

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 ———— 折笠秀樹 ありがさひでき



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ばらついていて、全体としてはどのようなか分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

なります。パワーが高いためです。

システマティック・レビューとガイドライン

今年（97年）にアメリカの内科学会誌に出たもので、システマティック・レビューとブラクティス・ガイドラインがどういふふうに関わってくるかを示した論文があります。

例えば不安定狭心症というやや重い狭心症ですが、その患者さんにアスピリンを投与するのは当然のことだけでも、それプラス最初の4～5日間は抗血栓剤も投与するのがいいのかという題材です。実はガイドラインの方に、まず、いいんだという指針が載ってしまった。その後メタアナリシスが出たという例もありました。しかし通常は逆であり、メタアナリシスが発表された後にガイドラインとなっていくことが多いようです。

メタアナリシスに関するもので、クリニカル・エビデミオロジー（Clinical Epidemiology）という本の第3版に、臨床疫学をよくやっているフレッチャー博士らによって、メタアナリシス論文の読み方が書かれています。

解決すべき問題の種類と研究デザインの選択

表-1には、どういふ問題を解決するときはどういふデザインが適しているか、その主なものを示しています。例えば、薬の副作用に関する関連性、あるいはO-157や食中毒の関連性を調べる場合には、CAUSEおよびRISKを調べたいわけですが、その場合にはRCTはまずできないし、ないわけです。その時はコホートかケース・コントロールを最大のエビデンスとして使って総合的に評価する。

臨床上の疑問	適切な研究デザイン（主なもの）
診断	断面研究（cross-sectional）
有病率	断面研究（cross-sectional）
危険性	症例対照研究（Case-control）
予後	追跡研究（Cohort）
治療	ランダム化比較試験（RCT）
予防	ランダム化比較試験（RCT）

（RCT：Randomized Controlled Trial）

表-1 臨床上の疑問に応じた研究デザインの選択

例えば、treatment（治療法）prevention（予防法）などに関しては、やはりRCTを行うべきで、行った成績に応じて総合評価すべきである。問題別に使っている研究、エビデンスというものをどこにするか、という設定が違ってくる。

そこで今日わたしの話で例題に出すのはtreatmentです。その場合にはRCTは行うべきですので、RCTの行われたものを使って話します。

問題になっている脳循環・代謝改善剤と、食後の過血糖を抑える作用を持つアルファグルコシダーゼ・インヒビター（阻害剤）というものに関しての国内のデータに関して総合評価を試みました。

わたし自身は、この薬は危険であるとか、この薬は効く、だとかいう片方のイメージを先に持つというより、どちらか分からないのだけでも公平な立場からデータに基づいて科学的な評価をしたい、という姿勢から始めました（編集部註：日本の臨床試験にはいろいろと問題があり、そもそもシステマティックレビューを行える程の質を備えているかどうかについて大いなる疑問はあるが、その問題はここでは一応保留にしておいて、全般改善度をエンドポイントとしてメタアナリシスをした場合にどのようなかについて報告を頂いたことを付け加えておきます）。

ただ、安全性の場合には、有効性と違ってシグナルが出た場合に、早くストップをかけることは重要であるという立場から言えば、迅速な対応をするときには、統計学的立場から言うと、

という言い過ぎ（誤り＝第1種過誤）の基準値を少し上げる（大きい値にする＝緩める）という見逃し（誤り＝第2種過誤）の基準値を少し下げる（小さい値にする＝厳しくする）という操作をして、危ないものをチェックする態勢をとるべきだと思います。

今回の分析は治療についてですから、分析すべき文献のソースはRCTに限ります。中でも主観的な尺度を扱うので、結果の信頼性や妥当性を確保する意味でプラシーボ対照のものを選びました。

アルファグルコシダーゼ阻害剤

まず最初に グルコシダーゼ阻害剤についてお示しします。1996年の臨床薬理学会で発表したものです。

グルコシダーゼ阻害剤というものが数年前から日本でも発売されて使われているのですが、日本のデータだけに限りまして、安全性と有効性を

評価してみました。いろいろな方法でそれに関する論文をチェックしました。プラシーボを対照にしたRCTに限定しております。

二種類のアルファグルコシダーゼ阻害剤が市販されています。アカルボース2試験とボグリボース3試験、合計5試験が1996年に調査した段階では検討の対象になりました。

表-2は研究デザインのサマリーです。作用機序は同じですが、同じ薬剤ではないのです。ドーズ（用量）に関しては、例えば、ボグリボースでは1回0.2mg、0.1mg、0.05mgのように一部異なりますが、そのくらいの違いは無視しています。患者さんの条件は、日本の試験というの

Design summary of 5 RDBPCTs of α -glucosidase inhibitors in JAPAN

Active drugs	Dose	Patients	Rx period	#Hosp.	Size
1. acarbose (Bay g5421)	100mgx3/d	PPPG \geq 200mg/dL	8wk	74	A: 122 P: 124
2. acarbose (Bay g5421)	100mgx3/d	PPPG \geq 200mg/dL	28wk	NA	A: 20 P: 20
3. voglibose (AO-128)	0.2mgx3/d 0.1mgx3/d 0.05mgx3/d	PPPG \geq 200mg/dL variationFBG \geq 30mg/dL	4wk	95	Ahi: 101 Amd: 104 Alo: 103 P: 108
4 voglibose (AO-128)	0.2mgx3/d	PPPG \geq 200mg/dL variationFBG \geq 30mg/dL	8wk	81	A: 119 P: 131
5. voglibose (AO-128)	0.2mgx3/d	PPPG \geq 200mg/dL variationFBG \geq 30mg/dL	28wk	NA	A: 44 P: 45

PPPG: Post-Prandial Plasma Glucose; FBG: Fasting Blood Glucose; Rx: Treatment; A: Active; P: Placebo; Variation of the FBG was assessed within a 2 to 4-week run-in period.

