

図 1 EBM に対する厚生労働省の取り組み

を行なうことが求められており、一定の明示的な作成プロセスを踏むことが重視されている。

2. 診療ガイドラインの構造と作成手順

ガイドラインを作成するテーマを設定したら、その問題の関係者を集めて作成グループを組織する。日本では学会の主導でガイドライン作成が進められることが多く、学会関係者という、ある意味では「仲間」、悪くいえば「内輪」で作業が進められることが多いが、欧米では stakeholders (利害関係者) を広く集め、さまざまな意見対立の中から合意を形成するプロセスが重視される。日本でも今後はガイドライン作成、とくに医療の領域でも社会性の高いテーマについては、このような取り組みの重要性が増していくであろう。

EBM の手法による診療ガイドライン作成は、まず推奨を示すべき臨床的課題 (clinical questions) を挙げ、それに対してシステマティック・レビューの方法に準じて、関連文献の系統的検索と吟味を行なう³⁾。臨床的課題は、EBM の第 1 ステップである「疑問の定式化」の方法を参考に、「どんな患者・疾患・病態に対し (Patient), 何をを行い (Intervention), どんな臨床的効果を期待するのか (outcome)」という視点で作成する。これに「何と比較して (comparison)」を加え、EBM では頭文字から“PICO”と称される。

推奨度の決定は、診療ガイドラインの核心である。あ

るテーマの文献が収集され、その内容が抽出された「診療エビデンス (clinical evidence) 集」は、方向性を示す推奨を含まない点で診療ガイドラインと決定的に異なる。推奨度は得られたエビデンスの質 (レベル) に大きく影響されるが、それだけで決定されるものではない。推奨度は、表 1 の要素を勘案して総合的に判断されるものである⁴⁾。

推奨度の表示形式を表 2 に示す。国内では 1990 年代前半に米国 AHCPR (Agency for Health Care Policy and Research, 現 Agency for Healthcare Research and Quality: AHRQ) が提案した方式が多くのガイドラインで用いられているが、判定が困難な場合も少なくない。特に「推奨度 C: 根拠が無いので推奨できない」の解釈については混乱が多いため、今後はこの AHCPR の 4 段階方式ではない表示方法を考慮することが望ましい。

日本脳卒中学会の合同ガイドライン (2001 年) では C

表 1 推奨度の決め方 (以下の要素を勘案して総合的に判断)

1. エビデンスのレベル
2. エビデンスの数と結論のバラツキ (同じ結論のエビデンスが多ければ多いほど、そして結論のバラツキが小さければ小さいほど勧告は強いものとなる。必要に応じてメタアナリシスを行う)
3. 臨床的有効性の大きさ
4. 臨床上の適用性
5. 害やコストに関するエビデンス

表 2 推奨度の形式

-
1. 米国 Agency for Health Care Policy and Research (1994 年)
- A: 行いよう強く勧められる
 - B: 行いよう勧められる
 - C: 行いよう勧めるだけの根拠が明確でない
 - D: 行わないよう勧められる
2. 脳卒中合同ガイドライン委員会 (2001 年)
- A: 行いよう強く勧められる (少なくとも 1つのレベル I の結果)
 - B: 行いよう勧められる (少なくとも 1つのレベル II の結果)
 - C1: 行いようを考慮してもよいが、十分な科学的根拠はない
 - C2: 科学的根拠がないので、勧められない
 - D: 行わないよう勧められる
3. 米国 臨床予防医学タスクフォース (2001 年)
- A: 臨床医は、対象となる患者に定期的介入 (ここでは予防サービス、以下介入とする) を提供することを強く推奨する。(USPSTF は、介入が重要な健康アウトカムを改善するとする強い科学的根拠を得て、利益は害よりはるかに大きいとの結論を下している。)
 - B: 臨床医は定期的に介入を対象となっている患者に提供するように推奨する。(USPSTF は少なくとも、介入が重要な健康アウトカムを改善するとする十分な科学的根拠を得て、利益は害より大きいとの結論を下している。)
 - C: 定期的な介入の提供についての是非に対する推奨は行わない。(USPSTF は、少なくとも介入が健康アウトカムを改善するとするかなりの科学的根拠を得たが、利益と害のバランスが、一般的な勧告を正当化するにはあまりにも接近しているとの結論を下している。)
 - D: 介入を定期的な非症候患者に提供しないよう推奨する。(USPSTF は少なくとも、介入が効果的でない、あるいは害が利益より大きいとするかなりの科学的根拠を得た。)
 - I: 介入を定期的に提供することの是非についての科学的根拠が不十分であるとする結論を下す。(介入が効果的であるとする科学的根拠の欠如、その質の低さや矛盾、また利益と害のバランスが決定できない。)
4. カナダ 予防医学に関するタスクフォース (2003 年)
- A: 予防的介入を推奨し得る十分な (Good) エビデンスがある
 - B: 予防的介入を推奨し得る適正な (Fair) エビデンスがある。
 - C: 既存のエビデンスは一致しておらず、予防的介入を肯定・否定する推奨は決定できない。しかし、他の要因が意思決定に影響し得るであろう。
 - D: 予防的介入を行わないことを推奨する適正な (Fair) エビデンスがある。
 - E: 予防的介入を行わないことを推奨する十分な (Good) エビデンスがある。
 - I: 推奨を決定するにはエビデンスが量的・質的に不十分である。しかし、他の要因が意思決定に影響し得るであろう。
-

