
本誌の評価表・評価法(総論)

ーアトピー性皮膚炎治療のシステマティック・レビュー作成に関してー

幸野 健¹⁾、羽白 誠²⁾

1) 市立吹田市民病院皮膚科、2) 大阪警察病院皮膚科

1. システマティック・レビューとは

臨床家の行動変革の指針として開発されたEvidence-based medicine (EBM) だが、現在では臨床医学実践の世界標準となっている^{1,2)}。

EBMの創始者SackettらはEBMの行動目標を5つのステップに分けた^{2,3)}。すなわち、

- 1) 問題点の抽出
- 2) 情報(文献)の収集
- 3) 批判的吟味critical appraisal
- 4) 結果の患者への適用
- 5) 以上のステップの評価

である。

多忙な日常診療において、個人の医師が2)、3)のステップを行うことは大変な労力を要するため、一般的にはレビュー・グループにより行われる。この作業及び結果をシステマティック・レビュー(systematic review, 以下SR)という⁴⁾。予め基準を設定した上で情報収集と批判的吟味を効率的に行い、最終的に一定の結論をもって要約する。SRにより「現在までに何が分かっており、何が分かっていないか、現在・未来において何をなすべきか」が明確になる。

皮膚科学における最大のEBM推進者、英国Nottingham大学のWilliamsらは英国医師会雑誌編集部のEBM2次資料“Clinical Evidence”に協力し、アトピー性皮膚炎に関するSRを発表している⁵⁾。また彼らは英国国民保健サービス(National Health Service, NHS)の医療技術評価プログラムの一環として予備的SRを公表している⁶⁾。

WilliamsらのSRも現時点では中間報告的な未完成なものであり、各研究のエビデンスとしての質のレベル評価は行われていない。また言語的な問題から本邦の情報がほとんど収集されていない。今回、我々はさらに新しい文献を徹底的に検索・収集し、質を評価し、本邦の実情に合わせたSR作成を行った。

2. システマティック・レビュー作成過程

SR作成過程は国際的なEBM支援組織コクラン共同計画により標準化されている(表1)⁷⁾。また、完成したSRはQUOROM(Quality of Reporting of Meta-analysis)声明により示された標準により評価される(表2)⁸⁾。今回のSR作成過程、即ち、対象論文の基準、検索方法、研究の質の評価、解析の要約においては、これらの国際標準を可及的に満たすよう配慮した。

A. 対象論文の基準

ランダム化比較試験(randomized controlled trial, 以下RCT)、対照のある非RCT試験(比較臨床試験、clinical controlled trial, 以下CCT)、コホート試験、症例対照試験(case control study)、症例集積研究を可及的に収集することとした。

B. 情報(文献)の検索と収集

PubMed、医学中央雑誌などのデータベースから電子サーチを行った。検索方法、検索式はできるだけ詳細に記載することとした。さらに、関連雑誌を可及的に手作業検索した。

C. エビデンスの質の評価

エビデンスの質の評価に関してはいくつかの分類表が考案されて来た(表3-5)⁹⁾。これらは簡明であるが、

表1. コクラン共同計画におけるシステマティック・レビュー作成過程 (文献7より)

	プロトコル	システマティック・レビュー	考慮すべき問題
1	タイトル	同左	
2		構造化要約	
3	背景	同左	病態生理, 疫学, 疾患の自然歴, 治療と経費
4	目的	同左	
5	対象論文の基準	同左	研究デザイン, 研究対象者, 介入, アウトカム評価
6	論文の検索方法	同左	データベース, 検索式, 未発表研究・学会抄録・副作用報告の扱い
7	レビューの方法	同左	論文翻訳について, 論文の選択, 試験方法の質の評価法, データ抽出と総合解析法など
8		情報収集の経過	
9		研究の質	ランダム化と選択バイアス, アウトカム評価の盲検化と測定バイアス, 脱落者の扱いと選択バイアス, 解析の単位, 試験期間
10		結果	
11		解析の要約	メタアナリシスに選ばれた研究の表と結果のグラフ
12		補足的データ	
13		考察	試験参加者と治療の要因, 患者集団の特徴, 統計学的な異質性, 出版バイアスと試験サイズ
14		結論	当該介入の意義, 今後研究すべき領域
15	謝辞	同左	外部レフェリー, 編集協力者, 翻訳者の明確化(コクラン共同計画では患者代表も加えることを推奨)
16	利益相反	同左	
17		対象となった論文の表	それぞれの特徴を記載
18		除外された論文のリスト	同上
19		進行中の臨床試験	同上
20	参考文献	同左	論文の書誌事項(対象とされた論文と除外された論文)
21		レビューアーと協力者名	
22		プロトコル作成日時とSR各部の改定日時	
23		キーワード	