表-2 グルコシダーゼ阻害薬の糖尿病に関するランダム化プラシーボ対照試験5報のデザイン要約

は画一的なところがあって助かるのですが、すべてPPPG（Post-prandial plasma glucose）、つまり食後2時間の血糖値が200mg/dl以上、下の3試験はそれに加えて空腹時血糖のバラツキが30mg/dl以下となっています。

治療期間は最短が4週、最長は28週ですが、あと2つは8週です。サイズは最少40例から最大でも400例で、全部でも1000例位しかありません。

血糖の改善効果の指標の定義は面倒ですがやり方は5試験とも全部同じです（表-3）。食後1時間値、食後2時間値、空腹時の3種類の血糖値を合成した指標を作っています。まずたとえば

〔1〕a)食後1時間の血糖値（mg/dl）の上がり

具合と、〔2〕a)最終的な血糖値（絶対値）とを

5段階の評価を全般改善度式に求めております。

それ自体はあまり賛成は

できないのですが、今回

すべての試験で同じよう

なエンドポイントを定義

しているということにと

りあえずこのエンドポイ

ントで進めたいと思いま

す。つぎに〔3〕b)下が

り具合と、最終血糖値の

組み合わせを点数化しま

Definition of the endpoint

〔1〕PG Lowering effect (in unit of mg/dL)		〔3〕Scoring of three items (PPPG1hr, PPPG2hr, FPG)	
a) PPPG1hr	b) PPPG2hr	Lowering Absolute effect	
A: -50 or more	A: -40 or more	A	B C D E
B: -49 to -25	B: -39 to -20	0	1 1 1 2
C: -24 to -13	C: -19 to -10	B	1 1 2 2 2
D: \pm 12	D: \pm 9	C	2 2 2 3 3
E: +13 or more	E: +10 or more	D	3 3 3 3 3
		E	4 4 4 4 4
〔2〕Absolute effect by achievement level of PG (in unit of mg/dL)			
a) PPPG1hr	b) PPPG2hr	c) FPG	
A: < 160	A: < 160	A: < 120	
B: 160 to 199	B: 160 to 199	B: 120 to 139	
C: 200 to 249	C: 200 to 249	C: 140 to 164	
D: 250 to 309	D: 250 to 309	D: 165 to 194	
E: \geq 310	E: \geq 310	E: \geq 195	
〔4〕Score of CGI = 11 x (PPPG1hr score) + 8 x (PPPG2hr score) + 6 x (FPG score)			
Score of CGI			
Remarkably improved:		0 to 19	
Moderately improved:		20 to 39	
Slightly improved:		40 to 69	
Unchanged:		70 to 89	
Worsening:		90 to 100.	

表-3 糖尿病の臨床試験における合成エンドポイントの定義

す。b)食後2時間値や、c)空腹時血糖値についても同様に、〔3〕の点数を求めます。そして、〔4〕この3種類の得点にある係数をかけたものをCGI(クリニカル・グローバル・インプレッション)というスコアとします。そのスコアを5つに切って日本流の改善度に置きかえています。総合得点は、これに意味があるかどうかはわからないのですが、最終的に使ったのは総合得点が40点未満の有効例とそれ以外、というものに分けて解析をしました。欠損値の場合にはとりあえず無効の処理をして解析しております。

有効性については、先ほど定義したような有効、無効の定義をして、5つの試験について何例中何%で有効だったかという数字を出して、差

の指標に関して図-1に示します。1番目のアカルボースNo.1というスタディでは、30%くらいプラシーボよりも実薬が上回っている。図の横棒の両端は95%信頼区間です。2番目の試験もプラシーボより有意に優れている。3番目は五分五分。4番目の試験はやはり有意に優れて、5番目も有意です。これに関してはそんなに見解の不一致はないと思いますが、全体に20%くらいアルファグルコシダーゼ阻害剤というものはプラシーボよりも有意である。有効性のほうは適格違反などを除いていますから832例になっていますが、有意差が見られた。

安全性のほうは図-2に示しました。多少興味があったのですが、消化管からの吸収を抑えるというので、消化管にいろいろな有害作用が起きるのではないかと、また肝機能でも副作用の報告がなされていたので、それがどのくらい出ているかとむしろ興味がありました。有効性はまず間違いはないだろうと思っていました。

全体での安全性、有害事象がどのくらい出ているかという、アカルボースNo.1で118例中58例、プラシーボが120例中24例でした。これもリスク差を出していますが、29%も実薬で有害事象が発生しています。図-2で見ていただいたほうが分かりやすいでしょう。右側に行くほど実薬のほう危険である。