判定を「C1: 行いようを考慮してもよいが、十分な科学的根拠はない」と「C2: 科学的根拠がないので勧められない」に区別している。棚橋は、同ガイドラインにおける急性期治療の推奨 71 のうち、A が 0 件、B が 6 件、C1 が 53 件、C2 が 8 件、D が 4 件であり、C1 が 75% を占めていることを報告している⁵⁾。これは、必ずしも「日本の臨床医学は、高いレベルのエビデンスに基づいて行われていない」という否定的な意味ではなく、「実際に臨床で行われていることの多くは高いレベルのエビデンスで裏付けられていないものである」という臨床の現実を意味していると捉えることもできる。

近年の諸外国での推奨度の表現方法としては次のような提案がある。臨床における予防医学的テーマの包括的ガイドラインである、米国の Preventive Services Task Force (<http://www.ahcpr.gov/clinic/uspstfix.htm>) の第 3 版では、ABCD の 4 段階の推奨度に新たに I 判定を加えている。

推奨度 C, I とも明確な推奨が難しい課題であるが、C は複数の信頼し得る研究報告が存在するが、利益と害が拮抗しているため一定の推奨が困難であるのに対し、I はエビデンスが無いが、その質が低いために推奨が決定できないケースがあたるとしている。表に示すようにカナダの予防医学に関するタスクフォースも米国に類似した、しかし独自の推奨度分類を提案している。これは医療の非専門家も含め関係者の誤解・齟齬を減らす意味で有益な視点と思われる。同タスクフォースは、推奨度 C のようにエビデンスが明確でない場合の意思決定・行動の基準を下記に示している (http://www.ctfphc.org/ctfphc&methods.htm#Table_1)。今後の国内におけるガイドライン作成においても検討を進めるべき問題の一つであろう。

表3 エビデンスが明確でない場合の意思決定の基準 (カナダ・タスクフォース)

・意思決定における患者自身の参加を促すこと
・害を最小化すること
・強い必要性が明らかな場合に関してのみ、大きな変化を主張すること
・不要な「ラベリング」を避けること
・益の不確かな高価な手技を避けること
・疾病負担が大きい状況に焦点を当てること
・ハイリスクグループの特別のニーズに配慮すること

3. 診療ガイドラインと「医師の裁量」

診療ガイドラインは日常診療の対象となる多くの症例に対して参考になるが、診療ガイドラインを個々の患者の状況、特性に応じて柔軟に利用することが必要である。欧米で「診療ガイドライン」は“starting point for discussion”として当事者の意思決定を「支援する」役割を担う。しかし今後、医療者の想定を越える領域でも、医療訴訟の判断基準として「診療ガイドライン」が社会的に認識されていく可能性は否定できず、それを確固とした判断の指針と捉える医療消費者も増えていくことが予想される。そうなれば臨床医が過剰な“defensive medicine”への傾向を強めることも懸念される。「診療ガイドライン」の意義と適切な社会的位置付けは医師だけではなく、法律家や一般の人々とも協力して取り組むべき重要課題である。

診療ガイドラインの意義が期待と共に語られるのと同程度に、診療ガイドラインに対する不安、疑問も投げかけられている。その一つがガイドラインという存在の持つ拘束力である。ガイドライン (guideline) は指針とも言われるが、この拘束力については国内では十分な議論がされていない。疫学辞典 (第4版) によると^{6,7)}、関連する用語として、指令 (directive)、勧告または推奨 (recommendation) 挙げられ、「指令は勧告 (推奨) よりも強く、勧告 (推奨) は指針よりも強い、北米では指針と勧告 (推奨) は同等である」と述べられている。この定義からも明らかなように、ガイドラインの役割はあくまで意思決定の支援であり、現場における個別の臨床行為・意思決定を規制するものではない。参考までに指令 (direct) よりも拘束力強いのは規制 (regulation) であり、この場合はそれに従わないと罰則を受けるなど、直接的な不利益が生じることが多い。Recommendation は従来「勧告」と訳されることが多かったが、「勧告」は医療法にも見られる法律用語であり、暗黙の拘束力が意図されている。一般の人々や法律家を交えたガイドラインに関する社会的議論に向けては、誤解を避けるために「勧告」

