

第8章 信頼できる関連エビデンスをすべてレビューすること

1つの試験で果たして十分か？

端的に答えると1つの試験で十分なことは「めったにありません」。1つの公正な治療法の比較から、治療法の選択に関する決定の基準にするのに十分信頼性のあるエビデンスが得られることはめったにないのです。しかし、時には1つの試験から信頼できるエビデンスが得られることはあります。そのようなまれな試験の例として、心臓発作時にアスピリンを服用すると早期死亡のリスクが低下することを示した試験¹や、急性外傷性脳損傷の患者にステロイドを投与するのは致死的であることを明らかにした試験（下記および第7章 p73-74 を参照）があげられます。また、早産児の脳性麻痺を防ぐことが知られている唯一の薬物はカフェインですが、それを特定した試験も同様です（第5章 p47-48 を参照）。しかし通常、単一の試験は同じような課題を扱う複数の試験のうちの1つにすぎません。したがって、個々の試験から導き出されたエビデンスは、類似する他の試験のエビデンスと併せて評価すべきです。

公正な試験の先駆者である英国の統計学者 Austin Bradford Hill 氏は 1960 年代に、試験の報告は次の4つの質問に答えるものでなければならないと述べています。

- なぜその試験を行ったか？
- 何を行ったか？
- 何を発見したか？
- 結局のところ、試験は何を意味しているか？

なぜその試験を行ったか？

「臨床研究の科学的および倫理的妥当性のためには、答えを必要とする課題を扱うこと、意味のある答えが得られるようにデザインすること以上に基本的な原則はほとんどない。この2つの原則の前提は、関連する先行研究が適切に洗い出されていることである。（中略）既存のエビデンスの把握が不十分であると、研究の参加者から提供してもらう情報が必要なものであり将来他の人の役に立つという、参加者との暗黙の倫理的な契約を破ることになる」。

Robinson KA, Goodman SN. A systematic examination of the citation of prior research in reports of randomized, controlled trials. *Annals of Internal Medicine* 2011;154:50-55.

この4つの質問は今日でも重要な意味を持ちますが、これらを検証することは不十分であったり、すっかり見過ごされたりしがちです。最後の質問「試験は何を意味するか？」への答えは治療の決定や将来の研究に影響を及ぼす可能性があるため、特に重要です。

例えば、早産になりそうな女性にステロイド薬を短期間投与するという費用のかからない治療法があります。この治療の最初の公正な試験は1972年に報告されており、母親にステロイドを投与したところ、新生児の死亡率が低下したということです。その後の10年間に複数の試験が行われましたが、規模が小さいうえに個々の結果は紛らわしいものでした。いずれの試験も類似の先行試験を系統的に検討していなかったからです。もし検討していたなら、ステロイドの有益な作用を裏付けるとも強力なエビデンスが浮かび上がっていたでしょう。実際に系統的な検討がなされたのは1989年のことで、それまでの間、この治療にそれほど効果があることがほとんどの産科医、助産婦、小児科医、新生児看護師に認識されていませんでした。その結果、何万人もの未熟児が苦しみ、そして不幸にも命を落としたのです²。

「試験は何を意味しているか？」という質問に答えるには、ある公正な治療比較のエビデンスを、類似する公正な治療比較のエビデンスと並べて解釈しなければなりません。

研究から得た情報を統合する

百年以上前に、英国学術協会長の Rayleigh 卿は、新たな研究の結果を関連エビデンスの文脈に置く必要性について論評した。

「もし科学が、巷で考えられているように、労力をかけて事実を積み重ねるだけのものにすぎなければ、すぐに行き詰まり、それ自体の重さで潰れてしまうであろう。（中略）そこには、新たな知見を受け入れる過程と、既存の知見を消化し同化する過程の2つが並行してはたらいっている。この両方の必要性を受け入れ、相対的な重要性に関する議論を進めることができる。（中略）発見と説明を同時に進め、新たな事実を提示するだけでなく、既知の事実との関連を指摘することは、最大級の称讃に値するが、残念なことに必ずしも受けいれられていない。」

Rayleigh, Lord. In: Report of the fifty-fourth meeting of the British Association for the Advancement of Science; held at Montreal in August and September 1884. London: John Murray, 1884: pp3-23.