表2. QUOROM声明のシステマティック・レビュー・チェックリスト(要約) (文献8より)

	章	トピック	記述項目
1	タイトル		「システマティック・レビュー」または「メタアナリシス」という用語を含む
2	抄録		構造化抄録を使用
3		目的	臨床的疑問点を明示的に記載
4		データ源	データベースのリストと他の情報源
5		批判的吟味の方法	選択基準(対象集団, 介入, アウトカム, 研究デザイン), 妥当性評価, データの要約, 研究の特性の記載. 定量的データ統合(定量的メタアナリシス)の方法
6		結果	除外されたRCTの特性. 点推定値と信頼区間. サブグループ解析
7		結論	主な結果
8	緒言		明示的な臨床的疑問. 介入の生物学的合理性, レビューの合理性
9	方法	文献の検索方法	情報源の詳細(データベース, 何らかの登録, 個人のファイル, 専門家からの情報か? 研究機関, ハンドサーチの詳細)
10		選択基準	批判的吟味への組み入れと除外の基準(対象集団, 介入, 主要アウトカム, 研究デザイン)
11		妥当性評価	使用された基準と評価過程(盲検化, 質の評価, その結果)
12		データの要約	使用された評価過程(独立して行ったのか, 二重に行ったのか?)
13		研究の特性	研究デザイン, 被験者の特性, 介入の詳細, 転帰の定義など. 臨床的異質性の評価は?
14		定量的データ統合(定量的メタアナリシス)	主な効果指標(リスク比等), 結果の統合法(統計学的検定と信頼区間), 欠測値の処理. 統計学的異質性の評価, 感度解析と層別解析の事前設定の根拠, 出版バイアスの評価.
15	結果	試験のフローチャート	収集された試験の吟味過程での流れ(フローチャートで図示)
16		研究の特性	それぞれの試験の特性を記載(年齢, 症例数, 介入, 内容, 用量, 介入期間, 追跡期間など)
17		定量的データ統合(定量的メタアナリシス)	試験の選択や妥当性評価の一致度. 単純な要約の結果(各試験での各群, 各主要な転帰について), ITT解析に用いられる効果サイズと信頼区間の計算に必要なデータ(2x2分割表, 平均と標準偏差, 割合等)
18	考察		主な知見の要約. 内的・外的妥当性に基づく臨床的推論. 入手可能なエビデンス全体の中での結果の解釈. レビュー過程において可能性のあるバイアス(出版バイアス等). 将来の研究テーマの提案

治療、予後、検査等の領域に応じて研究デザインが異なるにも関わらず、一律に評価しているという問題点がある。そこで、現在発表されている分類表として最も詳細かつ厳密なOxford大学EBMセンター分類を参考にすることとした(表6)⁹⁾。しかし、複雑に過ぎるため、何度も検討を加え、その要約版を考案した(表7)。

エビデンスの質分類には、いまだ専門家の間でも論議がある。Williamsらも、そのシステムティック・レビュー^{5,6)}においてエビデンスの質のレベルを評価することを躊躇しているが、我々はあえて今回のレビュー事業においてエビデンスの質評価を行うこととした。ただし、これは現時点での、あくまでも暫定的な評価であり、今後の世界動向により評価が変化することもあり得る。

表3. 米国予防医療サービス特別研究班の分類(1991)

I	RCT
II-1	非RCT
II-2	コホート研究または症例対照研究
II-3	時系列研究,非対照実験研究
III	権威者の意見,記述疫学

表4. 米国保健政策研究局(旧AHCPH,現AHRQ)の分類(1993)

Ia	複数のRCTのメタアナリシス
Ib	少なくとも一つのRCTによる
IIa	少なくとも一つのよくデザインされた非RCTによる
IIb	少なくとも一つの他のタイプのよくデザインされた準実験的研究による
III	比較研究,相関研究,症例対照研究など,よくデザインされた非実験的記述的研究による
IV	専門家委員会の報告や意見,あるいは権威者の臨床経験

表5. Gray(Oxford大)の分類(1997)

A	I	明瞭な結果を持つ大規模RCT
B-1	II	不確実な結果を持つ小規模RCT
B-2	III	同時観察対照群のある非RCT
C-1	IV	記録対照群のある非RCT
C-2	V	対照群なく症例群のみのある試験研究