ということは、1番目の試験でいうと30%

くらい副作用は実薬のほうが高かったわけです。2番目は少ない症例数なので信頼区間が広がっています。3番目も有意差があって実薬のほう危険である。4番目、5番目は有意差はないけれども多少実薬のほう危険である。

全体を総合したのがこの菱形ですが、ここで問題なのは、メタアナリシスでただ単に菱形だけを見れば十分ではないということです。この場合、アカルボースという最初に出たものと後に出たボグリボースでは、かなり安全性のプロファイルが違うというのがこれで読み取れます。スタディで相当違う。これが統計学的にはヘテロジェニアティ(heterogeneity)というテスト、異質性の検討というのですが、そういうものを使うと出てきますし、図を見れば一目瞭然です。

安全性のどういうものが出ているかと言うと、やはり消化器系のシンプトム(症状)がいっぱい出ておりました(図-3)。腹部膨満感などもこれは5つの試験を併合してのメタアナリシスを行ったあとでのリスク比(副作用の比の指標)に直していますが、プラシーボより2倍以上というリスク比が出ております。放屁の増加というのも3倍近く、軟便はバラツキが多くて有意差にはなってありませんが3倍くらいです。下痢も少々、トランスアミナーゼは数字としては多少リスクが高いほうにいらしていますが、人によってバラ

ツキが多くてははっきりしたことは言えない。右側へ行くほど実薬のほうプラシーボよりも危険ということです。

第1番目のアルファグルコシダーゼ阻害剤というのはわずか5試験で1000例弱の症例数なのです。日本で試しにやってみたということです。もともと外国の薬なのですが、外国になりますとこういう薬剤の有効性の効果となりますと、ほぼ血糖値になります。HbA1cというのが糖尿病のコントロール状態をみるのに一番重要とされている(サロゲート)エンドポイントなのですが、日本の場合は、その値に関する数値はあまりたくさん示されていないので、論文から使うことはできない。しかし、個別のデータを取り寄せればそういう解析をできる可能性もある。これは日本独特の一貫したエンドポイントがあったので、それに対してやってみたという一つの例です。(編集部: HbA1c値が未だに、糖尿病のサロゲートエンドポイントとして採用されていない日本の臨床試験論文は問題であろう)

脳循環・代謝改善剤

次は脳循環・代謝改善剤です。こちらのほうは試みてはみたものの、脳循環・代謝改善剤の効果に関していろいろ疑問が上がっています。メタアナリシスというのはあまりディシシティブ、結論的なものというよりも探る方に使うのがいいのではないかと、という個人的な気持ちを持っています。脳循環・代謝改善剤の中には、いろいろの薬理作用をもつものがあります。抗血小板作用を持つものやカルシウム拮抗作用を持つものなど

