

総説

未来の臨床研究 大規模ランダム化比較試験の限界と克服

植田 真一郎*

真のアウトカムを評価する大規模ランダム化臨床試験の意義

A. ランダム化の必要性

臨床研究の目的は暴露因子や介入とアウトカムの間の因果関係の有無を評価することであるが、両者に関連する交絡因子とその混入の結果生じるバイアスを除去する必要がある。交絡因子の除去のためのいくつかの研究デザインや統計解析の手法が存在するがランダム化割り付けは最も有用な手法のひとつである。観察研究においても補正や propensity score の使用などで交絡因子を除去することは可能であるが、未知の因子については対処できない。例としてホルモン補充療法はかつて 1990 年代に大規模な看護師を対象としたコホート研究によって心筋梗塞のリスクを減少させると報告された¹⁾。しかし後にランダム化比較試験 (Randomised controlled trial: RCT) を実施するとプラセボ群よりもむしろ心筋梗塞のリスクは高いと逆の結果になった²⁾。この理由については色々議論されているが、

ホルモン補充療法を受けるという患者側の因子 (健康に関心が高い, 学歴, 経済状況), 医師の患者選択 (より健康な対象者に投与) などが交絡因子として働いたと考えられる。後述するように RCT には種々の問題はあるが, ある治療介入の妥当性を検証するには必要な研究デザインであることは異論を待たない。

B. 大規模であることの必要性

「大規模」とは十分な検出力をもつ症例数を意味する。確立した疾患の治療介入がない時代においては有効性を証明することは比較的容易であり, 必要症例数も少ない。高血圧のはじめての RCT である VA 研究ではわずか一群 70 人程度の症例数で降圧治療の有効性を証明できた³⁾。しかし治療の必要性が認識され, 治療薬の開発, 使用が進むにつれ, ある治療法がこれまでの治療法と比較して優れていることを示すことは困難になってくる。またそのような状況では単独の薬剤がある疾患のリスクを劇的に減少させることも考えにくい。例えば表 1 に示すように循環器領域で十分なエビデンス

表 1 臨床試験で報告された, 循環器薬の死亡リスク低下

	疾患	患者数	相対リスク低下	絶対リスク低下
アスピリン	心筋梗塞	18,773	23%	2.4%
β遮断薬		28,970	18%	1.8%
ACE 阻害薬		101,000	13%	1.3%
アスピリン	2 次予防	54,360	15%	1.2%
β遮断薬		20,312	21%	2.1%
スタチン		17,617	23%	2.7%
ACE 阻害薬		7,105	23%	6.1%
β遮断薬	心不全	12,385	26%	4.0%
スピロノラクトン		1,663	30%	11%

*琉球大学大学院医学研究科臨床薬理学

スを持つとしてガイドラインに掲載されるような薬剤でもリスク減少はプラセボとの比較においてせいぜい20%前後である。つまり良くても20%程度の差を両群とも少なくともこれまでの標準的な治療を行った上で、検出しなければならないので、多くの症例数を必要とすることになる。

C. 真のアウトカムの評価する必要性

通常治療では代用のエンドポイントを用いて薬効を評価する。降圧薬なら血圧、高脂血症薬はLDL コレステロール、抗不整脈薬であれば不整脈の消失が薬効とされる。しかし診療におけるこれらの薬剤を用いた治療の目的は脳卒中、心筋梗塞、心筋梗塞後の不整脈死亡のリスク減少である。治療ではあくまで代用のアウトカムの改善が真のアウトカムを改善させるという前提で評価されている。この前提が必ずしも正しくない場合があり、したがって市販された後、真のアウトカムへの効果を評価する臨床試験が必要とされる。実際CAST試験では心筋梗塞後の不整脈に対して有効な抗不整脈薬とプラセボを、死亡をアウトカムとして比較したが、むしろ抗不整脈薬は死亡リスクを高めるという結果になった⁴⁾。また生物学的な理論はかならずしも現実の診療にあてはまらない。レニン-アンジオテンシン系は動脈硬化性疾患の発症、進展に関与しているとされ、事実基礎的な生命科学研究による実験的エビデンスは多い。この理論に基づいてACE阻害薬とアンジオテンシンII受容体拮抗薬の併用によるこの系のより厳格な抑制が動脈硬化性疾患をもつ患者の予後を改善させる、という仮説を検証するためのRCTが実施された。しかしACE阻害薬単独の治療と比べ予後は改善されず、むしろ安全性の問題が生じることが判明した⁵⁾。これらの事例は治療法の妥当性に関しては理論や代用アウトカムでは不十分でかならず真のアウトカムで評価しなければならないことを示唆している。