新たな試験結果を報告するときに、関連するエビデンスを考慮した解釈（系統的レビュー）を行わないと、有用な治療と有害な治療を見分けるのが遅れ、不必要な試験の実施につながります。

信頼できる関連エビデンスをすべて系統的にレビューする

1つの試験の結果は信頼できる他の関連エビデンスと並べてレビューすべきであると言うのは簡単ですが、これはさまざまな意味で難しい問題です。レビューが重要なのは人々の判断の拠り所にするためであり、系統的に行わないと人々の判断を誤らせてしまいます。

系統的レビューの重要性

「系統的レビューとメタアナリシスは、医療においてますます重要になっている。臨床医は自分の専門分野の最新情報を取り入れるためにそれらを利用し、臨床診療ガイドラインを作成する出発点として利用することも多い。研究助成〔資金援助〕を行う団体は、研究を進める妥当性を確認するために系統的レビューを要求することがあり、また、一部の医療専門誌はこの方向に進んでいる。研究がみなそうであるように系統的レビューの価値も、何を行ったか、何を発見したかと、報告の明確さによって決まる。また、他の出版物と同様に、系統的レビューの報告の質にもばらつきがあり、レビューの長所と短所を評価する読み手の能力を必要とする。

Moher D, Liberati A, Tetzlaff, Altman DG. The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement (www.equator-network.org), 2009.

系統的レビューによって、治療に関する同じような課題を扱っているように見えても、異なる結論に達することがあります。その原因は、取り組む課題が微妙に異なることや、研究者によって用いる手法が異なるためです。また、研究者が都合のよい解釈を結論に持ち込む場合もあります。そこで重要となるのは、自分の関心をもつ治療の課題に似た課題を扱うレビューを扱っていること、バイアスの影響や偶然の働きを上手く減らすように準備されていること、示されたエビデンスを反映するようにして正当な結論にたどり着いていることです。

系統的レビューでバイアスを低減する

バイアスは、個々の治療試験の結果を歪めて誤った結論を導き出すことがあります。同様にエビデンスのレビュー結果も歪める可能性があります。

例えば、研究者が自分たちの主張に沿った治療を支持するような試験を「都合良く選ぶ」こともできるのです。このような問題を回避するには、個々の試験と同様に、系統的レビューの計画も研究計画として作るべきです。この研究計画には、レビューの準備段階で、バイアスと偶然性の影響を低減するためにどのような方法をとるかを明示する必要があります。

これには、レビューで取り扱う治療の課題を具体的に示すこと、試験をレビューに組み入れるための適格基準、適格な試験の候補を見つける方法、組み入れる試験の選択とデータ解析の際にバイアスを最小限にするための手順などがあります。

系統的レビューの関連エビデンスをすべて洗い出す

系統的レビューの関連エビデンスを言語や報告形式を問わずすべて洗い出すことは常に困難を伴います。特に、関連エビデンスの中には発表されていないものもあります。過小報告の原因は主に、結果が期待外れだった場合に研究者が報告を書き上げないか、出版社に投稿しないことです。また、製薬企業は自社製品に都合の悪い試験を公表しません。

学術誌にも、結果が十分に「刺激的」でないと判断されて投稿論文が却下されるというバイアスの傾向がみられました³。

バイアスによる過小報告は非科学的、非倫理的であり、今ではこれが重大な問題であると周知されています。特に、治療を選択しようとしている人に誤解を与える可能性があります。それは、「期待外れの」あるいは「否定的」な結果が出た試験ほど報告が少なく、刺激的な結果の出た試験ほど「過剰報告」になるためです。