ではなく、「推奨」を用いることがより適切であろう。

診療ガイドラインがカバーするケースについて、Eddy は 60-95%の患者にとどまると述べ、95%以上の患者に適用される「スタンダード」、反対に 50%ほどの患者にしか適用されないものは「オプション」と区別している⁸⁾。しかし同じ「診療ガイドライン」でも、慢性疾患で予後への影響要因が多様なものと、進行がんに代表されるように特定の治療が予後に強く影響するものと、科学的な意味での拘束力は異なるであろう。従って、診療ガイドラインに関する議論は総論と共に、特定の臨床状況における各論的な検討を並行して深めていく必要がある。その意味でさまざまな状況で生じる診療ガイドラインをめぐる関係者の齟齬や葛藤、それらへの対応の事例を蓄積していくことは、診療ガイドラインの適切な利用・普及の視点から大いに重要と言える。

診療ガイドラインは臨床現場に大きな影響を与えるものであるが、臨床医の「プロフェッショナル・フリーダム」との整合が問題とされることが多い。長沢・津谷は日本語の「プロフェッショナル・フリーダム」と英語の“professional autonomy”の語義を、歴史的、文献計量学的に明らかにすることによって、「自律」をベースとする「積極的自由」は職業団体によって重視されるべきもので、日本語でも「プロフェッショナル・オートノミー」と称されることが適切であるとしている⁹⁾。この提案によれば医師はプロフェッションとして、積極的に診療ガイドラインの作成や、改訂のための意見表明を行なうことが期待されることを意味し、「診療ガイドラインは医師の裁量を拘束する」という従来からの危惧に対して、新しい視点を示すとも言える。

4. 臨床医は診療ガイドラインをどう見ているか？

Farquhar らは 1990-2000 年の刊行された英文論文から基準に合った 30 編のシステムティック・レビューにより、臨床医の診療ガイドラインに対する認識・態度を定量的に検討している¹⁰⁾。その結果、ガイドラインはアドバイス源として有用であることについて 75%、ガイドラインは医療の質を向上させることに 70%が同意している。一方で 34%が「ガイドラインは医師の裁量を減じ、医学を過度に単純化する」とし、41%が「医療訴訟を増加させるであろう」とし、さらに 53%が「ガイドラインは医療費抑制のためのものである」と回答していた。医療関係者は概ね診療ガイドラインに好意的な態度を示しているが、その実効性や医療費抑制、医療訴訟への影響についての懸念は少なくない。筆者はあるメーリングリストに参加しているプライマリケア医を対象として同様

の調査を行ない、614名中469名(76%)から回答を得た¹¹⁾。この結果、『「診療ガイドラインは医師の自由裁量を拘束する」と思うか?』という問いに対して、「まったく思わない(10%)」「思わない(48%)」に対して、「どちらともいえない(30%)」「思う(10%)」「大いに思う(1%)」と約4割が、診療ガイドラインが医師の裁量を拘束することに何らかの懸念を感じていることが推測される。

他の臨床領域においても、診療ガイドラインに対する臨床医の懸念とその理由・背景を明らかにし、それに応える説明とコミュニケーションを積み重ねていくことは、適切な形で診療ガイドラインが現場で普及していくために欠かせない過程となるであろう。

5. 診療ガイドラインの評価

有効な診療ガイドラインの要件として、Grimshawは①内容が妥当か? ②普及し利用可能か? ③実際に利用されているか? ④患者アウトカムの改善を挙げ、診療ガイドラインの臨床行為への影響・患者アウトカムの改善の視点からレビューを行なった¹²⁾。その結果、59の診療ガイドラインのうち55で臨床行為の改善が見られたこと、患者アウトカムへの影響を検討した11論文では2論文を除いてアウトカムが改善したことを報告している。小山らは1991年から2000年にかけて発表された、学会または国家機関レベルで作成された治療に関する診療ガイドラインが患者アウトカムに与えた効果を、ランダム化比較試験または準実験的研究デザインにより検討した研究のシステマティック・レビューを行った¹³⁾。17篇の論文が同定され、うち10篇(59%)で患者アウトカムの改善が示されていた。また、患者アウトカムの改善が示されなかった研究のうち、2篇ではガイドライン遵守率が低く、3篇ではガイドライン遵守率が報告されていなかった。本研究の結果は、診療ガイドラインを診療上利用することにより患者アウトカムが改善される可能性を強く示唆している。