ですが、いわゆる脳循環・代謝改善剤とくくられているものを集めました。

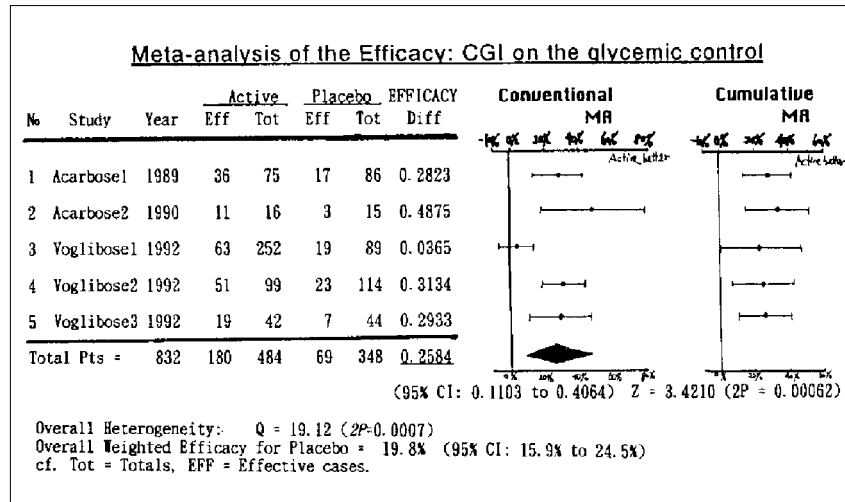


図-1

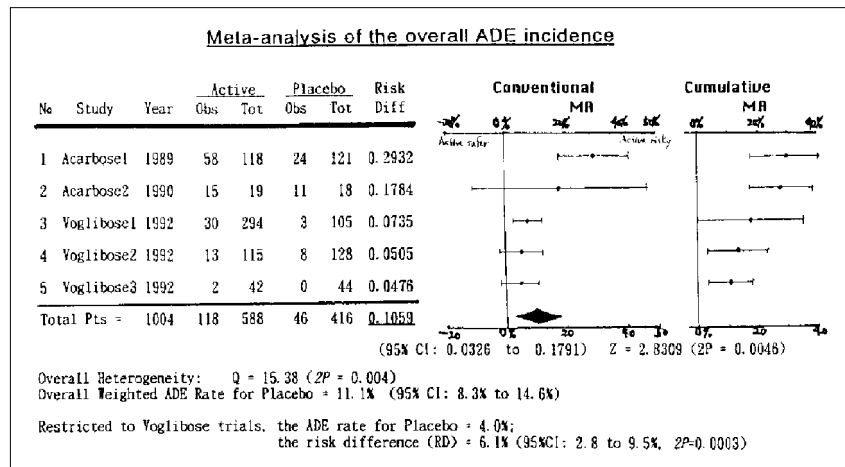


図-2

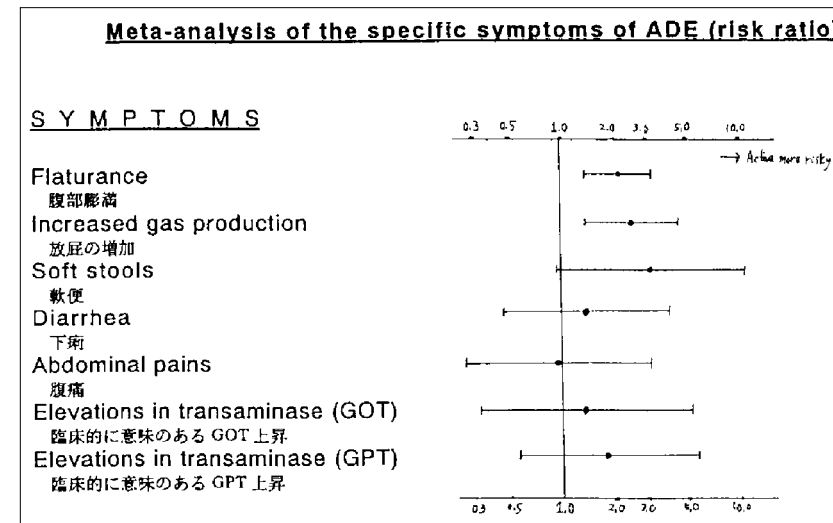


図-3

本メタアナリシス研究に用いた脳循環代謝改善薬とプラシーボのランダム化比較試験14報の一覧。

対象論文および薬剤	患者層	症例数	施設数	投与期間	出版年代
1 シンナリジン ⁹⁾	CI, CH, CA 発作後3年以内	A : 46 P : 46	12	4 weeks	1972
2 ベンシクラン, シンナリジン ⁹⁾	CI, TIA, CH, SAH, CA 発作後3カ月以上	Aベ:107, Aシ:108 P : 113	50	8 weeks	1975
3 メグロフェノキセート, 塩酸ピリチオキシン ⁷⁾	CI, CH, FIA, CA 発作後1カ月以上	Aメ:199, Aピ:205 P : 205	90	4 weeks	1975
4 酒石酸イフェンプロジル ⁹⁾	CI, TIA, CH, SAH, CA 発作後1カ月以上	A : 132 P : 134	30	8 weeks	1976
5 塩酸ピリチオキシン ⁹⁾	CI, TIA, CH, SAH, CA 発作後1カ月以上	A : 204 P : 221	55	8 weeks	1978
6 マレイン酸シネバジド ¹⁰⁾	CI, TIA, CH, SAH, CA 発作後1カ月以上	A : 163 P : 164	23	8 weeks	1979
7 GABA, 塩酸ピリチオキシン ¹¹⁾	CI, CH, SAH, CA 発作後1カ月以上	AG : 142, Aピ : 142 P : 147	50	8 weeks	1980
8 ホパテン酸カルシウム, 酒石酸イフェンプロジル ¹²⁾	CI, CH, SAH, CA 発作後1カ月以上	Aホ:118, Aイ:120 P : 117	54	8 weeks	1981
9 アマンタジン ¹³⁾	CI, CH, SAH, CA 発作後2カ月以上	A : 173 P : 170	52	8 weeks	1984
10 ニルバジピン ¹⁴⁾	CI, CH 発作後1カ月以上	A4mg : 93, A8mg : 95 P : 93	73	8 weeks	1991
11 アンラセタム ¹⁵⁾	CI, CH 発作後1カ月以上	A : 173 P : 159	118	12 weeks	1991
12 オキラセタム ¹⁶⁾	CI, CH 発作後1カ月以上	A : 139 P : 134	85	8 weeks	1991
13 塩酸ピノコナート ¹⁷⁾	CI, CH 発作後1カ月以上	A lo : 116, A hi : 119 P : 119	75	8 weeks	1992
14 ネブラセタム ¹⁸⁾	CI, CH 発作後1カ月以上	A lo : 105, A hi : 105 P : 96	69	8 weeks	1992

患者層: CI=Cerebral Infarction(脳梗塞), TIA=Transient Ischemic Affack(一過性虚血発作),
CH=Cerebral Hemorrhage(脳出血), SAH=Subarachnoid Hemorrhage(クモ膜下出血),
CA=Cerebral Arteriosclerosis(脳動脈硬化症)
症例数: A=Active(実薬), lo=low dose(低用量群), hi=high dose(高用量群), Aの後の文字は薬剤の識別に用いた。