D. 評価フォームの作成

基本的にはWilliamsらのSR^{5,6)}の表を参考にしたが、上述のコクラン共同計画のSR作成過程表(表1)⁷⁾、QUOROM声明のチェックリスト(表2)⁸⁾、さらにRCT論文評価の国際標準となっているCONSORT(Consolidated Standards of Reporting Trials)声明(表8)¹⁰⁾の各項目を可及的に満足する評価フォームを作成した(表9)。

E. 批判的吟味

前記評価フォームに則って、各文献を評価し、一定の結論を出した。Williamsらにより指摘されている以下の問題について、特に検討した^{5,6)}。

- 1) RCT自体が少ない
- 2) 脱落例が多い
- 3) 現代臨床疫学では常識となっている、割り付けのままの(脱落例込みの)解析(治療企図解析、intention to treat analysis, ITT)が行われているものが少ない
- 4) 盲検化(マスク化)、ランダム化したと記載されていても、その方法が示されていない
- 5) データの実数や信頼区間が示されていないものが多く統計学的評価が困難
- 6) アウトカム評価項目が一定しておらず結論の導出が困難

F. 左右比較試験(left-right comparison)の問題

皮膚科領域には同一個体内比較(intra-individual comparison)あるいは左右比較を行った論文が多い。これらのエビデンスの問題点に対し、しばしば批判が投げかけられて来た。Williams教授との対談により、今回はこれらの試験も通常のRCT、あるいは対照試験と同様に扱うこととした(付録参照)。

表6. Oxford大学EBMセンターの分類(文献9より)

勸告度	レベル	治療:その治療は効能があるか(efficacious)/効果があるか(effective)/有害か(harmful)	治療:その薬剤は同クラスの他の薬剤より優れているか	予後	診断	鑑別診断/有病率研究	経済学的分析と決断分析
A	1a	RCTのSR(同質性*あり)	直接比較型RCTのSR(同質性*あり)	前向きコホート研究のSR(同質性*あり);複数の集団で検証された臨床判断規則†	レベル1の診断研究のSR(同質性*あり);複数の臨床施設からの1b研究による臨床判断規則†	前向きコホート研究のSR(同質性*あり)	レベル1の経済学的研究のSR(同質性*あり)
	1b	個別のRCT(信頼区間の狭いもの‡)	臨床的に重要なアウトカムのある一つの直接比較型RCT	個別の前向きコホート研究で80%を超える経過観察が行われたもの;1つの集団で検証された臨床判断規則†	良い†††参照基準のある検証のための**コホート研究;1つの臨床施設で検証された臨床判断規則†	良い経過観察****が行われた前向きコホート研究	臨床的な目的に適った費用あるいは代替物に基づく分析;エビデンスのSR;多元的感度分析を含んでいる研究
	1c	全てか無か(all or none)§		全てか無かの症例集積研究	絶対的SpPins及びSnNouts††	全てか無かの症例集積研究	絶対的により良い価値,或いはより低い価値の分析**
B	2a	コホート研究(同質性*あり)のSR	検証された代理アウトカム**を持つ直接比較型RCT	後向きコホート研究またはRCTにおける無治療対照群のSR(同質性*あり)	レベル2以上の診断研究のSR(同質性*あり)	レベル2b及びそれ以上の研究のSR(同質性*あり)	レベル2以上の経済学的研究のSR(同質性*あり)
	2b	個別のコホート研究(低質なRCTを含む;例えば経過観察例が80%未満)	同一もしくは別の患者に対して行われた,異なる薬剤対プラセボの複数のRCT(臨床的に重要なアウトカムや検証された代理アウトカムを持つもの)の比較	後向きコホート研究またはRCTにおける無治療対照群の経過観察;split-sample analysis§§§でのみ誘導されたか検証された臨床判断規則†	良い†††参照基準のある探索的**コホート研究;split-sample analysis§§§またはデータベースのみから誘導,検証された臨床判断規則†	後向きコホート研究,或いは経過観察の貧弱なコホート研究	臨床的な目的に適った費用あるいは代替選択肢に基づく分析;エビデンスの限定的レビューか単一症例研究;多元的感度分析を含んでいる研究
	2c	アウトカム研究;生態学的研究		アウトカム研究		生態学的研究	監査 ¹⁴⁾ あるいはアウトカム研究 ³⁾
	3a	症例対照研究(同質性*あり)のSR	レベル2bの群のサブグループ分析		レベル3b或いはそれ以上の研究のSR(同質性*あり)	レベル3b及びそれ以上の研究のSR(同質性*あり)	レベル3b及びそれ以上の研究のSR(同質性*あり)
	3b	個別の症例対照研究	同一もしくは別の患者に対して行われた,異なる薬剤対プラセボの複数のRCT(代理アウトカムが検証されていないもの)の比較		非連続的研究;或いは参照基準が一貫して適用されていない研究	非連続的コホート研究,あるいは対象が非常に限定されたコホート研究	限定された選択肢,費用,低質なデータ推定値に基づくが,臨床的な目的に適った変動を取り入れた感度分析を含む解析
C	4	症例集積研究(及び低質なコホート研究あるいは症例対照研究§§)	臨床的に重要なアウトカムを持つ非RCT(観察的研究や行政的データベース研究)の比較	症例集積研究(及び低質な前向きコホート研究)***	症例対照研究,参照基準が貧弱か独立していないもの	症例集積研究,あるいは参照基準が入れ替えられた研究	感度分析が行われていない解析
D	5	明確な批判的吟味が行われていないか,生理学・実験室での研究や”確固たる信念”に基づいている専門家の意見	明確な批判的吟味が行われていないか,生理学・実験室での研究や”確固たる信念”,代理アウトカムが検証されていない非RCTに基づいている専門家の意見	明確な批判的吟味が行われていないか,生理学・実験室での研究や”確固たる信念”に基づいている専門家の意見	明確な批判的吟味が行われていないか,生理学・実験室での研究や”確固たる信念”に基づいている専門家の意見	明確な批判的吟味が行われていないか,生理学・実験室での研究や”確固たる信念”に基づいている専門家の意見	明確な批判的吟味が行われていないか,生理学・実験室での研究や”確固たる信念”に基づいている専門家の意見

