

シンポジウム 1 ● 臨床試験における信頼性とエビデンス

2 臨床試験の質とは何か

東京大学大学院医学系研究科公共健康医学専攻生物統計学分野

(現 東京大学名誉教授/中央大学理工学部人間総合理工学科生物統計学)

大橋 靖雄

臨床研究の目標は、ノースカロライナ大チャペルヒル校の生物統計学看板教授 Koch によれば、いかに精密 (precise) に、バイアスを少なくして、セレクションバイアスを避け、より広い対象に適用できるかとされている (表 1)。

臨床試験の「質」とは何か (表 2)。バイアスがないことは当然の条件であるが、プロトコルが精緻なことは条件になるのか。試験治療のコンプライアンスが高いことが常に必要なのか。エンドポイントの客観性が高いのがよいかといえ、QOL をプライマリーエンドポイントにとった試験では客観性は不要である。だから、表 2 でいえば、プロトコルが精緻であるというあたりから、議論が多くなる。これまでの臨床研究を例示しながら、臨床試験の質とは何かについて述べてみたい。

1 方法論が不適切だった一例

方法論が不適切であり質が明らかに低かった例を表 3 に示した。がんの臨床試験で、封筒法を採用していたことがある。ランダム化の結果を封筒に入れ、担当医に渡す。すると、臨床現場では、封筒の紛失など、さまざまな事象が起こりうる。開封してしまい、患者をみってから決めることさえある。それでは、ランダム化が損なわれてしまう。問題なのは、開封する順序が変わった理由が、同一施設内で複数の医

表 1 臨床 (試験) 研究の目標 (by Koch G)

- ・誤差的バラツキを小さくすること (精度を高くすること) → Clarity : Minimizing random error
- ・偏り (バイアス) を小さくすること → Comparability : Minimizing bias, Internal validity
- ・広い対象に適用できる結論を得ること、一般化可能性 → Generalizability : External validity

表 2 臨床試験の質 Quality of Clinical Trials

- ・臨床試験の質とは何か？
 捏造・改竄は論外であるが……
 バイアスがない (小さい) こと、当然の条件ではあるが？
 プロトコルが精緻である？
 試験治療のコンプライアンスが高い？
 ハードエンドポイントの試験である？
 (研究者主導試験の場合) 企業からの影響はない？
 SDV (source document verification) が徹底している？
 品質管理が徹底している。何のための品質管理？
- ・質を高めるためにはどうすべきか？

者に分けたというようなバイアスにつながらない理由か、意図的に選んだからか、識別できないことである。それでは、品質保証ができない。

それなら、封筒法は常に不適切かといえ、そうではない。薬剤師など、第三者が封筒を管理し、医師が適格性をチェックし、第三者に連絡のうえ開封

How Do We Have to Secure High Quality of Clinical Trials and Research?

Yasuo Ohashi : Department of Biostatistics, School of Public Health, Graduate School of Medicine, The University of Tokyo (the present affiliation, Emeritus Professor, The University of Tokyo/Department of Integrated Science and Engineering for Sustainable Society, Faculty of Science and Engineering, Chuo University)

表 3 ランダム化手法が不適切であった例

消化器癌を主体とする化学療法 122 例/12 施設
 非治癒切除適格例 A：化学療法対照 41 例
 B：化学療法+SSM 40 例
 年齢、性、部位、ステージ、検査値などに群間差は認められない

封筒法割付けの A, B の分布			施設 #
	6	BBBBBBBB	
	5	B	
AAAA	10	BBBBBBBBBBBB	
AAA	9	BBBB	
AAAAA	3	BBBBB	
AA	1	BB	
A	12	B	
	7*		
AAAAAAA	8	BBBBB	
AAAAAAA	2*	B	
AAAAA	11	B	
AA	4*		*世話人施設
41	計	40	

するなどすればよい。

ちなみに、表 3 の SSM は丸山ワクチンである。丸山ワクチンは、種々の社会的条件から現在も認められていない。私たちは丸山ワクチンで二重盲検プラセボ対照臨床試験を行い、最近、かなり良いデータを出し論文化した。子宮頸癌で、エンドポイントを生存とした 7 年がかりの試験である。次の臨床試験が進行中で、将来どのように評価されるのか、非常に楽しみである。