著者らは1997年から2001年にかけて国内で公表された48診療ガイドラインをShaneyfelt¹⁴⁾らの方法で評価した。その結果、平均値は25点満点で8.49点(標準偏差4.28点、範囲:0~18.5点)であった¹⁵⁾。1985~97年に発表された米国の279診療ガイドラインは25点満点で10.77点であり、それと比べると国内で作成された診療ガイドラインの内容はまだ十分ではないことがわかった。25項目の評価ポイントの中で80%以上の遵守率が5項目、遵守率が10%未満であったのは8項目であり、内容別に見ると「ガイドラインの作成方法と様式について」

の10項目は平均遵守率が60.0%、「エビデンスの検索・要約について」の10項目は平均遵守率が13.0%、「勧告の作成方法について」の5項目は平均遵守率が23.8%であった。

診療ガイドラインの評価に関する国際的な取り組みとして、European Union(EU)を中心に発足したAGREE共同計画(“Appraisal of Guidelines Research and Evaluation” <http://www.agreecollaboration.org/>)がある。AGREEのグループは発表された診療ガイドラインの質を調べるため、次に示す6領域23項目と総合評価から成る評価手法を提示している^{16,17)}。

範囲と目的(項目1-3):ガイドラインの全体的な目的、具体的な臨床問題と対象とする患者集団に関する項目。利害関係者の関与(項目4-7):ガイドラインがそのユーザーとして意図された者の見方をどれほど代表するものであるかに関する項目。

開発の厳密さ(項目8-14):根拠を集め集約するのに用いられた過程と、推奨を導き出す方法、その更新に関する項目。

明快さと提示(項目15-18):ガイドラインの言葉と形式に関する項目。

適用性(項目19-21):考えられるガイドライン適用、組織的、行動的、経済的影響に関する項目。

編集の独立性(項目22-23):ガイドライン作成グループが利害の衝突を生じる影響力から独立していることに関する項目。

内容を見れば分かるようにAGREEによる診療ガイドラインの評価は臨床医の行動や患者アウトカムの変化ではなく、診療ガイドライン作成の「枠組み」の評価に焦点を当てている。2003年9月には米国が中心となって行なわれたThe Conference on Guideline Standardization(COGS)の成果が報告された¹⁸⁾。COGSの提案は作成された診療ガイドラインを報告する際のチェック項目であり、既に発表された診療ガイドラインの評価法としてのAGREEの提案と表裏一体である。表4にCOGSの提案とAGREEの方法の対応を示す¹⁹⁾。COGSの提案は推奨が導き出された過程の透明性を確保するだけでなく、作成者、目的、資金提供者に始まり、可能な場合は有効期限を示すことを求めるなど、その診療ガイドラインの作られた枠組みに関する情報を重視している。これは情報学でいうメタデータ、すなわち「データに関するデータ」の一種であり、個々の情報の持つ意味を利用者が判断する際の有益な手がかりとなる。このような国際的な標準様式に沿った形で診療ガイドラインを作成し、報告することが、今後、質の高い診療ガイドラインを世に出すための必須要件となっていくことが予想される。

表 4 診療ガイドライン作成・報告のチェックポイント (COGS と AGREE の対応)

