

## (2) 医学論文の質の評価方法と実例

### RCTとCCT

つたにきいちろう  
津谷喜一郎



東京医科歯科大学、研究者

もともと漢方薬や鍼灸など東洋医学の臨床評価の方法論に関心があり、日本での臨床試験の開拓者の一人である佐久間昭氏に学んだ。「二重盲検法ではなくその基本にある無作為化が大事なんだ」と佐久間氏に言われてから20年になる。臨床試験の分野というのは当時も今も人不足で、東洋医学だけでなく近代医学の評価、さらにはそれらのシステムティック・レビューを行うコクラン共同計画にまで活動が広がった。

#### 1. エビデンスとは

しばしばEvidence-Based Medicine (EBM)は何か、ということが議論されます。私は漢方医学にも関心がありますので、そちらのほうの言葉を使って説明してみましよう。「親試実験」という言葉があります。江戸時代の古方家が用いた言葉です。つまり、空理空論ではなくて本当に効くのかどうかを試してみようという考えです。また、「实事求是」という言葉があります。こちらは中国語で、清朝の考証学者が使った言葉らしいのですが、物事の事実を求める、エビデンスを見る、というわけです。この言葉は、毛沢東も使ったようです。

今回のセミナーの2日目に「よい臨床試験とは」というタイトルで話しました。その中で医療において意思決定するとき、いろんな情報がモザイクのようになっているとお話しました。医学情報のモザイク・モデル(mosaic model)です。

図-1は、エビデンスの強さにレベルがあることを示すものです。臨床疫学のテキストにはよく載っていると思いますが、図の下へ行くほどエビデンスレベルが高い、最も高いのはメタアナリシスというものです。医学情報のリニア・モデル(linear model)です。

ただし、注意しないとイケないのは、よく「二重盲検で効いた」といいますね。ところが、みなさんもお存知のように、二重盲検法をもち

いた臨床試験といっても、その質にはいろいろバラつきがあるのです。

また、システムティック・レビューではその統計学的な部分をメタアナリシスとよんでいますが、広い意味でのメタアナリシスについても、質のバラつきがあるわけで、質の悪いメタアナリシスやシステムティック・レビューというのは、一見、エビデンスが高そうに見えてしまう分だけ、よりたちが悪い、ない方がいい、ともいえます。

#### 2. RCTとCCT

今まで何度もRCTやCCTという言葉が出てきましたが、randomized controlled trialとcontrolled

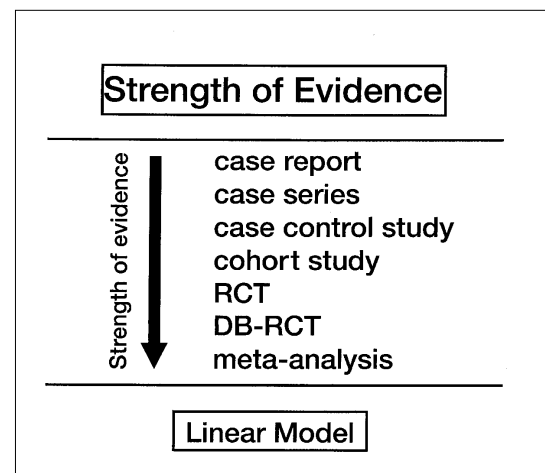


図-1 エビデンスの強さ

clinical trialの略称です。米国医学図書館(National Library of Medicine: NLM)の作成しているMedlineという医学データベースがあります。この中ではMeSH(Medical Subject Headings)という、キーワード集が使われています。また各書誌データのフィールドの一つに、publication type、つまりその医学研究がどのような研究デザインでなされたものかを示すものがあります。

ここで、1995年に一部、修正がなされ、RCTの再定義がなされました。図-2に示しましたように“Using mathematical techniques such as a use of random number table”、つまり、乱数表を使ったり、コンピュータで乱数を発生させたり、そういったmathematical数学的な方法によるものをRCTと呼ぼうというわけです。

そしてそれ以外の方法によるもの、コインを投げて表か裏か、患者さんの生年月日などの奇数が偶数か、あるいはsocial security numberこれはアメリカ的なものですが、国民皆番号のようなものを使う、曜日、例えば月曜だったらA群、火曜日だったらB群に分ける、カルテ番号で振り分けるなど。

しかし、コインが曲がっているかもしれませんし、月曜の患者と火曜の患者は明らかに違いますね。つまり、ランダム割付けに関して少しレベルが落ち、セレクション・バイアスが入る可能性が高いものです。pseudo(偽)とかquasi(準)、いわば、「RCTもどき」となっているということで、こういったものはRCTとは呼ばず、controlled clinical trial(CCT)なる用語を作って、これで呼ぶようになりました。