2 最近の不祥事について

1) 統計家の重要性

日本で昨今話題となった不祥事が起きた理由を表 4 に示した。私たち統計仲間では、特に中立で経験のある統計家の参画がなかったことが問題になっている。

私が会長を務める日本計量生物学会を紹介したい。歴史は非常に古く、1980 年に設立されている。統計家自体が日本では少ないため、会員数は 500 人である。企業やアカデミアの試験統計家はほとんど参加しているが、あの不祥事に関係した会員はいなかった。私たちは約 2 年前から統計家の行動基準の作成を開始しており、日本軽量生物学会としてこの行動基準を採択する予定である。そこには、約 10 の守

表 4 なぜ不祥事は起きたか

- EBM の名を借りて大規模臨床試験が（特に競争の激しい降圧薬で）次々に実施され、その「成果」が販売拡張に使われた。しかし、
- 2007 年から学会発表が宣伝に使えなくなり、販売拡張のために大規模試験の論文化が必要であった。
- がん領域を除けば、研究者主導の臨床試験を実施する基盤はなかった。
- がん領域など一部を除けば、臨床試験の品質管理・品質保証の概念は PI を務める医師にはなかった。
- 「製薬会社と研究者あるいはそれを支える財団等が契約を交わし中立の立場で臨床試験を行う」という考え方は一般的でなかった。大学紛争以来、委任経理金という寄付で臨床試験は実施されることが多かった。
- 寄付金が少額でとても CRO を本格的に使うような試験はできなかった。
- 試験を請け負った小規模 CRO が破綻した？
- 中立で経験のある統計家の参加はなかった。
- 日本ではプラセボ試験は市販後にはできない、という思い込みがあった。もちろん製薬会社の営業の立場からプラセボ試験は好ましくない。

るべき基準をあげている。最も重要なのは次の 4 つである。

- ①情報を適切に扱う。
- ②法やガイドラインを遵守する。
- ③不正行為を予防する。
- ④利益相反による弊害を防ぐ。

今回の不祥事で、意図があれば、統計家は不正を行えることが明らかになった。統計家は、自身のモラルをあげなければいけない。また、特に多施設大規模臨床試験の場合、統計家の存在のみならず、中立なデータセンターが重要である（図 1）。

2) 不祥事を生んだ背景

「誰も教えてくれなかった癌臨床試験の正しい解釈」（中外医学社 2011）には、統計家について「ほとんど幻ではないかと思われるような存在」と表現されている。本当にそうだろうか。

日本科学技術連盟が、製薬会社の試験統計家の教育を始めたのが 1989 年である。すでに 25 回実施され、約 1400 人が卒業した。東京大学に疫学・生物統計学の講座ができたのが 1992 年、3 つの薬剤について大規模な試験を行った厚生省の薬剤疫学的手法検討研究事業は 1994 年に開始されている。さらに統計学の講座は、北里大学 1999 年、京都大学 2000 年に設置され、他の大学も続いている。統計家を探そうと思えば可能だったはずである。

ただ、医療系のアカデミア、大学には統計家はあ

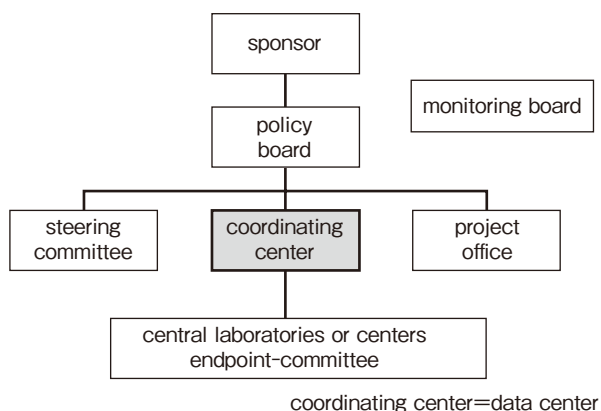


図1 NHLBIモデルによる臨床試験

NHLBI: National Heart Lung Blood Institute

大橋靖雄, 医学研究のための情報システム, 医学のあゆみ 1988; 146: 11-3.