COGS の項目	説明	対応する AGREE 評価ポイント
1. 要約 (Overview material)	ガイドラインのリリース日、ステータス (初版, 改訂版, 最新版), 活字情報及び電子情報の出典などの情報を含む構造化抄録を提出すること。	
2. 焦点 (Focus)	ガイドラインの対象となる基本的疾患・健康状態や介入・サービス・技術に関する記載を加えること。ガイドラインで採択された以外の予防上, 診断上, 治療上の介入で, ガイドライン作成段階で検討にのぼった介入についても明示すること。	2. ガイドラインが取り扱う臨床の課題が明確に記載されている 16. 管理に関する異なるオプションが明確に示されている
3. 目標 (Goal)	ガイドライン遵守によって達成されるべき目標を記載すること。なぜこのトピックに関するガイドラインを作成するのかについての理由説明も加えること。	1. ガイドライン全体の目的が明確に記載されている
4. 利用者・利用場面 (Users/setting)	ガイドラインの利用対象者 (医療提供者の種類, 患者) 及びガイドラインが利用されるべき場面について記載すること。	6. そのガイドラインの対象となる利用者が明確に示されている。
5. 対象集団 (Target population)	ガイドライン推奨の対象となる対象患者集団について記載し, 除外基準があれば全てこれを記載すること。	3. ガイドラインの対象として想定された患者が明確に記載されている
6. 作成者 (Developer)	ガイドライン作成に責任を持つ組織, 及びガイドライン作成に携わる人員の氏名, 資格, 利害関係の衝突の可能性を明示すること。	4. ガイドライン開発グループの個人個人は, 該当する全ての専門家集団からなっている 23. ガイドライン開発メンバーの利害関係が記録されていた
7. 資金提供者・スポンサー (Funding source/sponsor)	資金提供者・スポンサーを明示し, ガイドライン作成・報告における役割を記載すること。利害関係の衝突が考えられる場合は, これを開示すること。	22. ガイドラインは編集に関してその資金提供元から独立している 23. ガイドライン開発メンバーの利害関係が記録されていた
8. エビデンスの収集 (Evidence collection)	学術文献を検索するのに用いた手法を記載し, 検索の期間, 検索したデータベース, そして収集した文献を絞り込むのに用いた基準についての記載も含めること。	8. 根拠の検索には系統だった手法が用いられた 9. 根拠の選択基準が明確に記載されている
9. 推奨度決定基準 (Recommendation grading criteria)	推奨の根拠となるエビデンスの質を評価するのに用いた基準, そして推奨度を記述するためのシステムについて記載すること。推奨度は, 推奨への遵守がどれだけ重要かを示すものであり, エビデンスの質, そして予測される利益と害の程度に基づいて評価される。	10. 推奨を導き出した方法は明確に記載されている 11. 推奨を導き出すにあたって, 健康上の利点, 副作用, 危険性が考慮された
10. エビデンスの統合のための手段 (Method for synthesizing evidence)	推奨作成に際してエビデンスをどう用いたかを記載すること。たとえば, エビデンスの表, メタ分析, 判断分析など。	
11. リリース前のレビュー (Prerelease review)	リリース前にガイドライン作成者がどのようにガイドラインをレビューし, 及び/又は, 審査したかについて記載すること。	13. そのガイドラインは出版前に専門家による外部評価を受けた
12. 更新の計画 (Update plan)	ガイドライン更新の計画があるか否かを記し, 適用可能な場合は現在出ている版の有効期限を記すこと。	14. ガイドラインの更新手順は示されている
13. 定義 (Definitions)	一般的でない用語を定義すること。また, 間違っ て解釈される可能性のあるガイドラインについては, これが正しく適用されるために確実に理解されなければならない用語を定義すること。	
14. 推奨及び理由説明 (Recommendations and rationale)	推奨される対処法及びこれが実行されるべき具体的な状況を正確に記すこと。推奨と, 推奨の根拠となるエビデンスとの関連性を説明することにより, 各推奨の正当性を示すこと。9に記載されている基準に基づき, エビデンスの質, 推奨の推奨度を示すこと。	12. その推奨とその基となった根拠との関連が明確である 15. その推奨は具体的で, あいまいさのないものである (17. 容易に重要な推奨を見分けられる)
15. 考えられる利益と害 (Potential benefits and harms)	ガイドラインの推奨を実施することによって予測される利益, そして考えられる害について記載すること。	(20. その推奨の適用に伴う費用に関して考慮された)

16. 患者の希望 (Patient preferences)	推奨が患者の選択や価値観に大きく関わるものであった場合、患者の希望の扱い方について記載すること。	5. 患者の視点や選好は考慮された
17. アルゴリズム (algorithm)	適切であれば、ガイドラインで取り上げられている診療における段階や意思決定を図表に表し提供すること。	18. ガイドラインは適用手段によって指示されている
18. 実施における検討事項 (Implementation considerations)	推奨の適用において予測される障壁について記載すること。医療提供者と患者のために、推奨実施を円滑にするような関連文書については全て参考指示すること。ガイドライン実施による医療内容の変化を計測するための評価基準を提案すること。	19. 推奨を適用するにあたって考えられる組織の障害が検討された 21. ガイドラインには監視・監査のための主要な評価基準が示されている
		7. ガイドラインはエンドユーザー利用者によって試験的に用いられた