表-4

日本の脳卒中の薬剤の臨床試験は、わたくしが95年に調べたときには105件ほどRCT、ランダムに割り付けされている臨床試験がありましたが、その中で73件は慢性期のものでした。ここで対象となるのは脳卒中の発作後1ヵ月を過ぎた慢性期の後遺症に対する薬剤の使われ方です。

実薬同士の対照試験というのは評価にバイアス(有効の方向へ)が入る可能性があるということで、実薬が対照の試験は全部除外し、プラシーボを対照とした試験に限定しております。対象となる研究を一応網羅的に検討したつもりなのですが、どなたかの発表で18研究とありましたが、わたしのほうでは14研究、論文でした。

検討はわたし個人でやったというよりも、神

経内科の医師が入って一緒に評価したほうがいいのではないかとということで、ここ(検討対象論文の検出)に関してはそのドクターに責任を持ってやってもらいました。14論文でもし違つたとすれば、今現在使われていない薬剤もあるらしくて、それを入れると論文としては18くらいあるかもしれません。しかし使われていない薬剤を評価してもしょうがないだろうということで、それをその医師は除いた可能性はあります。『脳卒中』という雑誌に結果がでました(19:308-317, 1997)ので、それを読んでいただいて批判があれば、それをアップ・デイトしてまた検討したいと思います。

エンドポイントは全般改善度を用い、中等度

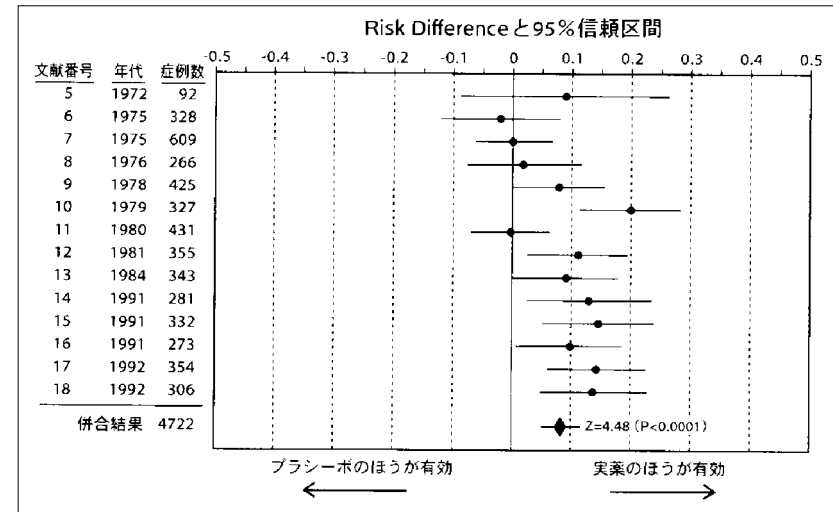


図-5

以上を改善とし、2つに分割しました。ほかの諸個別症状に関しても調べました。実薬群は合計で2800例、プラシーボ群は1900例です。数に違いがあるのは、実薬群のほうに異なる用量の試験も含んでいるためです。

14試験の古い薬剤からシンナリジン、ベンシクラン、メグロフェノキセート...と示しました(表-4)。

指標の出し方で神経内科の医師とも随分揉めたのですが、全般改善度がどのくらい効いているかということ、わたしは差の指標で出したのですが、オッズ比のほうがいいとその医師に言われまして、それで示したのが図-4です。14研究並んでおりますが、1よりも右側にあるほうがこの場合には治療薬が有効である、という数字です(編集部:通常、メタアナリシスの図では、イベント発生率でオッズ比を求めため、1よりも左側になれば介入群が有効であることを示している。この場合は、有効率のオッズ比を求めているために、1より右側の方が介入群が有効であることを示すようになっている)

古い順に並べております。最初の方は効いているものもあり、効かないものもありという状況でしたが、1980年代からの成績は良いものになっています。

全体で約1.6というオッズ比でした。臨床医は1.6倍効いたと言いたいらしいのです。欧米の論

5)。1980年以降の試験は(プラシーボよりも)10%を少し上回っているくらいの有効率を出しているのが多いかなというところ。14研究の間でここではその違いを変量に加えた変量効果モデルで分析しました。

自覚症状に関してオッズ比で表すと、全体で1.8倍有効と計算されました。神経症候に関するものは少し弱くて1.5倍くらいですが、これはバラツキが大きい傾向があります。やはり最近の1991年以降のスタディが良いのはちょっと気になると言えれば気になります。精神症候はかなり良くて2倍近くの有効率が出ておりますが、同じく結果は相当ばらついてます。ADLは1.5倍くらいです。これは試験論文の年代を問わず一定の状態がずっと続いています。

図-6は、脳梗塞の後遺症と脳出血後遺症の試験が混ざっているのが幾つかありまして、それを分けてサブ・グループで解析したものです。脳梗塞は全部で1480例、脳出血は376例しかないのですが。脳梗塞のほうで見ますと、オッズ比で1.6倍くらい。先ほどと同じような数字です。ただしこれも見て分かりますように、1991年以降4試験ありましてこれはいいですね。古い試験はよくないというデータが出ております。これがどういう意味を持つのか、論文を読んだだけでは分からないです。