[コメント]

1. 英国国民保健サービス (National Health Service) のEBM研究開発(Research & Development: R & D)センターのメンバーにより作成 (Bob Philips, Chris Ball, Dave Sackett, Brian Haynes, Sharon Straus, Finlay McAlister)。
2. 以下の理由により結論を出せない場合、レベル表示にマイナス記号“-”を付けること。即ち、
 - ・信頼区間が広い結果が得られた時 [例えば、RCTにおける絶対リスク減少 (absolute risk reduction: ARR) は統計学的に有意ではないが、信頼区間からは、臨床的に重要な便益や害を除外できない場合]、
 - ・システマティック・レビューに問題のある (しかも統計学的に有意な) 異質性が認められた時。
3. 勧告のグレードはエビデンスのレベルと直接にリンクさせて示している。しかし、エビデンスのレベルは研究の妥当性を言及しているものであり、その臨床的適用性を示すものではない。勧告のグレードを決定する前に、その他の要因 (例えば、経費、実行のし易さ、疾患の重要性) を考慮する必要がある。グレードは、現在、エビデンスの妥当性に密接に関連付けて示されているが、これらは、指針や関連するCATs (critically appraised topics: 批判的吟味されたトピックス) において強調される関心を反映して変更される可能性がある。

*	同質性homogeneityとは、それぞれの研究結果が、その方向と程度において問題となる程のパラッキ (異質性) がないこと。統計学的に証明された異質性が常に問題となるものではなく、逆に、問題となる異質性が常に総計学的に証明されるものでもない。問題となる異質性がある場合は、評価の後に“-”を付けること。
†	臨床判断規則Clinical decision rule (これらは予後判定或いは診断カテゴリーを導くアルゴリズムまたはスコア・システムである)
‡	広い信頼区間を持つ臨床試験、あるいはその他の研究をどのように理解し、段階付けし、用いるかについてのアドバイスに関してはコメント2を参照。
§	その治療法が得られる前には全ての患者 (all) が死亡していたが、現在はその治療法で何人かが生存している場合、及び、その治療法が得られる前は何人かの患者が死亡していたが、現在はその治療法で誰も死亡していない (none) 場合。
§§	低質なコホート研究とは、次のいずれかまたはその組み合わせ： 比較する群の定義が明確でないもの； 曝露群と非曝露群において、曝露と転帰の測定が同様に (可能ならば盲検的に)、客観的に行われていないもの； 既知の交絡因子について、その同定や調整が適切に行われていないもの； 患者を十分な長さの期間、完全に経過観察していないもの。 低質な症例対照研究とは、次のいずれかまたはその組み合わせ： 比較する群の定義が明確でないもの； 症例群と対照群において、曝露と非曝露の測定が同様に (可能ならば盲検的に)、客観的に行われていないもの； 既知の交絡因子について、その同定や調整が適切に行われていないもの。
§§§	Split-sample analysisによる検証は、サンプル全体で全ての情報が収集され、人工的にこれを誘導群と検証群に分割することで行われる。
††	絶対的SpPinsとは診断法の結果が陽性であれば、診断を確定 (rule-in) できるほど特異度が高いこと。 絶対的SnNoutsとは診断法の結果が陰性であれば、診断を除外 (rule-out) できるほど感度が高いこと。
‡‡	より良い価値の治療とは、同じくらい良くて安い、または同じ或いはより低い費用でより良い治療。 より低い価値の治療とは、同じくらい良いがより高い、或いはより悪くて、費用は同じかより高い治療。
†††	良い参照基準とは、対象となる診断法とは独立して、盲検的あるいは客観的に、全患者に適用されるものである。 貧弱な参照基準とは、その診断法とは独立しているが、でたために適用されるものである。独立していない参照基準を使用すれば (その「診断法」が「参照」に含まれるか、「診断法を用いること」が「参照」に影響を与える場合)、レベル4の研究ということになる。
**	検証的研究とは、以前に得られたエビデンスに基づき、ある特異的な診断法の質を研究するものである。探索的研究とは、情報を収集し、どの因子が有意か見出すため、データを底引き網漁業のように探索する (つまり回帰分析を行う) ものである。
***	低質な前向きコホート研究とは： サンプリングにおいて、標的とする転帰を既に持っている患者に都合の良い方向へバイアスが入っていたり、 転帰の測定が対象患者の80%未満でしか行われていなかったり、盲検化していないなど、非客観的な方法で行われていたり、交絡因子の調整が行われていない研究のこと。
****	鑑別診断における良い経過観察とは： 別の診断が成されるまでの十分な時間 (即ち、急性疾患では1-6か月、慢性疾患では1-5年) の経過観察が80%以上行われていること。
‡‡‡	Surrogate outcome (代理アウトカム、代理帰結) と臨床的に重要な転帰との関係が長期・複数のRCTにより確立された時に、代理転帰は検証されたと考えられる。