オリジナルは Greenberg Report (1967)

http://sph.unc.edu/files/2013/07/greenberg_report.pdf

まりいないのである。旧帝大にはすべて講座が設置されたものの、中核病院で統計家を採用しようとすると、なかなかみつからない。1990年ころまで、医師研究者への文献検索と統計解析サービスは、製薬会社営業部門の重要な仕事であった。これは営業の一環で、利益相反 (COI) であったが、利益の供与として認識されたのは1991年のことである。1991年に改正独占禁止法が施行され、その業界版として日本製薬工業協会の公正競争規約が作成され、医療関係者に対する無償のサービス提供を禁じた不当景品類防止法が発効して、文献検索・統計解析サービスは影をひそめたはずであった。だが、実は2000年ころまでは実態として残っていた。そのことを不審に思わないのは、企業側ではなく、むしろ医師側であったと思われる。それが今回の不祥事の本質、温床だろうと、私は考えている。

このような癒着、企業の影響あるいはCOIの問題は、別に日本だけでなく、欧米でははるかに深刻にとらえられていた。そのようなことが日本になかなか伝わってこなかったというのが実情であろう。

3 臨床試験における中立性の確保 (表5)

2001年のLancetに、企業主導の臨床試験は諸刃の刃であるという論文が発表された。このころには、米国NIH (National Institutes of Health) の資金が少なくなり、資金提供者がアカデミアから民間、製薬

表5 臨床試験における中立性の確保

■企業資金と学術的独立性

Montaner JS, O'Shaughnessy MV, Schechter MT. Industry-sponsored clinical research: a double-edged sword. Lancet 2001; 358 (9296): 1893-5.

研究資金がアカデミアから民間・CROへシフト

計画・解析・解釈・出版への資金提供者の影響

■ICMJE (International Committee of Medical Journal Editors)

臨床試験登録の推進 (2005年7月からの試験は登録義務化)

2010年にCOI宣言を含んだ統一投稿規定

■JAMA (The Journal of the American Medical Association) の投稿条件 (2008)

ICMJEの認めるサイトへの事前登録 (日本はUMIN)

ガイドラインCONSORTに沿った論文記述

アカデミア統計家による解析 (現在は緩和)

会社に移ってきた。それで、製薬会社が試験結果に口を出し、出版を差し止めるという事態が起きていたのである。

ICMJE (International Committee of Medical Journal Editors) がCOIのことを真剣に検討し始めたのが2003年ころである。その成果として、COI宣言を含む統一投稿規定が2010年に発表された。ICMJEは、2005年に出版バイアスを避けるために臨床試験登録制度を開始している。欧米でも、COI問題の対処にははるかに長い時間が必要だったのである。したがって、日本が遅れたのも、ある意味でしかたないとも思われる。

私の経験でいえば、医療関連雑誌のなかで査読が最も厳密なのがJAMA (The Journal of the American Medical Association) である。投稿中の学会発表が禁じられているほどである。もちろん試験の事前登録が前提であるが、日本での登録はUMINでよい。また、ガイドラインCONSORT (後述) に沿った論文記述が必要とされ、アカデミア統計家の解析結果しか受け付けていなかった。つまり、製薬会社が行った治験でも、製薬会社の統計家が解析した研究は認めなかったのである。

4 過去における臨床試験の好例

1) EBCTCGのメタアナリシス

乳癌の術後補助療法に関しては、世界的なメタアナリシスグループ (Early Breast Cancer Trialists'

Collaborative Group : EBCTCG) が構築されている。これはメタアナリシスの嚆矢ともいえ、世界中の乳癌研究者がこのグループに、論文ではなく、個票のデータを必ず提出している。それらの解析は 1984 年ころから始められ、1992 年、Lancet に最初のメタアナリシスが 2 論文として発表された。ひとつは、タモキシフェンは再発を 20~40% 減らすという論文で、これには、日本の ACETBC (Adjuvant Chemo-Endocrine Therapy for Breast Cancer) グループの研究も含まれている。メタアナリシスは、タモキシフェン投与 1 年、2 年、5 年をグループとして行われ、徐々にリスクリダクションが大きくなるという結果であった。2 サイクルのメタアナリシスで 5 年投与がよいことになり、現在、5 年投与が世界標準となっている。メタアナリシスがまさに世の中を変えたわけである。

さらに、タモキシフェンの使用は 5 年で十分なのか。estrogen receptor (ER) 陽性の乳癌は 5 年を経過しても再発する。ホルモン治療の効果は、中止しても数年間は持続するが、その後に再発するのである。それでは 5 年以上、薬剤を使用したらよいのではないかと、ATLAS (Adjuvant Tamoxifen : Longer Against Shorter) 研究が開始された。36 カ国で 15000 人が登録され、ER 陽性に限定した 6800 人が解析された。その結果、10 年間使用したほうがよいことが示された。死亡率でも差がつき、未治療に比べると 10 年後にはリスクが 52% になる。ER 陽性の乳癌のホルモン剤術後補助療法は 10 年以上となるであろう。世界中の教科書が変わるはずである。さらに、イギリスで行われた ATOM (Achievement of improved survival by molecular Targeted chemotherapy and liver resection for not Optimally resectable colorectal liver Metastases) トライアルが、この結果を確認した。