脳出血の方は結果がばらついております。た

文でも、オッズ比が1より大きいときに有効と言っているものを見かけますが、ふつうはイベント発生などの危険性のときにオッズ比を使うようです。

改善率の差の方がいいのではないかと検討し直したのですが、そうすると平均で約9%実薬の方がプラシーボよりも優れているという成績が出ました(図-

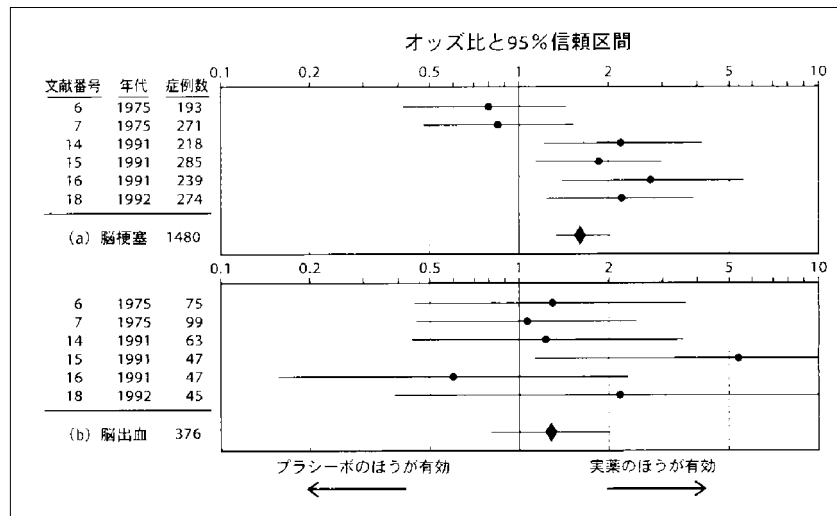


図-6

だし脳梗塞に比べると、脳出血の有効率が低いことは、これを見る限り間違いないのではないかと思います。

こういうふうに4000例くらいのデータで総合的にメタアナリシスという手法で脳循環・代謝改善剤の効果を再検討してみたわけですが、議論のポイントは3つほどあります。

1番目は開発が中止された同種の薬剤は含んでおりません。つまり途中で消えていったものは日本の臨床試験の特徴ですが、論文に発表されない。発表されないけれども同種同効の薬剤はあったはずですが、それが発表されていないということはネガティブの可能性が強いということで、そのパブリケーション・バイアスがあるだろう、あるかもしれないというのがポイントの第1です。

2番目のポイントは、使用されたエンドポイントは一貫しております。一貫という表現は適切ではないかもしれませんが、CGIといってクリニカル・グローバル・インプレッション、外国に説明するときはそう説明していて、脳卒中を扱う神経内科の医者は分かるんですが、それはいいんですが、その信頼性・妥当性は分からない。今、それが検討されている状況だと思います。

3番目のポイント。これが一番大事かと思いますが、最初に示したように、スタディの期間はほとんどが8週間なのですね。長いのも12週間

死亡例を増やさないというところの評価は残念ながら行われていない。それにも関わらず、再評価で8週間というスタディをいろんなところでまたやっているようなのですが、どうせ8週間のスタディをするくらいなら長期のスタディをきちりやってもらいたいと思いますつまり、長期効果については何とも言えないということです。

というように、2ヵ月から3ヵ月間をフォローしてその間に自覚症状とか精神・神経症状とかが良くなったかどうかを見ているわけですが、実は、脳循環・代謝改善剤というのは、後遺症の患者さんに結構長期で使っている。

長期で使ってどこがいいのかということ、例えば将来的に再発作を起こさないとか、

■ 質疑／討論 ■

別府：時間がなくなりかけていますが、質問を。

柳：（医師、内科）アカルボースに関して。エンドポイントが血糖値であるべきかどうかは問題のあるところですが、それはおいといて、プラシーボを対照にした試験のメタアナリシスだということですが、そういう試験しかなかったのでしょうか、プラシーボを対照にしているということをお考えですか。私は納得できないのです。実際に血糖を下げる薬としては効果の確認されているものが存在するわけです。それよりも優るといふ証明をしなければ、実際の臨床的価値はないのではないかというのがわたしの直観ですが。

折笠：従来のSU剤については、ふつうは併用禁止として（血糖降下がエンドポイントなら）、食事療法と運動療法はすべての人に与えておいてするのは。

柳：それがプロトコールとして確認されているかどうかです。試験計画の中に組み込まれていたかどうかを問題にしているのです。プラシーボを投与するというだけでは、議論の対象になっていないということです。本当に両群にきちんとした食事療法がなされていてプラシーボだけが違うのであれば同意します。

折笠：それは患者背景を見ていただければ分かると思います。

柳：そういうプロトコールでアナリシスしないと本当の治療効果は言えないのではないかと思うのですが。

折笠：それは臨床試験をやるときに、ランダムに割り当てておけば、ある程度の症例数が集まって来ると似たような内容になる。

柳：そうするとですね、標準治療を受けているかどうかということについては無関係だということでしょうか。患者さんに対して食

事指導をきちんとしているかどうかは議論の対象ではないとお考えですか。

折笠：それは重要なことです。食事治療とか、教育的なことは非常に重要なことですが、そこは両群でばらついており、その違いが結果に影響を及ぼしたとは考えにくい。RCTだからです。