3. 評価フォームの読み方

今回、本邦皮膚科において初めてシステマティック・レビューが作成されたことには大きな意義があると考ええる。種々の問題点が残っていることを承知の上でエビデンスの質のレベル評価を行っていることを理解して頂きたい。今後、これらは新しい論文の追加に伴い更新されて行くであろう。

臨床疫学的にRCTが最も妥当性が高いのだが、RCTが行い難い臨床分野があるのも当然であり、そのような場合、次善のエビデンスから順次適用して行くべきである。

EBMでは、エビデンス (要素1) だけでなく、医師としての技能 (要素2)、患者の価値観・好み (要素3)、保健資源 (要素4) といった要素を統合することが目指されている (図1)。この点をよく理解した上で、評価フォームに記載された治療を適正に運用して頂きたい。

表7. Oxford大学EBMセンター分類の要約版

(RCT: randomized controlled trial, ランダム化比較試験, SR: systematic review)

勧告のグレード	エビデンスのレベル	I 介入研究(治療研究)	II 予後/診断/有病率研究
A	1	RCTのSR 高質なRCT(脱落例20%未満, 真のアウトカム [#] を測定したもの)	前向きコホート研究のSR 個々の前向きコホート研究, 1つまたは複数の集団で検証された臨床判断規則 [†]
B	2	以下のSR 個々のコホート研究, 質の劣るRCT [§] (脱落例20%以上や代理アウトカム [#] を用いているもの), 同時対照群のある非RCT	以下のSR 後向きコホート研究, RCTにおける無治療対照群の経過観察
	3	以下のSR 個々の症例対照研究, RCT内のサブグループ解析, 記録対照群のある非RCT	以下のSR 非連続的コホート研究, 対象が非常に限定されたコホート研究
C	4	症例集積研究, 対照群のない研究, 質の劣るコホート・症例対照研究 [§]	症例対照研究, 症例集積研究, 質の劣るコホート研究 [§]
D	5	明確な批判的吟味が行われていないか, 基礎研究や信念に基づく専門家の意見	明確な批判的吟味が行われていないか, 基礎研究や信念に基づく専門家の意見

訳注:上記の表がI,II列に分けられる理由は,本質的に介入試験の行い難い研究分野が存在するからである。

説明

#	真のアウトカム(転帰)とは,重症度,癌判定など,代理アウトカムとは,IgE (RIST,RAST),LDH,各種サイトカインやsIL-2R,好酸球数などの測定
†	臨床判断規則Clinical decision rule(これらは予後判定或いは診断カテゴリーを導くアルゴリズムまたはスコア・システム) 訳注:gold standardとみなされているスコア,分類表やよく検証された各種ガイドラインもこの中に入ると考えられる。
§	質の劣る研究とは,次のいずれかまたはその組み合わせ:比較する群の定義が明確でないもの,曝露群と非曝露群において,曝露と転帰の測定が同様に(可能ならば盲検的に),客観的に行われていないもの,既知の交絡因子について,その同定や調整が適切に行われていないもの,患者を十分な長さの期間,完全に経過観察していないもの,サンプリングにおいて,標的とするアウトカムを既に持っている患者に都合の良い方向へバイアスが入っているもの,アウトカムの測定が対象患者の80%未満でしか行われていないもの。