過小報告の程度は驚くほどで、全臨床試験の実に半数以上が正しく報告されていないのです。このような過小報告はバイアスであり、大規模試験にも小規模試験にも当てはまります。この問題に取り組むためにとられた手段は、試験を開始するときに登録する体制を確立することと、研究者に研究計画を公開するよう促すことでした³。

バイアスによる過小報告は人命に関わることさえあります。1993年に英国の一部の研究者が、13年前に実施した臨床試験の結果を報告する決意をしたのは称讃に値することでした。

マーケティングに基づく医療

「製薬業界の内部資料は、公表されたエビデンスの根拠が製品に関する裏付けデータを正確に表していないことを示唆している。製薬業界とメディカルプロモーション企業は、医学文献への掲載は主に市場での関心を引くためであると述べている。否定的なデータの抑制と都合のよい解釈、論文の代作〔第10章 p102-103 参照〕は、製品の販売に最適となるよう医学雑誌での発表を支援する手段として出現した。それと同時に、医師に対する病気の不適切な宣伝や市場の細分化も効率的に利益を最大化するために利用されている。エビデンスに基づく医療は崇高な理想であるが、マーケティングに基づく医療が現状の姿である。」

Spielmans GI, Parry PI. From Evidence-based Medicine to Marketing-based Medicine: Evidence from Internal Industry Documents. *Journal of Bioethical Inquiry* 2010;7(1):13-29. Available online: <http://tinyurl.com/Spielmans>.

この試験は、心臓発作をきたした患者を対象に心調律異常を軽減する新規の薬物に関するものでした。この薬物を服用した患者のうち9人が死亡したのに対し、対照群で死亡したのは1人のみでした。著者らはこのように書いています。「1980年にこの試験を行ったとき、被験薬群での死亡率の上昇は偶然の働きだと考えていた。当該薬物〔ロルカイニド〕の開発は商業的な理由で中止されたため、本試験が発表されることはなかった。これは現在では『発表バイアス』の典型的な事例となっている。ここに記載した結果は、（中略）その後起こった問題の早期の警告になっていたかもしれなかった」⁴。ここで言及されている「その後起こった問題」とは、そのロルカイニドの類似薬が最も多用された時期に米国だけで年間何万人もの早すぎる死を招いたことを指しています（第12章 p12を参照）⁵。

系統的レビューで偶然の動きを減らす

第7章 p74で、類似する別々の試験のデータを統合する「メタアナリシス」と呼ばれる手法によって、偶然性をどのように減らせるかについて説明しました。例としてあげたのは、5つの国で計画され、別々に資金提供を受けた5つの試験でした。早産児が重大な身体障害なく生存できる可能性を最大にするために、どの程度の血中酸素濃度が必要かという、60年来の難問に対処するものでした。この例では、試験がまだ終わっておらず結果が公開される前にこの手順がいかにかに計画されたかが示されましたが、一連の類似する試験が完了した後にも同じ手順を使用できます。

例えば、1974年にスウェーデンの医師が、乳がん手術に放射線療法を追加する場合としない場合の結果を比較する試験の系統的レビューを実施しました⁶。この医師は、検討対象とした全試験で放射線療法を受けた群の女性のほうが死亡率が高かったことに気づきました。メタアナリシスを用いて、このエビデンスをすべて統計学的に統合すると、この死亡率の高さは偶然の動きである可能性は低いことが明らかになりました。その後、個々の患者のエビデンスに基づいてさらに詳細な解析を行い、この時代の放射線療法が死亡率を上昇させていたことが確認されました⁷。この発見によって、安全な治療が確立されたのです。

