、同じ一つの統計的方法論を採用している研究者の間でも、必ずしも同じ証拠の基準が受け入れられているわけではない。例えば、ECMO治験を行った三つの研究グループは、いずれも古典統計学の方法論を用いていたが、彼らが暗黙裡に採用していた証拠の基準は必ずしも同一ではなかったのである。ハーバード・チームは、バートレットらが「よい証拠」として提示したミシガン治験の結果を、自らの基準に照らして不十分だと見なしたがゆえに、第二の治験を実施した。またミシガン治験、ハーバード治験のいずれもが「よい証拠」を提供していないとして、より大規模な治験を実施したイギリスの研究者たちは、ミシガン・グループ、ハーバード・グループのいずれとも異なった証拠の基準を採用していたと見なせるのである⁶⁹⁾。

統計的方法論とそこで採用されている証拠理論の違いに由来する方法論的な議論は、現代医学の他の領域においても見出すことができる。先に言及した、EBMにおける「証拠のヒエラルキー」は、古典統計学の方法論を最も重視する立場で組み上げられているものである。近年、そのヒエラルキーに挑戦する試みもいくつか見られるが、それら批判の多くは、そのヒエラルキーにおいて低い位置しか与えられていなかった、

69) もちろん、同一の研究者、研究グループが、複数の証拠の基準、ひいては証拠の理論を、便宜的に使い分けたり、併用することもままある。特に、証拠理論そのものの研究者—例えば統計学者—よりも、そのユーザーたる現場の科学者—例えば医学者—にその傾向が強い。一例を挙げれば、一方でRCTを実施しつつ他方でデータベースをも構築していたミシガン・チームも、ランダムマイゼーションを必須の要件とする古典統計学の方法論と、その他の証拠理論を併用していたと言えるのである。実際、彼らは、RCT、ランダムマイゼーションを中止した後も続けられた治験、さらにはデータベースの構築を、事前情報や倫理的問題という、その都度の状況に応じてデザインを変えていく、一つの連続した証拠立て作業と捉えていた(Cf. Bartlett and Cornell, 1991, p.64)。ここで見られる、証拠の基準・理論に関する「実践的な折衷主義」とでも呼べる態度は、現場の科学者に広く見られる立場である。

臨床疫学の方法論を支持する立場からなされているのである⁷⁰⁾。

統計学という不可視の装置

しかし、私としては、ここで、これらの論争において各々の論者が陰に陽に採用していた証拠の基準をいちいち見ていくことよりも、彼らが例外なく、しかも必ずしも明示的に主張することなく抱いていた一つの考え、いわば現代の医学界における「語られざる共通了解」とも言える事柄に目を向けてみたい。

ハーバードECMO治験の妥当性を巡っては、ランダム化や同時対照群を用いる古典統計学の方法論そのものにも疑問の目が向けられ、別の方法論（我々の言う証拠の理論）を臨床治験に導入する可能性が論じられた。その際、古典方法論のオルタナティブとして候補に挙げたのは、ベイズ方法論や臨床疫学の方法論であった。またEBMにおける「証拠のヒエラルキー」を巡っては、古典統計学を重視し臨床疫学の方法論を軽視する立場と、後者に対してより高い地位を与えるべきだとする立場の間で議論が戦わされていた。

ところが、先にも触れたように、ここで名前が挙げられている方法論は、いずれも広い意味での統計的方法論であることには変わりがない。つまり現代の医学界において繰り広げられている、臨床治験の方法を巡る議論は、せいぜい、各種の統計的方法論の間の論争にしか過ぎないのであり、そこでは（例えば、古典統計学かベイズ統計学かはともかくとして）何らかの統計的方法論の適用が問題視されることは、ほとんどなかったと言えるのである。

—臨床治験は、不特定ではあるが何らかの統計的方法論に則って実施されるべきであり、治療の選択をも含めた医療行為全般もまた統計的証拠にもとづいて行われるべきである。

このような考えは、現代の医学界において、ほとんど疑問に付されることもなく、

70) Concato, J., Observational Versus Experimental Studies: What's the Evidence for a Hierarchy?, *The Journal of the American Society for Experimental Neuro Therapeutics*, 2004, vol.1, pp.341-347.

広く受け入れられている。この考えを、「証拠」や「証拠の基準」に即して言い換えれば、つぎのようになるだろう。医療・医学は、何らかの統計的方法論の分析を経て得られた証拠、即ち統計的証拠にもとづかなければならない。データが証拠と言えるかどうか、その証拠がどれだけの質を持つと見なせるのかは、何らかの統計的方法論に即した何らかの基準によって判定されるべきである一と。このように、不特定の統計的方法論としての統計学（以下、「統計学」という言葉を、個々の統計的方法論に限定されない、一般的・不特定の意味で用いる）は、現代医学において、「データを証拠と化す装置」として認定された上、そのような装置としては、他の証拠理論を差し置いて、独占的ないし特権的な地位を与えられているのである⁷¹⁾。