Social security numberというのはアメリカ的だと言いましたが、こういった手法も各国の状況を反映しているものです。日本では封筒法というのがこれまで結構使われてきました。

その結果の一例を図-3に示します。中央の縦に書いたのが施設番号。左側が実薬(Active:A)が投与された患者数。右側がプラシーボ(Placebo:P)が投与された患者数です。よく見ると、一番上の6番の施設は実薬のみ、5番の施設も実薬のみ。10番の施設は実薬が12例、プラシ

### MeSH/Publication type - 1995

#### RCT (Randomized Controlled Trial)

Using mathematical techniques such as the use of a random number table

#### CCT (Controlled Clinical Trial)

Employing treatment allocation methods such as coin flips  
odd-even number  
patient social security numbers  
days of the week  
medical record numbers  
or  
other such pseudo- or quasi-random process

図-2 RCTとCCTの定義

Distribution of treatments		
AAA AAAAA	6	
A	5	
AA AAAAA AAAAA	10	PPPP
AAAA	9	PPP
AAAAA	3	PPPPP
AA	1	PP
A	12	P
	7*	
AAAAA	8	PPPPP PP
A	2*	PPPPP PP
A	11	PPPPP PPPPP
	4*	PP
40		41

\* Leading Institutions ?  
津谷 喜一郎; 薬効評価と情報の役割, 薬学図書館, 38 (3) : 145 - 156, 1993

図-3 封筒法による結果の例

ーボが4例ということです。なぜこういうことが起きるかという封筒法だからです。封筒を破って、Aと書いてある、では使おう。ところがPとあると、これはまずい、この患者さんにはAを使いたい。と、つぎの封筒を破ってしまう。出てこないとまた破る。Aが出てくるまでいつまでも破る(会場笑い)。

図-3の下の方には逆にプラシーボが多い施設が出てくる。これはなぜだと思いますか? これは全体として2群の例数を合わせるためなのです。つまりある施設は、プラシーボが出るまで封筒を破っている。この例数を見ただけで、当然2群の背景因子にバイアスがあると思われる。実はこれは丸山ワクチンの中では一番質の良いトライアルなのです。他にはるくなものがない。封筒法では、割付け表(randomization list)

医薬品評価の方法 その(3) (M・N・O)

は、通常、コンピュータなどでmathematicalに作られます。しかし実際の現場ではそのとおりには使われないということです。さきほどのMeSHのRCTとCCTの定義は、割付けの順序(randomization sequence)の方法での区別です。したがって、封筒法はRCTとなりますが、その実施の段階でランダム化が崩れているということになります。日本ではこういうことにも注意をはらうべきです。

### 3. コントロールになにを？

コントロールに何を持ってくるか。3種類あります。第1に、無治療群。第2に、プラシーボ群。第3に、アクティブ・コントロール群です。

第1の無治療群はその名の通り何も治療をしないものです。ただ薬物の臨床試験では、あまり使われません。ブラインドネスが保ちにくいのと、臨床試験へのコンプライアンスが悪くなる、つまり、何もされないと患者さんは病院へ来なくなるためです。すでにある治療をしていて、その「上乘せ」で、片群は実薬、他群は無治療というのはたまにあります。

第2はプラシーボです。日本ではプラシーボというとネガティブな意味で使われることが多いようです。図-4を使って説明しましょう。元々、*placebo*は、ラテン語の*placere*の一人称、単数、未来形です。キリスト教の夕べのお勤め、晩課で、死者を弔う時の聖歌の一章です。神に召されて喜ばしいといった意味の言葉です。“I shall please”というわけです。でも、洋の東西を問わず死ぬことは喜ばしくはないですね。それで、歴史的に意味が変わって来て、中世には人を喜ばせるために、おべっかを使ったり、へつらったり、嘘をつく、そういうネガティブな意味が生じました。いわば、ダブル・ミーニングとなったわけです。

医学の領域では、1811年にHooperが、“more to please than benefit”、つまり、「ベネフィットを与えるよりもむしろ喜ばせるもの」と定義しています。

近代の薬理学は19世紀後半から始まりました。

昔の生薬や漢方薬は、飲んで効くとき、その「効き目」の説明が、寒熱や乾湿のバランス、陰陽五行説といった形而上的なものではなされていきました。それが近代科学にもとづく、解剖学、病理学、生理学、生化学などの用語を使って説明されるようになりました。物的なメカニズムに働きかけて効く、「作用機序」(action mechanism)をもつという考えが出てきたのです。そうするとaction mechanismに働きかけるものではない、「アクティブ(active)ではないもの」、inactiveという概念が生まれます。それがプラシーボの意味になったわけです。この流れで、ドイツ語では、空の物質、外見のみの物質、示唆する物質などの用語が使われたのです。現在ではプラシーボ(Plazebo)が普通です。