系統的レビューの利益相反および情報の適格性を認識する

もしレビューを行う立場の人（査読者）が、レビューやその解釈に影響を及ぼしうる利益相反を有していたとしたらどうでしょうか？ 査読者がその新しい治療の開発元企業からお金を受け取っていたとしたら。湿疹に対する月見草油の効果についてエビデンスを検証した例では、査読者と製造会社に繋がりがあった場合、そのような商業上の関係のない査読者に比べ、はるかにその治療に対して有利な結論を述べました（第2章 p15-17を参照）。しかし、商業上の利益だけが審査のバイアスに繋がるわけではありません。研究者、医療従事者、患者も同様に、私たちはみな、先入観のために同様のことをやりかねません。

残念ながら、利益相反を有する人は、ある治療法を実際よりもよく見せるためにバイアスを利用することがあります（10章も参照）⁸。また、商業上の理由のみにかかわらず、実際に存在するエビデンスを意図的に無視する研究者もいます。彼らは、特定の治療研究を自らの研究結果で有利に見せようと設計、分析、または報告するのです。これは、1990年代、抗うつ薬 Seroxat（パロキセチン）において起こったことです。実際に、この薬を若者に投与することにより若年患者の一部でうつ病を苦しむ自殺を考える症例が増加したと示唆されましたが、同薬の製造会社はその重大なエビデンスを隠蔽したのです⁹。

過剰報告も問題です。「サラミ・スライシング」と呼ばれる現象は、研究者が、単一の試験から結果をいくつも抽出し（1つのサラミから同じスライスを切り分けるように）、報告がそれぞれ独立した研究の結果ではないことを明示せずに複数の研究報告として提出することです。こうして、「肯定的な結果」を示した単一の試験が、いくつものジャーナルに異なる論文として発表され、バイアスを誘導します¹⁰。ここでも、すべての研究に固有の識別子を使用して臨床試験を登録すれば、このような事態を誘導するのを減らすことができます。

信頼性が高く、関連性あるすべてのエビデンスでの評価がなされていないとどうなるでしょうか。

治療法の公正な検証のためには、信頼性の高いすべての関連あるエビデンスを系統的にレビューすること、そしてすでにわかっていることはなにかを把握するために、動物または他の研究室実験であるか、よくある健康なボランティアによる新治療法の検証であるのか、または患者を対象とした過去の研究の報告であるか、などを確認する必要があります。このステップが省略されたり、ずさんに行われたりすると、深刻な結果を招くことになりえます。一般の患者や研究参加者らが苦しんだり、時には不必要に死亡したりする可能性もあり、医療と研究の両方の貴重な資源が無駄に費やされることになります。

回避可能な患者の不利益

30年間にわたって教科書に掲載されていた心臓発作時の推奨治療と、同期間に報告された治療を公正に検証した結果の系統的レビューが行われていれば考慮されていたであろうエビデンスとの比較が行われました¹¹。

科学は累積的なものなのに、科学者は科学的に証拠を蓄積することをしない

「25年も前から、学術研究者のあいだでは、『累積メタアナリシス』について語られてきた。それは、ある介入に関するメタアナリシスを定期的実施することであり、1つの臨床試験が完了するごとにデータを差し代えて最新の結果を示す方法である。それにより、結果がどこに向かっているのかを把握することができる。また最も有用なことは、次の臨床試験が発表されれば累積的な統計学的有意性がすぐに明らかになることで、これ以上、不要な研究で人の命を危険に曝さずに済む」。

Goldacre B. Bad Science:How pools of blood trials could save lives.
The Guardian, 10 May 2008, p 16.