柳：両群が均等に受けているということ。それが前提とならないでプラシーボがあるからそれでいいと言っているのでしょうか。

折笠：それは倫理的な問題ですか。

柳：倫理的ではなくて実証的な問題としてです。

折笠：デザインとしては、そういうベーシックな治療をするのは当然のことです。

柳：それがきちんと記載されているかということですか。

折笠：それは記載されています。食事療法、運動療法はすべての人に行うのが前提となっておりましたが、確かに行われていたかについては、誌上報告されていませんでした。問題ないとは言えないと思います。

近藤：（医師、放射線科）日本のいろいろなランダムイズド・トライアルというのは質が劣っているということが言われているし、封筒法をやっても自分の思うのが出てくるまで破くとか、そういう状況で、こういうメタアナリシスすることに矛盾を感じないのですか。もう一つ、全般改善度という摩訶不思議なものを数値化することは、ほとんど意味がないような気がするのですが、この2点をお聞きしたい。

折笠：いい加減な試験と一言で言ってしまうのはどうかと思うのですが、いろんな問題のあるところというのは、日本の試験だけでなく各国でも指摘されています。ただ、個々の試験をポイント、ポイントとして出すだけで、それがどうなっているのかは見えないと思う。それをつなげて、現状の成績としては

どの程度のものなのかと、見るための一つの道具なのではないかと思えます。

メタアナリシスをしたところ結果がおかしいから、何か違う操作をしていたのではないかとということも見ることは出来るわけです。個々の試験をレビューするだけで、この試験は悪い、あの試験は...も出来ますけれども、総合的に見ることも大事ではないかと。

近藤：こういうメタアナリシスをやっていると、留保を付けても、「結果がよかった」と発表していくことが、世の中の誤解を招いていく、一般の人を惑わしていくことにつながるのではないかと問題意識があるのです。

別府：司会者の独断ですが、時間がないので、質問者はマイクの前にいらっしゃる3人に限らせていただきます。これ以上増やさないでください。発言はお一人1分にしてください。折笠さんには申し訳ありませんが、3人の質問をメモしていただいて後でまとめて返事していただくということにさせていただきます。

橋本：(医師、内科)まず、全般改善度がまったく信用できない。と言うのは、個別評価項目で104項目調べて、その内有意差があったのはゼロだった。ところが結論としては全般改善度で有意差があったと。そういう臨床試験のデータを集めて解析するのはどういう意味があるのか。症例数についても各個別項目は、例えばセラクロールは全例で259の症例を集めているのですが、ところがめまいを訴えたのはわずか33例、わずか13%。これは統計的には脱落と考えるべきだと思うのですが、そういう不思議なスタディのやり方と思うのですが、どうお考えですか。

革島：(医師、外科)メタアナリシスをやることに関しては何らかの蓄積されたデータを利用するという意味でどんどんやられたらいいと思うのですが、個々の「RCT」がベストであるという主張に対しては疑問があります。

と言いますのは、インフォームド・コンセントとの関わりになるのですが、先ほど大橋さんがおっしゃっていた日本でもやるというタモキシフェンの臨床試験ですが、これを日本でやるとして、2年間の使用の後それからの2年間あなたはランダムにプラシーボになるかもしねません、でも現在までの感触では合計4年やったほうがいいようですよ、と患者さんに話をして納得する人がいるのかどうか。もっと切実な問題は、がんの臨床試験でランダムに割り付けてから患者さんに、新しい治療法だけど受けてくれませんか、というふうにインフォームド・コンセントをする割付後承諾という方法がありますが、その場合、断られたケースもITT(インテンション・トゥ・トリート)で解析しなければならない。ということは脱落例になるわけです。脱落例がたくさん入るということは、「RCT」の信頼性が落ちてしまって、ほとんど無意味になるのではないかと考えるのですが、その点について何か方策をお持ちかどうか。

名郷：(医師)先ほどの柳ドクターのコメントに対するコメントです。ダブル・ブラインドのトライアルを(折笠氏は解析対象)選んだということなので、ダブル・ブラインドが守られる限り、全体としての治療は同じであると考えていいと思うのです。ただ、日本のトライアルはそのダブル・ブラインドが守られないから問題であるというちょっと次元の違いだったのではないかと思います。

アカルボースについては、追跡率とかフォローアップ期間による感度分析が必要ではないか。アカルボースの治療論文を見ますと、フォローアップ期間によって実は全然効果が違って、3ヵ月だとかなり大きな差があるのですが、6ヵ月だと非常に小さな差しかないという論文がかなりあったので、そういう点が問題ではないかと考えます。

折笠：最初の近藤さんのご質問の、こうい

うものが社会にインパクトを与えるのではないかと、という点のご指摘の通りだと思います。慎重にやらないといけません。

第2番目の、橋本さんの質問ですが、個別症状ではいいところがないのに総合評価手は有意といいますが、個別でもいいところは少しあります。それらの症状が総合評価へ強く影響したものと想像します。これについては、総合評価の構造を解明する必要はあるかと思えます。こういう主観的データでしか評価できない領域はあるかと思えますので、こういうアプローチで進めたわけです。