表8. CONSORT声明のRCT論文チェックリスト(要約)(文献10より)

章	トピック	記述項目	
1	タイトル・抄録	「ランダム化」されたものであることを表しているか?	
2	緒言	背景	科学的背景と合理的根拠の説明は?
3	方法	参加者	参加者の適格条件とデータが収集された設定とその場所は?
4		介入	介入の詳細と実施期間は?
5		目的	特定の目的と仮説は?
6		転帰(アウトカム)	明確に定義された主要・副次的アウトカムの評価項目は?
7		症例数	目標症例数が決定された理由,中間解析と中止基準の説明は?
8		ランダム化,割付順番の作成	作成方法は?
9		ランダム化割付の隠蔽化	割付が最後まで隠蔽されたか?
10		ランダム化の具体的実施法	割付順序作成,参加者組み入れ,参加者の各群割付を誰がしたか?
11		盲検化(マスク化)	参加者,試験実施者,評価者に対し盲検化されていたか?
12		統計学的方法	手法は?
13	結果	参加者の流れ	各群ごとにランダム割付された人数,目的とする介入を受けた人数,完了した人数は? (フローチャートでの図示を推奨)
14		参加者の募集	募集期間と追跡期間,日付は?
15		ベースラインのデータ	試験開始時における人口統計学的,臨床的特性(患者背景)は?
16		解析された人数	解析された各群の参加者数は? ITT(脱落者込みの解析)か?
17		転帰の結果とその推定	各群の結果の要約は? 精度は?(95%信頼区間など)
18		補助的解析	サブグループ解析などは?(多重性に言及すること) 事前に決定されていたものか,後付けの探索的なものか?
19		有害事象	各群の有害事象は?
20	考察	解釈	仮説,バイアス,精度低下の原因,解析の多重性を考慮して結果を解釈したか?
21		一般化可能性	外的妥当性は?
22		全体的なエビデンス	既存のエビデンスに照らしての解釈

表9.

文献	対象	デザイン・介入	評価項目	脱落例数	結果	有害事象	エビデンスのレベル	備考
(1) 報告者	(1) 総数	(1) RCT or 非RCT (対照群10例以上、比較群10例以上、計20例以上のものを取上げる)	(1) 主要アウトカム	可及的に詳細を記載	報告者の結論をそのまま記載	可及的に詳細を記載(結論に影響するもの、報告者のサマリーを参照)	表7を参照して1~5を決定	(1) 脱落例も含めて解析してあれば"ITT" それ以外は"ITTなし" または "不明"と記載
(2) 西暦年	(2) 年齢	(2) クロスオーバー or 同時対照 (parallel) or 記録対照	(2) 副次的アウトカム	わからない場合は"不明"と記載		わからない場合は"不明"と記載	担当者の治療法としての総論をもって勧告の代わりとする	(2) 詳細に記載されているもののみ"ランダム化の方法の記載あり"と表示
(3) 文献番号	(3) エントリー時における重症度	(3) (研究により) 前向き or 後向き						(3) 詳細に記載されているもののみ"盲検化の方法の記載あり"と記載
(4) 実施場所 (試験を行った場所が明確な場合のみ)	(4) その他ベースラインのデータ (患者背景、家族歴など絶対必要なもののみ)	(4) 各群の例数						(4) 併用療法について記載または不明とする
		(5) 実際の方法 (薬剤の名称・量・投与方法など)						
		(6) 比較対照の内容 (プラセボの名称・量・投与方法など)						
		(7) 追跡期間 (導入期間+試験期間) など						

4. 今後の展望

SRは定期的な更新が必要である。今後、既存の診療ガイドラインを補完するものとして、さらにグローバルスタンダードを満たすSRに発展させていく必要がある。今後の課題としては以下の点が挙げられる。

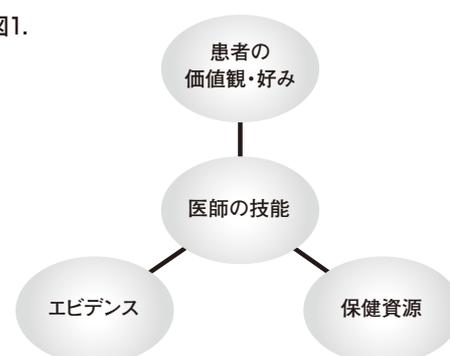
A. メタアナリシス meta-analysis

SRにおいて複数の研究結果の代表値を統計学的に統合すること (これを「定量的メタアナリシス」、SRを「定性的メタアナリシス」と呼ぶこともある)。代表的手法にPeto法、Mantel-Haenszel法、DerSimonian-Laird法等がある^{11,12)}。