以上のことから、これらの研究は、再現性もある良い研究なわけである。最初の疑問を思い出してほしい。プロトコルが精緻なほど研究の質は高いのかという疑問である。もしこれらの研究が通常の治験で行われたとしたら、どういう適格規準になるのか。乳癌であることが組織学的に確認され、治癒切除され、ホルモン陽性で、年齢は問わず、肝機能や腎機能はどの程度であるなど、1 ページは必要である。

だが実際には、プロトコルはきわめてシンプルで、適格条件は次の 1 文だけであった。「5 年間タモキシフェンを服用し、続けるべきか止めるべきか決定できない医者と患者」である。一般化可能性が非常に高く、だれが対象か、きわめて明白である。

臨床試験の概念でいえば、この試験は pragmatic (実際の) 試験といえる。通常の治験は、どちらかといえば explanatory (説明的) な、つまり臨床薬理学的な効果を検証するための試験である。pragmatic 試験はすでにある薬剤を臨床現場でどう使うべきか、リアルワールドのシミュレーションに近い。その意味では、これは質が高い試験といえる。プロトコルが精緻であることは必ずしも必要条件ではないのである。

【関連文献】

- 1) Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. Early Breast Cancer Trialists' Collaborative Group. Lancet 1992 ; 339 (8784) : 1-15.
- 2) Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. Early Breast Cancer Trialists' Collaborative Group. Lancet 1992 ; 339 (8785) : 71-85.

2) MEGA Study

2006 年、MEGA Study の論文が Lancet に掲載された。MEGA Study は、日本初の大規模な循環器系臨床試験で、エンドポイントは冠動脈疾患発生である。オープンで、HMG-CoA reductase 阻害薬 (プラバスタチン) + 食事療法併用群、食事療法単独群という 2 群を比較した、いわゆる PROBE (prospective randomised, open, blinded endpoints) 試験である。エンドポイントを評価する際、バイアスが入りやすい試験であった。

その結果、18% の LDL-C の低下が 33% のリスクリダクションを示した。論文査読時に Lancet から最終的に要求されたことは、コンプライアンスを表にせよ、というものであった。投与群のコンプライアンスは高く、90% を超えていたが、食事療法単独群のコンプライアンスは悪く、5 年ではプラバスタチン投与は 34% であった。その理由は、プロトコルに食事療法単独群でも、日本動脈硬化学会の基準であ

る 220 mg/dL 以上なら高脂血症の治療をしてよいとされ、プラバスタチンの使用が医師の裁量に任されていた。Lancet は、コンプライアンスが悪いことを理由にして不採択にはしなかった。なぜかといえば、この試験はきわめて pragmatic な試験で、コレステロールの低下作用はすでに臨床薬理学的に検証されているプラバスタチンについて、リアルワールドでの使い方、すなわち初めからプラバスタチンを投与するか、食事療法でできるだけがんばるか、を比較したからである。

UMIN の登録では、臨床試験のタイプ (表 6)、検証か探索かを入力しなければならない。その両方ということもありうるが、説明的か実践的になると、両方ということはあるえない。このタイプがきちんと書いてあるような試験は、研究者が方法論を理解しているといってもよいと考えられている。

【関連文献】

Nakamura H, Arakawa K, Itakura H, Kitabatake A, Goto Y, Toyota T, Nakaya N, Nishimoto S, Muranaka M, Yamamoto A, Mizuno K, Ohashi Y; MEGA Study Group. Primary prevention of cardiovascular disease with pravastatin in Japan (MEGA Study): a prospective randomised controlled trial. Lancet 2006; 368: 1155-63.