ランダム化比較試験による治療法の妥当性検証の限界

A. 患者や疾患の多様性への対処が不可能

RCTはあくまで確率論的な、集団でのリスクを比較する。つまり均一な集団としてのリスクを評価する研究手法である。しかし患者背景も疾患そ

のものも多様であり、そこには対処していない。どちらかという臨床試験の決定論的な実際の診療において個々の患者へ結果を適用することは必ずしも容易ではない。年齢、合併症、性別などのサブグループ解析は可能であるが、これはあくまで逆に治療効果の(背景の多様性を超えた)一貫性を証明するものであり、多様性の克服はできない。

B. 外的妥当性の問題

ランダム化比較試験は患者を選択しており、現実世界の診療での患者と異なるという議論がある。これが「選択バイアス」となるという意見もあるが、これには筆者は賛同しない。選択バイアスとはむしろ効きそうな患者に薬剤を投与するなど、対象者と薬剤の間に交絡因子が存在することにより発生するバイアスと考える。このバイアスはランダム化割り付けでは発生しない。ランダム化比較試験、特に新薬の薬効の評価が適応拡大のための臨床試験では生物学的ノイズを減らして薬効を明確に証明するため、また安全性を確保するため、患者の選択が行われる。除外の対象となるのは合併症を有する患者、多剤併用患者、高齢者、女性などである。しかし現実世界の診療の現場ではこのような患者が多い。試験の適合基準を満たし、除外基準に抵触しなくても、ランダム化の前に試験薬を(多くは少量)投与し、副作用が出現するようなら除外する、あるいはあまりにアドヒアランスの悪い患者は除外するなどの方法も用いられることがある。このような理由で、一部のランダム化比較試験は効果(効能)や安全性を過大評価しており、実際の診療ではかならずしも同じ結果が得られない、すなわち外的妥当性の問題が指摘されている。

C. 時間軸の問題(図1)

動脈硬化性疾患の場合、服薬する期間は少なく見積もっても10年以上になる。しかし多くのこの領域の臨床試験期間は3年または5年であり、長期間にわたる介入の妥当性を評価しにくい。短期間で評価するためにはエンドポイントとなる心血管イベント(心筋梗塞や脳卒中)を多く発生するハイリスク患者を対象とするか、重篤でない、客観性に乏しいエンドポイント(狭心症での入院)を組み合わせた複合エンドポイントを使用せざるを得ず、低リスクの患者での疑問の解決にはならず、結果の信頼性も落ちることになる。

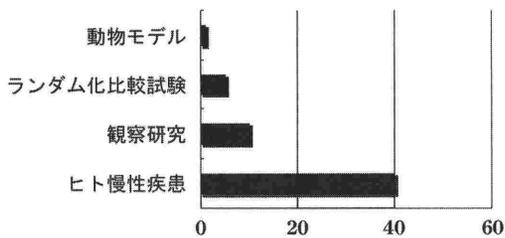


図1 臨床試験における時間軸の問題

ほとんどの慢性疾患の臨床試験は疾患の長期間にわたる過のごく一部をきりとり観察しているのにすぎない

限界を克服するための方法

A. 多様性の克服 ゲノムバイオマーカー

患者や疾患の多様性は結果として薬物応答の多様性を招く。例えばいくつかの薬物代謝酵素や薬剤の標的酵素には遺伝子の多型が存在し、薬物代謝および薬力学的な多様性の原因となる。ワーファリンは CYP2C9 で代謝されるが、変異が存在すると代謝が異なり、適切な抗凝固作用を得るための用量や維持料達成までの時間、致命的な出血までの時間が変わるとされている⁶⁾。また標的酵素であるビタミン K エポキシドリダクターゼの遺伝子変異も用量に影響を与えると報告されている。これらの情報に基づいた、ワーファリン用量設定のためのアルゴリズムも提案されている⁷⁾。また大腸がんの KRAS 遺伝子の変異が存在すると、転移性結腸、直腸がん患者においてパニツムマブは効果がない⁸⁾。HLA-B*1502 は漢民族におけるカルバマゼピン投与時のスティーブンジョンソン症候群出現を予測できる⁹⁾。このようにゲノムバイオマーカー(生殖細胞系列のみならず腫瘍遺伝子や遺伝子発現を含む)は薬剤の有効性、安全性の予測、副作用のリスク最少化、用量の選択に貢献できる可能性がある。薬剤の開発における効率化や安全性の確保に貢献するばかりではなく、実際の診療においても患者、疾患多様性に対処する一つの方法である。しかしマーカーとしての分析の妥当性、臨床的有用性や妥当性の確立には様々な手法の研究が必要である。そのマーカーの薬物代謝や薬力学への影響のみならず、実際の診療での用量反応曲線への影響、薬効、副作用との関連、アウトカムへの関連を示す必要がある。最終的にはそのマーカーに基づいた臨床的決断が予後を改善することを示す必要がある。