革島さんのご質問ですが、確かにそうだと思います。臨床試験におけるインフォームド・コンセントは現在もなかなか進まないという現状があるのですが、プラシーボを使うと言うと、患者さんたちは何か薬(にせぐすり)を飲まされて、普通の治療をやってくれないのではないかとというふうに思われますが、ほとんどのスタディというのは標準治療というものがベースにあって、それだけで本当はいいんだけど、それに上乗せして例えば薬を使っているのかどうか、というスタディがあって、プラシーボ・コントロール・スタディなわけです。標準治療がない場合を除けば、標準治療はだれに使ってもいいという前提なので、そこを患者さんにちゃんと説明されていないのではないかと思います。

追跡率で グルコシダーゼの有効率が違うというのは確かにその通りだと思います。それは今までの市販前のデータだけなので、長期でフォローされていないデータを解析に使っているの、そこまでできていませんが、外国のデータではできると思えます。

橋本：私の二つ目の質問に答えていただけていませんが。

折笠：259例中、めまいが33例しかベースラインでなかったのはしょうがないと思います。すべての症状をすべての脳卒中後遺症の

患者が有しているわけではありません。このように33例が対象になってしまうため個別症状では有意にならないわけでしょう。

別府：整理いたします。いったんジェネライズした後に層別化して細かく分けたものの中には非常に少数例しかないものがある。それには信憑性がないと橋本さんはおっしゃるんですね。それはまさにその通りだと思うので、しかしながら折笠さんが解析しているのは全般改善度を中心にやっていらっしゃるわけで、その全般改善度自体に非常に問題があるということはみなさんご指摘になっているわけです。そのことに関しては、わたし自身も聞きながら、元ネタを提供する臨床医の質の良さ、悪さというものが関係してくると思いました。折笠さんは、統計学者として非常にきちっと臨床試験論文を見てくださったのだけでも、実は相手方のことが見えていらっやらない、という気が少しするのです。

それに関しては、やはり非常に慎重に、近藤さんがおっしゃったように、Evidence-Based Medicine自体が逆に、やってもそれをただ宣伝に使うんじゃないかという危惧がある。

Evidence-Based Medicineやメタアナリシス自体を批判する人々がいる。しかし、きちんとしたシステムティック・レビューは、それ自体きわめてオリジナリティーの高い仕事だと思うのです。どのトライアルを採用し、どれを落としたり、ちゃんと書かなくてはならないようになっているわけです。そうすると下手なシステムティック・レビューをやると、その人の研究者としての質を問われることになると思うのです。そこがコクランのレビューのいいところだと思います。

決してエビデンスという言葉に幻想を抱いてはいけません。エビデンスは細心の注意を払って、非常にクリティカルにやらなければ得られないものだと思います。

司会としてまとめましたが、折笠さん、何

か追加してコメントいただくことは。

折笠：大変クリティカルな問題で、わたしも解析をやってて、医師が入ってるのでそこは気になっていたのですが、誤解を招かないように書いたほうが良いとは申したのですが、もしかすると変なふうに使われるかもしれないと心配はしております。一人歩きして、脳循環・代謝改善剤がいいんだ、というふうになると困るのですが、わたしは普通のやり方に従って解析をしてみるとこういう結果にな

ってしまった、と。しかしながら、それぞれのところでは問題点もありますので、それは考察にちゃんと書いていますので、それを読んでいただくとよいと思います。

別府：確かにエンドポイントの問題とか、折笠さんにご指摘になったと思います。今後、ちゃんとしたエビデンスを作っていくのだという共通の認識を確認することで、話をまとめさせていただきます。

抗痴呆剤：E2020の Protokol作成を例に

柳 元和



内科医師

医療問題研究会会員
コクラン共同計画レビュー担当
TIP 誌編集委員
薬害オンブズパーソンメンバー

患者の緊張をほぐす

コクラン共同計画の勉強で行ったラドクリフ・インファーマリーでの様子を少し紹介します。写真1は、レントゲンの待合室です。壁にペインティングを専門家にしてもらって非常にきれいです。待合室で待つ患者さんの気持ちを少しでも和らげるという配慮からなされています。患者さんに対してどのようなケアをしていくかということは、このように多くの人手と配慮とに裏打ちされています。小児科病棟のバスルームにも写真-2のようにペインティングがなされて、子どもの緊張をほぐす配慮がなされています。人員配置としても、看護婦さんの数が日本のだいたい3倍くらいです。そういう非常に手厚い看護の元で、ここは臨床試験を積極的に勧めていく病院としてNHS (National Health Service) から指定されています。



写真-1

社会の資源としての高齢者

あるいは、老人の方々、高齢者が非常にアクティブに働いておられます。そうして病院をもり立てている、ボランティアで支えておられます。そういう老人を社会の資本として非常に大事にしていく。そういう背景があって、抗痴呆剤の臨床試験が積極的に行われています。治験の評価マニュアルなどは、電話帳くらいあります。それを使って問診をしていく。患者さんに何分かかわからないくらい、じっくりと問診をやっていく。そういう臨床試験をしています。



写真-2