現代医学の共通了解として、「データを証拠と化す独占的・特権的装置」という地位を与えられている統計学は、誰もがそれを利用していながら、誰の目にも問題があるものとしては映らない、いわば「不可視の装置」と化しているとも言える。ハーバードECMO治験の是非、古典的方法論の是非は問題となっても、統計学の是非は問題とはならないのである。

しかし、この不可視の装置としての統計学こそ、ハーバードECMO治験の背後に控え、その実施を支え、その中における、あのような形での4人の死を準備していたものの一つなのである。ハーバードの研究者たちは、古典統計学を奉じ（言い換えると、古典統計学を正当な証拠産出装置として認め）、その方法論の証拠の基準に従って、治験開始時に与えられていた事前情報を、ECMO優位の「証拠」ないしは「よい証拠」と見なさず、同時対照群とランダム化を備え、かつ対照群からのインフォームド・コンセントを欠いた治験を実施したのであった。そしてこのような形でのECMO治験がチルドレンズで行われたからこそ、4人はECMO治療を受けることができないまま亡くなったのである。また、ハーバード治験を巡る議論の中で指摘された

71) ここで言う、一般的・不特定の意味での統計学とは、例えば古典方法論やベイズ方法論といった特定の方法論を－古典述語論理の用語で言えば－「存在一般化(existential generalization)」したものとも言える。従って、「臨床治験は（ベイズ方法論ではなく）、古典方法論に則って行われなければならない」という主張も、「臨床治験は（古典方法論ではなく）、ベイズ方法論に則って行われなければならない」という主張も、ともに、「臨床治験は、一般的・不特定の意味での統計学に則って行われなければならない」という主張を「(古典述語論理的に) 演繹的に含意する」と理解されるべきである。

ように、もしその研究者たちが、古典統計学の方法論を適用しなかった場合、ベイズ方法論やその他の統計的方法論によって、同時対照群をもランダム化をも欠いた実験が行われていた可能性がある。つまり、もしハーバードの研究者たちが、古典統計学の技法を用いていなければ、たとえ他の条件が全く同じであるとしても、必ずしも、あのような状況下での4人の死が発生したとは限らないのである。結局、古典統計学は、4人の死を引き起こした条件（必要条件）の一つなのである。

そして、古典統計学を証拠産出装置として受け入れるということは、不特定の統計的方法論としての統計学をもまた証拠産出装置として受け入れることを含意していた。端的に言って、統計学に証拠産出装置としての資格を認めるという現代医学における共通了解は、ハーバード治験での4人の死の背後にあった（哲学的読者のために言い換えると、「必要条件の一つによって含意されていた」となろう）事柄の一つなのである。ECMO治験におけるあのような形での4人の死をもたらした、現代医学において自明視されている事柄のうちには、間違いなく、このような語られざる共通了解が含まれていたのである。

「悲劇」としての四人の死

統計学を独占的ないしは特権的な証拠の産出装置と見なしているのは、なにも現代アメリカの医学界に限ったことではない。むしろ、そのような考えは、医学の科学化に際してモデルと目された物理学・化学・生物学などの自然科学において、近代以降、長年にわたって培われてきたのである。これらの諸科学のなかでも、統計学の導入において主導的な役割を果たしたのが物理学である。科学の現場で実際に使われるようになった最初の統計的方法論は最小二乗法（the least square method）であるが、それは、天体力学の摂動論や、測地学における子午線の測定といった、物理学の研究の現場の必要性に応じて、18世紀末から19世紀始めにかけて開発されたのである。（ちなみに子午線の測定の目的の一つは1メートルの長さの確定にあった。）天体観測において最小二乗法の利用が普及したのは1830年代、物理学一般における物理量の測定—例えば、光速度の測定—において統計的方法論の応用が定着したのは1860年代であると

されている。それ以降、今日にいたるまで、物理実験や観察の結果として統計的分析を施した値を報告することが常識として定着している⁷²⁾。

このように見てくれば、医学が物理学をモデルとする科学化のプロセスにおいて、統計学を、科学的証拠を独占的ないし特権的に生産する装置と見なす考えをも導入したのは、むしろ当然であるとも言える。医学の科学化という動向は、物理現象を対象とする物理学の研究に即して生み出され、定着してきた方法を、人間、なかならず患者を対象とする医学の分野に移植するプロセスでもあったという意味で、物理学帝国主義の帝国支配の拡張という側面をも持っている。しかし、それはまた、物理学における実験・観察の方法論であった統計学の観点から見れば、統計学の適用範囲をさらに一步広げるプロセス、即ち、統計学の帝国の版図拡大の一環とも言えるのである。

また、この物理学と統計学からなる二重の帝国による医学の支配は、物理現象といった、倫理的な配慮を要しない対象についての研究という、いわば没倫理的な環境の下で培われてきた、統計学ひいては実験・観察という方法を、倫理的な配慮が何よりも求められる人間、わけても病に苦しむ患者を対象とする医学に持ち込むプロセスでもあった。その成り立ちからして没倫理的な方法を、倫理的に極めて敏感な領域に導入すれば、そこにさまざまな倫理的あつれきが生ずることは自明であろう。医学の方法が物理学などの他の分野に波及したのではなく、その逆であったこと。そこに、人体実験なかならず RCT の倫理的問題の根があるのである。