この比較により、その教科書の推奨事項に多々間違いがあったことが判明しました。このことは、著者らが関連するエビデンスすべてを系統的にレビューしていなかったために起こりました。そしてこの一件が与えた影響は悲慘なものでした。一部の症例では、心臓発作を起こした患者に救命救急処置（例えば、血栓溶解薬）も行われず、また別の症例では、死亡に至らしめる危険があると公正な検証により示されている治療（例えば、心臓発作発症患者の心拍数異常を減らす薬の使用など）が、その後何年も続けられていました（第2章 p12を参照）。

新たなエビデンスが創出されたにもかかわらず、その系統的レビューの研究結果を実地医療に反映しなかった場合、患者に不利益をもたらし続けることとなります。冷蔵保存や交差適合試験を必要としない血液代替物は、出血の治療にとって本物の血液にとって代わる可能性のある、明らかに魅力的な発明です。しかし残念ながら、これらの製品は心臓発作および死亡のリスクを増加させます。さらには、1990年代後半以降発表のランダム化比較試験を系統的にレビューしたところ、その血液代替物の危険性は数年前に認識できた、もしくはすべきであったことが明らかになりました¹。

回避可能な研究参加者の不利益

信頼性の高い、関連あるすべてのエビデンスの評価をしなければ、研究参加者に回避することができたかもしれない不利益をこうむらせることとなります。研究者らは、未だに、有効と判明している治療であっても、その治療を行わない群を含めた試験を利用し、そのような研究を依頼しています。例えば、腸の手術を受けている患者に抗生物質を投与すれば、手術の合併症による死亡率が低下することが確かなエビデンスで示されました。しかし、研究者らはその後も比較試験の対照群として参加者の半分を割り付け、抗生物質を与えない比較研究を漫然と継続していました。研究者が、過去に結果が出ている事実を系統的に見直すことができなかつたために、既知の有益な治療法があるにもかかわらず、研究参加者の半数にその治療を施行しなかったのです。この重大な過ちは、研究資金を提供し

た資金調達機関や、研究計画を見直し研究者に指摘すべきであった倫理委員会によって、明らかに見過ごされたのだと言わざるを得ません。

これから行なおうとする治療の有効性についてすでにわかっていることを研究者が系統的に評価しない場合、リスクにさらされるのは治療を必要とする患者だけではありません。健康なボランティアにも被害を与える可能性があります。基本的に治療法を検証する場合、はじめに実施される第1相試験には、非常に少数の健康なボランティアが参加します。2006年、ウェストロンドンにある民間研究施設の若い男性ボランティア6人に、これまで人に用いられたことのない薬物を注射しました。そして彼ら全員が生命を脅かす合併症を被ったのです。1人は指と足指を失い、長期にわたり健康が損なわれました。同様の薬物に対する重篤な副作用の報告が公表されていれば¹⁴、そして研究者がその薬物の効果について、すでに公表されている研究結果の系統的レビューを行っていたら、この悲劇はおそらく避けることができた¹³でしょう¹⁵。研究者らははじめから試験を計画しなかったでしょう。もしくは実施したとしても治験薬を一斉に投与するのではなく個別に1回ずつ投与したことでしょう。また、健康なボランティアに対し、副作用の危険性を話すことができたでしょう¹⁶。

医療と研究における資源の無駄遣い

信頼できる研究のエビデンスを系統的にレビューしなければ、患者や研究参加者に被害を与えないまでも不利益を生じるのは確実です。

最初に過去のエビデンスを調べることで死亡を防ぐことができたか？

「回避できたかもしれない悲劇的状況を招いた例として、ジョンズ・ホプキンス大学のゼンソクの研究がある。2001年6月に24歳の健康なボランティアの女性 Ellen Roche さんが死亡した一件である。彼女は指示されたように化学物質を吸引した結果、肺および腎臓機能不全の進行により息を引き取った。彼女の死亡後、実験を行った研究者とそれを承認した倫理委員会は、Roche さんに吸入させる化学物質のヘキサメトニウムの危険性についての多くの手がかりを見落としていたことがわかった。この一件がさらに悔やまれるのは、過去に出版された文献を調べればその化学物質の危険性のエビデンスを容易に見つけることができたのである。ボルチモア・サン紙は、主任研究者である Alkis Togias 医師がこの薬の副作用の研究をするために『誠実に努力』したものの、彼の検索は明らかに1966年以降の文献のみ検索可能な PubMed などの数限られたリソースを中心に調査していたと結論した。しかしながら実際には、1950年代以前の出版物で、かつその後も引用されている文献に、ヘキサメトニウムに関連する肺損傷についての警告がなされていたのである」。