B. 決断分析decision analysisと費用効果分析cost-effectiveness analysis

メタアナリシスによりデータを統合して診療方針を決定するための統計学的手法を決断分析という。決断分析において決断樹を作成し、予測される転帰が達成できる費用を指標とする費用効果分析が行われる。メタアナリシス、決断分析、費用効果分析により医療情報が最終的に統合される¹¹⁾。

図1.



参考文献

- 1) Gray JAM: Evidence-Based Healthcare, Churchill Livingstone, a division of Harcourt Brace and Company Ltd, London, 1997 (邦訳. 久繁哲徳監訳: 根拠に基づく保健医療, じほう, 東京, 2000)
- 2) 厚生省健康政策局研究開発振興課医療技術情報推進室監修: わかりやすいEBM講座, 厚生科学研究所, 東京, 2000
- 3) Sackett DL, Richardson WS, Rosenberg W, Haynes RB: Evidence-based Medicine: How to Practice and Teach EBM, 2nd ed., Churchill Livingstone, a division of Harcourt Brace and Company Ltd, London, 2000 (邦訳. Evidence-based Medicine—EBMの実践と教育, エルゼビア・サイエンス, 東京, 2003)
- 4) 別府宏圀, 津谷喜一郎編: コクラン共同計画資料集, 第2版, サイエンティスト社, 東京, 1998
- 5) Smethurst D, Macfarlane S: Atopic Eczema, In Clinical Evidence, Issue 9, BMJ Publishing, London, 2002 (邦訳. Clinical Evidence翻訳委員会訳: クリニカル・エビデンス, 日経BP, 東京, 2004, p1972-1990)
- 6) Hoare C, Li Wan Po A, Williams H: Systematic Review of Treatments for Atopic Dermatitis, Health Technology Assessment, 4(37), 2000
- 7) Clarke M, Oxman AD (eds): Cochrane Reviewer's Handbook 4.1.5, In The Cochrane Library, Issue 2, Update Software, Oxford, 2002
- 8) Moher D, Cook DJ, Eastwood S: Improving the quality of reports of meta-analysis of randomized controlled trials: the QUOROM statement, Lancet, 354:1896-1900, 1999
- 9) Phillips B, Ball C, Sackett D, Haynes B, Straus S, McAlister F: Level of Evidence, March 2002, Evidence-based On-call homepage (<http://www.eboncall.co.uk/>), 2002
- 10) Moher D, Schultz KF, Altman DG: The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials, Lancet, 357: 1191-1194, 2001
- 11) Petitti DB: Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine, Oxford University Press, Oxford, 1994 (邦訳: 福井次矢, 青木則明訳: EBMのためのデータ統合型研究—メタ分析, 決断分析, 費用効果分析の理論と実際. メディカル・サイエンス・インターナショナル, 東京, 1999)
- 12) 丹後俊郎: メタアナリシス入門: エビデンスの統合をめざす統計手法, 医学統計学シリーズ第4巻, 朝倉書店, 東京, 2002

付録

Williams教授に聞く

—臨床試験における左右比較の問題点をめぐって—

2003年8月11日、セルリアン・タワー東急ホテル（渋谷）にて。

出席者：Hywel C. Williams（英国Nottingham大学Centre of Evidence-based Dermatology教授）、
古江増隆（九州大学医学部皮膚科教授）、大矢幸弘（国立成育医療センター・アレルギー科医長）、
幸野 健（市立吹田市民病院皮膚科部長）、林 千恵子（書記：医療英語教育翻訳センター）

（Jは日本側、WはWilliams教授）

J： 臨床試験における同一個人内での比較（intra-individual comparison）、つまり左側に実薬、右側に対照薬を外用して比較するといった左右比較試験（left-right comparison）に関する国際的なコンセンサスについてお話をうかがいたいのですが。

W： コンセンサスと言っても、エビデンスなしの「いわゆるコンセンサス」というものもありますが、これらは「ノンセンサス」とでもいうべきです（笑）。左右比較試験の妥当性について、英国では本当の意味でのコンセンサスができています。英国でも左右比較試験は多数行われていますが、きちんと行われたものである限り、それが問題になったことはありません。日本では、一体どのような時に問題視されるのでしょうか？