医学の科学化、物理学的・統計学的帝国主義の膨張過程。ECMO 治験の上に影を落としているのは、これら近代科学を貫く、大きな歴史的動向である。逆に言えば、ハーバードの ECMO 治験は、このような近代科学の巨視的な流れの一つの現れに過ぎないのである。その意味でも、ECMO 治験における、かの 4 人の死は、個々の研究者に

72) 最小二乗法の成立史については、Sheynin, O.B., の一連の論文、特に Sheynin, O.B., *Mathematical Treatment of Astronomical Observations*, *Archive for History of Exact Science*, 1973, vol. 11, pp.97 - 126, *On the History of Statistical Method in Astronomy*, *Archive for History of Exact Sciences*, 1984, vol.29, pp.151-199, *Theory of Errors*, pp.1315-1324, in Grattan-Guinness I., ed., *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, 1994, Routledge, London. さらに Stigler の著作、Stigler, S.M., *The History of Statistics: The Measurement of Uncertainty before 1900*, 1986, Harvard U.P., Cambridge MA, chap.1, *Statistics on the Table*, 1999, chap.17 を参照。

責を負わすべき事柄というよりも、個々の人間には止めようがない、宿命的とも言える歴史の流れの中で、起こるべくして起こった事態であると言えるのである。つまり、それは、医学の科学化や物理学と統計学の二重の帝国主義といった近代科学の歴史がもたらした一つの「悲劇」なのである。

しかしわれわれは、あのような形での4人の死が避けがたい悲劇であったことを認定した上で、それでもなお、冒頭に立てた問いを繰り返したい。即ち、なぜ4人は、事前情報に照らせば有望のように思え、また利用可能でもあった治療法を受けられずに死ななければならなかったのか—と。いまや、この問いの矛先は、統計学という科学における独占的ないし特権的な証拠産出装置そのものに向けられている。そもそも、「統計的方法論に則って計画・実施された実験・観察において統計的分析を加えられたデータこそが一定の証拠能力を持つ」という考えが、医学界、ひいては科学の世界において広く受け入れられているからこそ、ハーバードのECMO治療は行われたのであった。言い換えると、「「データが証拠と見なせるかどうか」、ないしは「どれだけ質の高い証拠と見なせるかどうか」は、統計学が提供する証拠の基準に照らして判断されるべきである」という考えが、科学の「常識」となっていたからこそ、4人はECMO治療を受けることなく死ななければならなかったのである。

—それでは何故、とわれわれは問うことにしよう。

なぜ、統計的分析をへたデータは「科学的証拠」であると言えるのか。統計学は、いかにして、データを証拠と化すのか。統計学における証拠の基準の「根拠」—そのような「根拠」があるとして—とは何か。そもそも、統計的データが科学的仮説の証拠となる、とはいかなることなのか。

これらの問題を問うことで、私は、統計的証拠を「証拠」の名に値しないものとして退けるつもりはない。統計学という装置の証拠産出能力を否定する意図はないのである。かたや、統計学が誤用も濫用もされうる可能性を踏まえ、それを正しく使い、その結果を注意深く評価しようという、EBM運動においても見られる一種の啓蒙キャンペーンの一翼を担う考えも、私にはない。さらに私は、統計的方法論に代わる、より役に立ちより無害な方法論を構想し提案しようとも思わない。現行の統計学とは別の何らかの証拠の基準や証拠の理論の構築を目指すことはしないのである。私が目指

すのは、現代の科学の現場で実際に用いられている統計学が、データを証拠化する仕組み・メカニズムを明らかにし、その仕組みの理由・根拠を同定し、その理由・根拠がいかなる仕方で正当化されているのか—ないしは、(必ずしも正当化とは言えない仕方で) 受け入れられているのか—を見定めることである。

もちろん、これらの事柄が解明されたからといって、ただちに現実の科学の実践、臨床治験のあり方に何らかの変化がもたらされることはないであろう。しかし、そのことで、少なくとも、なぜ4人があのような仕方で死ななければならなかったのかという、ここまでなんども繰り返し問うてきた問題に対して、より一歩踏み込んだ解答を与えることはできるだろう。データを証拠と化す「仕掛け」としての統計学を支えていたものはまた、ハーバードECMO治験をも支え、結果として、あのような形での4人の死を準備していたに違いないからである。統計学を支え、4人の死を準備していたものの「正体」を明らかにすることで、私は、生まれてすぐに死ななければならなかった、かの4人の新生児に対して、「なぜあなたたちは、あのような仕方で死ななければならなかったのか」を説明する責任を買って出ようと思う。それが、統計学者でも医学者でもない、哲学者としての私にできるせめてものことであるし、また、統計学者でも医学者でもない、哲学者としての私がやるべき事柄だと思うからである⁷³⁾。

73) 本稿を執筆するにあたって、西山慶彦氏(京都大学経済研究所)からハーバードECMO治験の統計実験としての性質について様々なご教示をえた。また神埼宣次氏(京都大学大学院文学研究科)には、数値計算をお願いした。ここに記して謝したい。