5 質の高い臨床試験とはどういうものか

1) レギュラトリーサイエンスの重要性

私は 1985 年、TQC (total quality control) 用語辞典の編集をし、統計部分だけはすべて確認していたものの、正直なところ、他の部分はよく読んでいなかった。JIS (日本工業規格) の品質の定義には、こう書かれている。「品物またはサービスが使用目的を満たしているかどうかを決定するための評価の対象となる固有の性質・性能の全体」。さらに、備考には、「品物またはサービスが使用目的を満たしているかどうか判定する際に、その品物またはサービスが社会に与える影響を考慮する必要がある」とある。近年、ネガティブな意味で、品質は社会に与える影響が大きくなっているが、これは品質の高さには影響の大きさも含まれるという意味である。そして、「品質は品質特性によって構成される」と続く。つまり、品質は複数の機軸によって評価される構成概念であるということである。構成概念は心理学の用語であり、似た例には知能や QOL がある。

表 6 UMIN 登録：臨床試験のタイプ

検証的 (Confirmatory) :

すでに探索的試験などにより仮説が形成されており、その仮説を検証するために実施される試験

探索的 (Exploratory) :

検証的試験の実施前に、仮説を形成するために実施する試験

説明的 (Explanatory) :

介入法の作用機序などを解明する目的で、実施条件をある程度厳しく設定して実施する試験

実践的 (Pragmatic) :

実施条件をゆるく設定し、日常診療に近い状況で介入法を評価するために実施する試験

ただ、だれがその品質を評価するのかが問題である。答えは顧客であるが、私たちの研究の顧客とはいったいだれかという、きわめて重要な問題が提示される。製薬会社にとっての顧客は PMDA (独立行政法人 医薬品医療機器総合機構) であるが、PMDA を満足させるために薬剤を開発しているわけではない。やはり、国民のためだと言いたい。研究者主導研究では、国民とのあいだに入るのが医学界のコミュニティである。つまり、レビューアーであり、彼らを満足させないと、論文が通らない。しかし、レビューアーを満足させるために、私たちは研究をしているわけではなく、その先に国民がある。

私たちの行う臨床研究の評価を困難なものにしてるのは顧客の二重構造であり、顧客間にしばしば乖離があるという問題がある。これをつめていかなければならず、そこにレギュラトリーサイエンスの使命があると、最近、考えている。

2) 「当たり前品質」と「魅力的品質」

狩野紀昭氏は、品質管理に関して世界的に大きな貢献をした。狩野モデルとよばれる「当たり前品質」と「魅力的品質」の考え方である。臨床試験を志す私たちも心にとめるべきことで、「魅力的品質」を目指すべきである。これを私たちの世界にも敷衍したい。そういう作業が、不祥事対策の次のステップではないかと思っている。

臨床試験の品質特性 (表 7) は、次の機軸、①標準治療に対するインパクト、②正確さと精密さ、一般化可能性、そしてプロセスとして、③プラン、④デザイン、⑤実施、⑥出版、⑦品質保証システムから構成される。

表 7 臨床試験の品質特性 (Domain of clinical trial quality characteristics)

①標準治療に対するインパクト
②正確さと精密さ、一般化可能性
③プラン 研究組織
④デザイン プロトコルの質 (2013 年 1 月 SPIRIT 発表)/relevant な仮説 pragmatic か explanatory か/エンドポイントの設定/症例数と適切な中間解析, 適切な解析方法
⑤実施 治療コンプライアンス/追跡と脱落, 欠損データ/データ管理とそのシステム
⑥出版 (CONSORT により品質特性は測定可能)

特にデザインに関しては, プロトコル作成のためのガイドラインとして, Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) が発表されている。きわめて質, 要求水準が高いチェックリストが作成されているので, これに 대응することが, 今後課題となる。出版に関しては, CONSORT (Consolidated Standards of Reporting Trials) が出ており, それを測ることは比較的容易である。つまり, 目標が明確になったということである。

各種のガイドラインに関しては EQUATOR というホームページを見ていただきたい。CONSORT も SPIRIT も掲載されている。また, ICMJE の論文の書き方のガイドラインも大きく変わり, お作法というより「出版とはどうあるべきか」というところにウエイトが移ってきた。著者とはなんであるか, これが最も重要である。研究結果の正確性と integrity に関する疑問, それに対する説明責任を有する人, 説明できる人と書かれている。これは難しい。これを理解させることが医師に対する重要な教育になると思っている。

質の高い臨床試験は, 検証的試験の場合, 次のようにまとめることができよう。

①目的・仮説が明確で, ②その検証のためのデザインが適切で, ③適切に実施され, ④その結果, 得られてデータの品質保証が適正なレベルでなされ, ⑤目的が達成できたかどうか明確で⑥以上のことが, 明確に論文化され, ⑦実地医療に (大きな) インパクトを与える試験。