6. 医学研究・診療ガイドライン作成における「利害の衝突」

「利害の衝突 (Conflict of Interest)」は海外では既に学術活動の公正さに対する大きな脅威として認識されており、国内でもより多くの関心を向けられるべき問題である。医学雑誌編集者委員会 (International Committee of Medical Journal Editors: ICMJE, <http://www.icmje.org/>) による「生物医学雑誌の統一投稿規程」では、「著者、査読者、及び編集者といったピア・レビュー (査読) 及び出版過程に関わる関係者が、自らの判断に不適切な影響を与え得る活動に関わっている場合、それが現実に影響を与えるかどうかは別として、寄稿された原稿に関する『利害の衝突』が存在する」としている。企業との財政上の関係 (例えば雇用、コンサルタント業、株式の所有、謝礼、専門家の立場での証言などを通じた関係) は、直接的なものでも直系の家族を通じたものでも、これは通常最も重大な「利害の衝突」としてみなされる。それ以外に個人的人間関係、学術的な競合、及び知的情熱なども原因となる。同規定では査読及び出版に関わる関係者らは自らの利害衝突の状況について明らかにし、他の関係者が、それによって自らが被るかもしれない影響を自分で判断できるように、そういった情報がアクセス可能であることを求めている。

原著論文における「利害の衝突」の影響を知るのに比べ、総説論文や論評では「叙述のニュアンス」に対する影響となるため、その察知はより困難である。個々の研究論文の研究デザインや、ある臨床的疑問に関する利用可能なエビデンスに対するエビデンス・レベルの判定はある程度、客観的に可能であるが、行動の方向性を示す推奨度の決定には多くの要因が影響し、作成者の恣意が入りやすい。多くの論文情報が集約され、推奨の形で方向性が示される診療ガイドラインでは、「利害の衝突」が

無視できない影響を与えることに、関係者はより多くの注意を向ける必要がある。Lenzer はガイドライン上での tPA 「格上げ」に AHA や医師への企業献金が影響していた可能性を指摘し²⁰⁾、また Choudhry らはガイドライン作成者の 58% は研究資金供出を受けていたことを報告している²¹⁾。これまで国内で作成されてきた診療ガイドラインでは、作成者の「利害の衝突」の開示はほとんど行われていない。ガイドライン作成の途中、または発表後に企業が直接、間接に自社製品の記述に対して影響を与えようとした事例が聞かれることも事実であり、今後、国内でも作成プロセスの一層の透明性向上が強く望まれる。現在作成中のいくつかのガイドラインでは、その記載が加えられつつあり (例: 日本褥瘡学会による褥瘡局所治療ガイドライン、日本耳科学会による急性中耳炎ガイドラインなど)、この方向性が広く認知されていくことが望まれる。

7. 診療ガイドライン作成における「患者参加」について

患者と医師の関係からは、患者・家族と医療者の対話の結節点としての診療ガイドラインが機能する可能性を探る意義は大きい。前述の AGREE による評価項目の一つとして、「患者の視点や選好は考慮されたかどうか」が問われており、「診療ガイドライン開発にあたって、患者の経験と期待に関する情報を知っておかねばならない。そのための方法として開発グループに患者の代表を含める、患者のインタビューから情報を得る、開発グループが患者の経験に関する文献をレビューする、などである。この手順が行われたという記述がなければならない」ことが求められている。

患者の代表者が診療ガイドライン作成に関与することは、厚生労働科学研究による診療ガイドライン作成の方針を示した「診療ガイドラインの作成の手順 ver. 4.3」

(<http://minds.jcqhc.or.jp/st/svc115.aspx>)⁴⁾でも述べられているが、国内ではその取り組みは遅れている。しかし欧米では、診療ガイドラインを作成するのは「その領域の権威」に限定されず、対象課題が設定されると、それに応じた様々な“stakeholders (利益関係者)”がそれぞれの立場で発言の機会を与えられることが一般的である。診療ガイドライン作成の初期から「重要な関係者」である患者の参加を得ることは決して容易ではないが、現場で臨床医と患者の双方の視点で有用な診療ガイドラインが作られていくためには、医療者が向き合うべき重要な課題である。

実際に患者の視点を取り入れるためには、準備期間が必要であろう。今のままで何人かの患者がその代表としてガイドライン作成の会議に参加したとしても、実質的な議論への十分な貢献は期待できない。それは患者にその力が無いのではなく、「必要な知識を事前に習得する機会が提供されていないこと」と「個人への依存度が高く、システム化されていないこと」が大きな原因と思われる。英国・イングランドでは、診療ガイドライン作成・普及に責任を持つ National Institute for Health and Clinical Excellence (NICE) の内部に“Patient Involvement Unit”という専門部局があり、診療ガイドライン作成に参加する患者代表をコーディネートすると共に、参加者に対して事前にトレーニングセッションを提供している²²⁾。診療ガイドラインが、限られた領域の専門医だけで作られるのではなく、社会的な役割を果たしていくためには、適切な“stakeholders”が初期の段階から関与していることが望ましい。海外の先進的な事例を学び、日本における今後の方向性を探る手がかりを得ることは大きな意味がある。