英語ではダミー(dummy)ということもあります。日本語では、偽薬、偽型薬、凝薬、擬薬、模倣薬、囮薬、姑息薬、気休め薬、だまし薬、かくし薬(会場笑い)。これ、みんな悪い訳ですね。なぜ悪いかといいますと、一番の理由はプ

Placebo		
Latin	: <i>placēre</i> → <i>placēbō</i> (1811 Hooper: more to please than benefit)	
German	: (modern pharmacology) “action mechanism” Leermittel (空の物質) Scheinmittel (外見のみの物質) Suggestivemittel (示唆する物質) Falsum (偽物) Plazebo	
English	: dummy	
日本語訳	偽薬	模倣薬
	偽型薬	だまし薬
	凝薬	囮薬
	擬薬	かくし薬
		姑息薬
		プラセボ
		気休め薬
		プラシーボ
中文訳	: 安慰剂 (ānwèijì)	
韓国語訳	: 偽薬 (위약, wüyak)	

〔津谷喜一郎：プラセボの日本受容—placeboはのり薬だ—。In: 山田慶児, 栗山茂久(編)：歴史の中の病と医学。思文閣出版, 1997〕

図-4 Placeboの各国語訳

ラシーボというコンセプトが日本では臨床試験からひろまったのです。臨床の場ではなく、二重盲検ランダム化比較試験の中で用いられたので、こういうネガティブな訳がついたのです。

中国語訳は、安慰剂。I shall pleaseという意味が入っていてよい訳だと思います。

世界中の華僑世界では安慰剤が使われています。韓国は日本と同じで偽薬で、これは日本からの影響です。つまり、日韓のみがplaceboに対してネガティブな意味の言葉を使っています。日韓で、世界中の良い意味合いのものに対抗している形です。プラシーボで誤解されることがあります。

プラシーボには副作用がないと思っている人が多いことです。しかし、プラシーボにも副作用があります。二重盲検ランダム化比較試験では、たとえば、実薬に消化器系の副作用が多いとプラシーボにも同じ様な副作用がでできます。被験者も主治医も人間ですから、結果が実薬の方に無意識に引張られるようです。

第3のアクティブ・コントロールとは何でしょうか。「実対照薬」と訳されますが、標準薬などが使われます。ICHの統計ガイドラインでは、このアクティブ・コントロールを用いるに当たってはつぎのことに注意しろと言っています。まず慎重に選ぶということです。広く使用されている治療法であること。十分に計画され記録されている一つ以上の優越性試験(superior test)によって、適切な適応に対する有効性がはっきりと確立され定量的に示され、現在計画されている臨床試験でも同じ有効性を示すことが十分に期待されるもの。こういった条件を満たすものをアクティブ・コントロールとすべきだということです。

こういった議論がICHで起きたというのも、別府さんや浜さんが、日本ではっきりしない標準薬を使ったトライアルが多いと批判したこともそれなりに影響を与えていると思います。これは世界的にもある問題ですが。

比較試験にも3種類あります。第1に、先にお話しした、優越性試験。つまり、あるものに優

るかどうかを比べる。ここで、あるものとは、無治療、プラシーボ、またはアクティブ・コントロール、そうしたコントロールに比べてよいか、superiorかと調べるものです。

第2は同等性試験です。コントロールに対してある幅を決めるわけですが、それを、もちろん劣ってはいけないしですし、優ってもいけないものです。幅の中央に入る、つまり効きすぎてもいけないのです。

第3は非劣性試験です。ある幅から劣ってはいかん、というものです。幅より効く分には構わない。昔は日本では第2の同等性試験とこの第3の非劣性試験と合わせて、同等性試験と言っていました。しかし同等性試験では効き過ぎてはいけないものです。幅の中に納まるのが同等性。非劣性試験というのは下限だけを決めています。そこで別に分けて考えるようになったのです。ここで、これらの幅は「その中に納まってもらいたい」と医学的に決めるものです。

### 4. エンドポイント

時間もなくなってきたので、エンドポイントについて簡単に触れます。全般改善度の話が昨日もありましたが(2日目のG.H、第4章など)、エンドポイントは、いくつかに分類することができます。一つの分け方は、代理エンドポイント(surrogate endpoint)と真のエンドポイント(true endpoint)です。例えばコレステロール値は代理のエンドポイントで、心筋梗塞の発症が真のエンドポイントです(編集部註)。このサロゲートとトルーの関係は、血圧値と脳卒中、血糖値と糖尿病に基づく合併症の発生、癌縮小効果と生存年、など色々なものについていえます。

もう一つ、ハードかソフトかの分け方があります。生死、各種の検査値、画像診断がハードデータといわれるものです。一方、むしろ患者さんの主観的なものを大事にしよう、心地良さだとかQOLといわれるものが、ソフトデータです。現在、QOLのためのスケールがいろいろ考えられたりしています。いずれにしても、患者の利益が反映されたエンドポイントでないとい