Perkins E. Johns Hopkins Tragedy. Information Today 2001;18:51-4.
that were more likely to be relevant to identifying improvements

なぜなら、医療および医学研究における資源の無駄になります。例えば、1980年代および1990年代に、合計8,000人以上の患者が、脳卒中の新薬候補を検証するいくつかの試験に参加しました。オランダの研究者は、同薬を研究した結果を洗い出し、系統的にレビューしましたが、有益な効果を見出すことができませんでした(第10章 p121を参照)¹⁷。その後、彼らは動物で以前に行われた同薬の実験結果を再検討しましたが、やはり、有益な効果は認められませんでした¹⁸。動物実験を実施した研究者および臨床試験を実施した臨床医が動物実験の結果の系統的レビューを行っていれば、何千人もの患者が臨床試験に参加せずに済んだでしょう。確かに、脳卒中患者の治療のため、また同疾患のより良い治療の特定につながる研究のためにもっと有効に資源を活用できたかもしれないのです。そして、これは全く特別な例ではありません¹⁹。

新たな研究報告は、系統的レビューで始まり、系統的レビューで終わる

ステロイド投与の急性脳外傷に対する有効性を評価する研究報告²⁰では、Bradford Hillの法則(4つの質問)すべてに回答するためのやり方が示されています。研究者らは、同治療の多様な臨床応用のエビデンスを含め、既存のすべてのエビデンスを系統的にレビューした結果、現在一般的に使われている治療の有効性について重大な不確実性がみられたため、この研究に着手したと説明しています。彼らは、レビュー開始時にはすでに研究計画の登録および発表を完了していたと報告しました。

医学雑誌「The Lancet」の編集者による、著者への研究成果を文脈に入れる

系統的レビュー

この部分では著者がすべてのエビデンスをどのように検索したかの説明が含まれていなければなりません。また著者らはどのようにエビデンスを選択し、どのようにそれらを組み込んだかなど、そのエビデンスの質の評価方法についても記載すること。

解釈

著者は、彼らの研究が過去の研究に加えられることにより、エビデンス全体に対して何が追加されるかをここに述べるべきである。

「8月1日以降提出されるすべての研究報告において、ランダム化比較研究であろうとなかろうと（中略）『考察』にて、その結果をエビデンス全体の背景の中に組み込むよう求める」

Clark S, Horton R. Putting research in context – revisited.
Lancet 2010;376:10-11.

彼らは、必要な数の患者を試験に組み入れることにより、バイアスを最小限に抑えて偶然の可能性を十分にコントロールし、重篤な脳損傷を有する患者に投与されていたステロイドが、実は患者が死亡する確率を高めていたことが自分たちの研究で示されたと発表しました。

最後ですが忘れてはならないのは、彼らは、現在一般的に使われている治療により何千人もの人が命を落とすことがないように行動を起こすべく、そのために必要なすべてのエビデンスを読者に提供したことです。つまり、以前の研究の数々を精査したこれまでの系統的レビューを更新し、自らの研究で見出した新たなエビデンスをそこに組み込んだのです。

キーポイント

- 単一の研究で、医療の治療選択を導くのに十分なエビデンスが得られることはほとんどない。
- 新しい治療の相対的メリットの評価は、関連性があり信頼性の高いすべてのエビデンスの系統的レビューに基づいて行われるべきである。
- 治療を検証した個々の研究同様、系統的レビューではバイアスや偶然性による誤りへ導く影響を減らすためにさまざまな措置を講じる必要がある。
- 系統的レビューの結果を考慮しなかったことにより、結果的に患者に回避可能な不利益を被らせ、医療と研究の資源を無駄にした。