国内では厚生労働科学研究による喘息診療ガイドライン作成班（主任研究者：宮本昭正）に、患者支援団体の一つである「アラジーポット」が参加し、患者・支援者向け情報作成に大きな役割を担い^{23,24)}、一般紙でも取り上げられた²⁵⁾。他の領域でもこのような取り組みに関する検討が進むことが望まれる。

稲葉は法律家の立場から、医療者と患者が共にガイドラインの限界と役割を理解し、医療者は責任と倫理を踏まえて患者の陥りやすい問題を把握し、診療ガイドラインを用いてインフォームド・コンセントを行ない、対話の中で治療方針を共に決めていく調和的な医療モデルを提案している²⁶⁾。患者の視点も取り入れて作られた診療ガイドラインは、それを活用すればインフォームド・コンセントを巡るトラブルを回避するだけでなく、インフォームド・コンセント自体を充実させ、その結果として医療の質・安全性、そして患者満足度を高めるための

中核的な手段となり得るであろう。EBMの手法による診療ガイドラインが、臨床現場にとどまらず、社会的にも適切な認知と支援を受けるためには、EBMを基盤として、そこに臨床倫理や法律的な視点からの検討を加えていくことが強く望まれていると言えよう。

8. EBMとエビデンスをめぐる誤解

EBMの認知の高まりと共に、頻繁に耳にするようになったことの一つが、「EBM」と「エビデンス」の混同である。EBMは、医師の専門性や経験、熟練・患者の価値観・エビデンスの3要素をバランスよく統合し、よりよい患者ケアのための意思決定を行うものである²⁷⁾。このように、エビデンス自体はEBMの要素のうちの一つに過ぎない。しかし、現在の日本では、エビデンス・レベルが高いとされるランダム化比較試験、とくに大規模臨床試験の知見が得られれば「EBM」が確立し、臨床現場の意思決定までもが決まってしまうという短絡的な解釈がなされているケースが見られる。近年、数多く出版された総説・解説書でも、著名な臨床試験の解説をもってEBMとしている場合が少なくない。改めて言うまでもなく、臨床試験の知見はエビデンスの一つであり、EBMそのものではない。EBMとの混同と共に見られる「エビデンス」の過大視は、これまでの臨床的意思決定が、担当医師の経験・熟練に負う比重が大きかったことへの反動かもしれない。Haynesらは2002年のBMJ誌でEBM時代の臨床医・患者が意思決定に際して考慮すべき要因を整理し（図2）、併せて“Evidence does not make

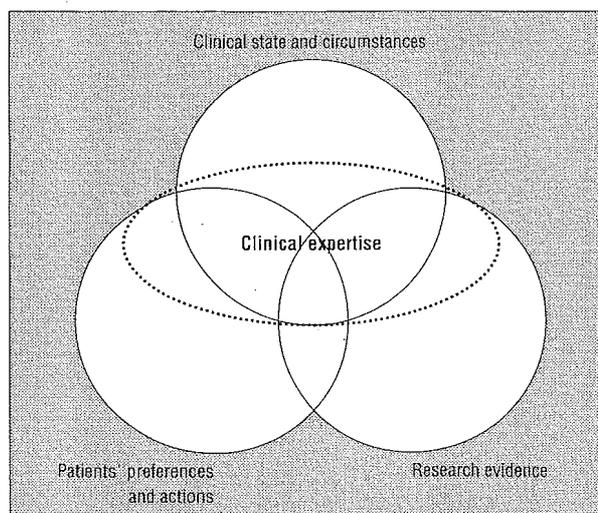


図2 evidence based practice における臨床医と患者の選択 (Haynes, et al. BMJ 2002).

decisions, people do”と強調している²⁸⁾。EBMの主唱者たちによるシンプルかつきわめて正当なメッセージであり、国内でもこの原点を再確認することが改めて望まれる。

9. ランダム化比較試験と観察研究の関係

EBMの普及と共に、広く知られるようになった概念の一つが「レベル・オブ・エビデンス（エビデンス・レベル）」である。個々の研究としてもっともレベルの高い、すなわち妥当性の高いエビデンスを提供するのはランダム化比較試験（RCT）であり、その次が非ランダム化比較試験、続いて観察的な疫学研究であるコホート試験やケースコントロール研究とされている。その下に位置する症例集積は多くの臨床研究報告で見られるが、対照群を持たないため、そこで観察された知見の一般化に大きな限界がある。さらに、高いレベルのエビデンスに基づかない権威者の個人的見解（エキスパートオピニオン）は、もっとも低いレベルのエビデンスとして位置づけられている。RCTの上に、個々のエビデンスを集約した2次情報であるシステマティック・レビューやメタアナリシス（データ統合型研究と呼ばれることもある）が置かれる場合もある。このように研究方法、すなわち研究デザインによって、得られるエビデンスの科学的妥当性に上下があるという考え方は、疫学・臨床疫学においては広く認識されていたが、それらの方法論に馴染みの薄かった多くの臨床医には大きな衝撃であったことが推測される。