J： 主に新薬認可の時に問題となります。左右比較では認可されないため。多数の患者さんを対象とする必要がありますし、個体差などのバラツキも多く大変な苦勞をしています。また米国の皮膚科雑誌でも、紫外線療法の試験など、左右比較ではアクセプトされないようですし、FDAも問題視するようです。

W： なるほど。今日用意して下さったMEDLINEの資料を見ても、左右比較の論文を掲載しているのはヨーロッパの雑誌が多いですし、著者も主にヨーロッパ人ですね。しかし、一体どのようなことが問題になるのでしょうか？

J： ある部位に実薬を使用すると血中に移行し、他部位すなわち対照薬使用部位などでも作用する、あるいは相互作用が起こるのではないかという全身的影響の問題、また患者さんが左右の使用薬を間違える、あるいは混合してしまうというコンタミネーションの可能性の問題などです。

W： 全身的影響やコンタミネーションというのは、比較する薬剤、たとえば実薬とプラセボで有効性に有意差が出なかった時に、本当に差がないのか明確でないため、問題となりえます。しかし、有意差が出たのであれば、なんら問題にならないと思います。ただし比較する治療法の有効性や副作用に差があり過ぎると、患者さんが自分でどちらが実薬か気がつき、盲検化が破れるという問題点は確かに残るでしょう。長期的なものは行いにくいかも知れませんが、短期的な試験なら問題ないと思います。しかし、全身的影響について、一体どんなエビデンスがあるのでしょうか？

J： 強いエビデンスはないと思います。ただ、同一個人内での比較あるいは左右比較に関するコンセンサスを明確に記載した国際的な文書がないということが問題だと思います。

W： 確かにそうですね。それは出すべきですね。

J: 我々は現在、厚生労働省科学研究「アトピー性皮膚炎の既存治療法のEBMによる評価」を実施しています。多数の臨床試験論文をOxford大学EBMセンターのエビデンスの質分類表 (http://www.cebm.net/levels_of_evidence.asp) によって分類しているのですが、左右比較試験はどのように分類したらよいとお考えでしょうか？

W: Oxfordの10レベルからなるヒエラルキー分類は非常に複雑で、我々は英国皮膚科学会の診療ガイドライン作成のために、より皮膚科に適合した分類を作ろうとしています。Dr. Cox (英国Cumberland病院) と共にBritish Journal of Dermatologyのeditorial (2003; 148: 621-5) でこの問題を提起しました。従って、我々としては、コクラン・レビューにおいても、英国・国民保健サービス (National Health Service; NHS) の医療技術評価 (Health Technology Assessment; HTA) プログラムのシステマティック・レビューにおいても、Oxford分類は使用していません。もっとも私の考えとしては、Oxford分類を使用する場合、その左右比較試験がランダム化されているのなら単純にランダム化比較試験 (レベル1あるいはグレードA)、ランダム化されていないのなら非ランダム化試験 (レベル2bあるいはグレードB) としてよいと思います。第一、あれだけ詳細なOxford分類でも、「左右比較は認めない」とか「レベル・グレードを下げること」などとはどこにも記載されていないでしょう。

J: では我々が分類した表での左右比較試験の扱いに「Williams教授も同意」と記載してよろしいでしょうか？

W: もちろんです。しかし、日本や米国で左右比較が問題視されているのなら、これは捨て置けない問題です。英国に帰ってから、まず手始めに、私が編集長をしているコクラン・スキングループのホームページ (<http://www.nottingham.ac.uk/~muzd/>) で我々の左右比較に関する考えを發表しましょう。そして外用薬の臨床試験に詳しいJ. Ferguson教授 (英国・Dundee大学)、ヨーロッパ皮膚疫学ネットワーク (Europe Dermato-Epidemiology Network; EDEN) のメンバーであるL. Naldi教授 (イタリア・Bergamo大学)、米国のR. S. Stern教授 (Harvard大学) らと相談し、Oxford大学EBMセンター、コクラン共同計画の方法論グループの面々からも意見を求めて、何らかのコメントを發表する用意をしましょう。

最近私は、M. Bigby (Harvard大学)、T. L. Diepgen教授 (ドイツ・Heidelberg大学) らと共に、“Evidence-based Dermatology” (BMJ Publishing Group) という本を編集しました。これは私のライフワークというべきものです。この本からスタートして今後継続的にインターネットを通じて改定して行く予定で、ホームページも作ってあります (<http://www.evidbasedderm.com/>)。今回のような重要なトピックスについては、日本からもどんどん提案して、この本にcontributeして行って下さい。私が最も重視しているのはこのような問題点・疑問点に関する国際的な協力関係です。

(文責：幸野、林)