B. 時間軸の克服 サロゲートバイオマーカー

時間軸の問題を解決するための一つの方法はサロゲートバイオマーカーの開発である。動脈硬化性疾患の診療はいわば血圧、コレステロールといったサロゲートマーカーを指標にして行われる。これは再現性があり、比較的容易に測定され、これらの指標と脳卒中や心筋梗塞発症リスクが関連し、指標を改善させることがリスクを減少(正確にはコレステロールの場合スタチン系薬剤でのみ)させることが証明されているからである。そしてこれらの指標をもとに動脈硬化の進展を抑制し、将来の脳卒中や心筋梗塞などの心血管イベント発生リスクを減少させることが本来の降圧薬や高脂血症薬、糖尿病薬投与の目的である。しかしこれらの薬剤の評価はかならずしも動脈硬化そのものを指標として行われているわけではない。現在の医学では動脈硬化そのものを可視化することは日常診療レベルでは不可能であり、一部の間接的に可視化される動脈硬化巣にしてもそれを頻回に実施して、その進行や治療効果を推定することや、将来の心血管イベントを予測することは容易ではない。結局時間軸の問題を解決するためには臨床試験においても診療においても動脈硬化そのものを反映するサロゲートマーカーが必要となってくる。ただし表2に示すようにサロゲートマーカーとして満たすべき条件は厳しく、結局そのために(Surrogacy の確立のために)疫学研究や臨床試験が必要になるなど時間がかかる。腎機能障害そのものを反映している可能性の高い蛋白尿ですら FDA は、surrogacy は十分に確立していないという見解である。また心不全の治療の指標とされる BNP はほぼサロゲートマーカーとしての必要条件を満たすが、BNP ガイドの治療が BNP を用いない治療と比較して予後を改善できるか否かは未だ明らかではない¹⁰⁾。

表2 バイオマーカーの surrogacy 確立のために満たすべき条件

・ 簡便に測定可能、安価、再現性良好
・ 十分な科学的正当性をもつ
・ 真のアウトカムをみるよりも短期間、少人数に試験で薬効を評価可能
・ 介入によるマーカーの変化と予後の改善が量的にも相関
・ マーカーガイドの治療が予後を改善

C. 外的妥当性の問題 観察研究をどのように行うか

冒頭に述べたように観察研究の欠点は交絡因子により原因と結果の取り違えが生じる可能性があることである。しかしより選択されていない、現実世界の診療で遭遇する患者が対象となるため結果の外的妥当性が高いという利点がある。結局選択された患者において、厳格な RCT で安全に、ノイズの混入なく薬剤の効能を評価した後、観察研究やより現実的な RCT で治療法の効果を評価し、より外的妥当性の高い結果を得ることが必要になる。もし両者の結果が一致すればその治療法に関する評価は信頼性が高いと言える。

我々は冠動脈疾患患者における積極的脂質低下、降圧が標準的な危険因子管理と比較して優れているかどうか検証するため、冠動脈疾患患者レジストリを基盤とした観察研究と同じ集団でよりリスクの高い集団(心筋梗塞の既往を有する患者)を対象とした RCT を計画している。前者では血中 LDL コレステロール値や血圧値を時間依存性変数として取り扱い、積極的管理へと移行した群とそうでない群との比較を行う。後者は積極的脂質低下、降圧群と標準治療群にランダム化割り付けし、3年間の観察を実施する。この両者の結果が一致すれば結果の信頼性の点からも外的妥当性の点からも積極的脂質低下、降圧は妥当と言える。

文 献

1) Stampfer MJ, Colditz GA, Willett WC, et al: Postmenopausal estrogen therapy and cardiovascular disease. Ten-year follow-up from the nurses' health study. *N Engl J Med* 1991; 325: 756-62.

2) Writing Group for the Women's Health Initiative Investigators: Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002; 288: 321-33.

3) VA Cooperative Study Group on Antihypertensive Agents: Effects of treatment on morbidity in hypertension. Results in patients with diastolic blood pressures averaging 115 through 129 mmHg. *JAMA* 1967; 202: 1028-34.

4) Echt DS, Liebson PR, Mitchell LB, et al: Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med* 1991; 324: 781-8.

5) ONTARGET Investigators, Telmisartan, ramipril, or both in patients at high risk for vascular events. *N Engl J Med* 2008; 358: 1547-59.

6) Higashi MK, Veenstra DL, Kondo LM, et al: Association between CYP2C9 genetic variants and anticoagulation-related outcomes during warfarin therapy. *JAMA* 2002; 287: 1690-8.

7) International Warfarin Pharmacogenetics Consortium. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med* 2009; 360: 753-64.

8) Allegra CJ, Jessup JM, Somerfield MR, et al: American Society of Clinical Oncology provisional clinical opinion: testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. *J Clin Oncol* 2009; 27: 2091-6.

9) Hung SI, Chung WH, Liu ZS, et al: Common risk allele in aromatic antiepileptic-drug induced Stevens-Johnson syndrome and toxic epidermal necrolysis in Han Chinese. *Pharmacogenomics* 2010; 11: 349-56.

10) O'Donoghue M, Braunwald E: Natriuretic peptides in heart failure: should therapy be guided by BNP levels? *Nat Rev Cardiol* 2010; 7: 13-20.