けないわけでは

コクラン共同計画の一つの特徴は、消費者の関与です。運営委員会 (Steering Committee) にも入っていますし、各レビューグループなどにも入っています。例えばレビューをするときに、何をエンドポイントにしてレビューするかは患者の意見が反映されるべきだ、ということです。

おわりに

図-5にあるMCPDBRPCTとは何でしょうか？私は、漢方や鍼の評価にも関心がありますが、コクラン共同計画の相補代替医学 (complementary and alternative medicine: CAM) のフィールドでディスカスしている時に、冗談で言いはじめたものです。Multicenter Cooperated Parallel Double Blind Randomized Placebo Controlled Trialの略です。多施設共同平行二重盲検無作為化プラセボ比較試験。仰々しい限りですが、この手法の中には、いろんなバイアスを避けるための手だてが入っているということです。

しかし、実は臨床試験の原則というのはそれほど難しいものではありません。基本的な要素を押さえれば、容易に理解できるものなのです。

MCPDBRPCT ?	
M : Multi-center	多施設
C : Cooperated	共同
P : Parallel	平行
D : Double	二重
B : Blinded	盲検
R : Randomized	無作為化
P : Placebo	プラセボ
C : Controlled	比較
T : Trial	試験

多施設共同平行二重盲検無作為化プラセボ比較試験

図-5

編集部註：日本人はコレステロール値が低い方が死亡のリスクが高く、240mg/dl以上が最もリスクが低いので、コレステロール値を下げて心筋梗塞が減少しても死亡全体は増えることになる。したがって、日本人にとってコレステロール値に関する真のエンドポイントは全死亡であると考える；TIP誌14号No.6, 61-70, 1999)

## メタアナリシスの具体例

脳循環・代謝改善剤とグルコシダーゼ阻害剤を例に ———— ありがさひでき 折笠秀樹



富山医科薬科大学医学部、臨床疫学研究者

医学統計学を専門としながら、臨床医学研究者といる臨床研究を行っている。著書に『臨床研究デザイン』(真興交易医書出版部 1995)、『ニュースの統計数字の正しい読み方』(「ニュースの統計数字を正しく読む」、ヴィクター・コーン著、折笠秀樹訳、バイオスタット社 1996)、『臨床試験とは何か』(Curt Furberg 著、南江堂 1994) など。

メタアナリシスとシステムティックレビューについて  
今日は、メタアナリシスはどういうものかという話をまずしまして、その後でふたつのメタアナリシスを試みとしてやってみたので紹介したいと思います。メタアナリシスというのはいろんな呼ばれ方をしており、システムティック・レビューという言葉を使う人もいますし、論文のタイトルにも使用されます。

普通のレビュー(総説)とシステムティック・レビューの違いを言います。普通のレビューは定性的にやるのに対して、システムティック・レビューはきちんと文献検索をして網羅的に関連する文献を見つけ、それを定量的にレビューする、という点が違います。言葉の使い分けは人によって違いますが、わたしなりに整理しますと「系統的な文献の収集に始まり、定量的なデータを統計学的方法で併合して何らかの総合的な数字を出すもの」がシステムティック・レビューと言えます。総合評価(オーバービュー)と言われることもあります。

こういうようなシステムティック・レビューなりメタアナリシスというものが、最近、クリニカル・プラクティス・ガイドライン、診療指針として学会や政府団体で使われておりますが、いろいろ利用され応用されてきているのが現状だと思います。

システムティックレビューの目的

目的にはいろいろあります。サブ・グループ解析を十分なパワーで行うのも一つの目的なのですが、治療効果の結果がいくつかの試験ではらついていて、全体としてはどのようなことになるかわからないという場合に役立ちます。もしも結果がずっと同じ傾向で出てくる場合はより信頼の置ける情報となる。ただし、これには条件があります。ステイブルな(安定した)情報が集まるというのは、違う研究者がやっても同じような結論がずっと続いて来るといときにだけ当てはまります。違う結果が出てきたときには、また議論の余地が出てきます。

例えば、数年前の例ですと、急性心筋梗塞に対してマグネシウムの静注という治療法、これは過去のデータではずっと良いというデータが出たんだけど、ISIS-4(アイシスフォー)という4万人くらいのスタディでは悪いというデータが出ています。その時に、深く考え直す材料を提供するということがあります。メタアナリシスでは治療法の違いは許容してやっておりますので、そういう違いも許容した上で結論がどうか、というのが出ますので、一般性があるのですが、真実については分かりにくくなります。

ある特殊な患者層に対して効くのか、別な人には効かないのかという解析が、多数の症例になってきますと少し詳細な検討ができるように