エビデンス・レベルに関する第一の誤解は、AHCPRによる治療の有効性に関するエビデンスのレベルが、国内のガイドラインムーブメントの初期に広く紹介されたため、治療の有効性評価以外の臨床的課題についても、それと同じエビデンス・レベルの考え方が誤って用いられてしまっている点である。Oxford EBM centreでは、対象課題を“Therapy/Prevention, Aetiology/Harm”, “Prognosis”, “Diagnosis”, “Differential diagnosis/symptom prevalence study”, “Economic and decision analyses”の5カテゴリーに分け、それぞれの課題の特性にあった適切なエビデンスを考慮してレベルを付与している (http://www.cebm.net/levels_of_evidence.asp)。その結果、“Therapy/Prevention, Aetiology/Harm”では、RCTのシステマティック・レビューが最上位となるが、“Prognosis”, “Diagnosis”, “Differential diagnosis/symptom prevalence study”ではコホート研究などの観察研究のシステマティック・レビューが最上位とされている。この点については国内のガイドライン作成者も注意を払う必要があるだろう。

エビデンス・レベルの概念の明確化を通して、研究デザインの重要性が認知されるようになったのは好ましいことではある。しかし、それが行き過ぎて、「RCTでなければならない」あるいは「RCTでありさえすればよい」というような新たな誤解が生まれたのも確かであろう。RCTは一般的に何らかの介入の有効性を評価する目的で用いられる研究デザインであるが、少なくともこれまで国内では、治験に代表されるように薬物療法の評価という文脈でもっとも受け容れられてきた。それに対して、たとえば外科領域では、2種類の方法を施術者と患者の意図を考慮せずにランダムに割り付ける介入試験を行うこと自体が実施困難な場面が多い。そこから、「外科ではEBMを行えない」という話も時々聞かれる。しかしEBMの実践のためには、それぞれの領域の特性を考慮して、そこで得られる最良のエビデンス、すなわち“best available evidence”を利用すれば良いのであり、RCTによるエビデンスがなければできない、というものではない。一般的にRCTは科学的妥当性の高いエビデンスを提示する研究デザインであるとされているが、一方で、ランダム配置に伴う倫理的な制約、厳密な組み入れ・除外基準によって選ばれた対象者での（限定的な）知見であるという一般化可能性の問題、次々と開発される新しい手法・手法の後手に回りかねない適時性の問題（ある手法の有効性についてRCTで結論が得られた時点で、その手法は陳腐化しており、次の新しい手法が現場で利用され始めている）などがある。また、費用対効果を考えれば、すべての臨床的課題をRCTで検証することは現実的ではない。一方で、エビデンス・レベルとしてはRCTに比べて低いとされている観察研究は、上記のRCTの欠点を補うものとして臨床現場で有用となる場合もある。1970年から80年代にかけて国内外で、観察研究がRCTに比べて介入の有効性を過大評価することを指摘する報告が続き^{29,30)}、観察研究は各種のバイアスを制御しきれないため、有効性評価の目的で用いるのは望ましくないという考え方が一般的に受け容れられた。臨床疫学、EBMの台頭と共に、その認識はさらに広く定着していったと言える。しかし、Benson³¹⁾、Concato³²⁾らは2000年のNew England Journal of Medicine誌に、近年報告された観察研究とRCTの結果をメタアナリシスの統合値として比較すると、有効性の評価に大きな違いは見出されないという発表をそれぞれ独立に行った。観察研究（特にコホート研究）とRCTの関係について改めて問題を提起したと言えるが、それに対する反論も多い^{33,34)}。エビデンスを作る立場としては、RCT、観察研究、いずれの方法を選ぶことが良いか、それぞれの臨床的課題の特性、意思決定に必要とされるエビデンスのレベル、それぞれの研究

の実行可能性など、総合的に勘案して判断を行う必要があるだろう。

10. 補完医療・代替医療領域での診療ガイドライン

改めて言うまでも無く、補完医療・代替医療への社会的関心は大いに高い。その関心、期待に応えるための科学的検証、EBMの適用の必要性は関係者が等しく共有されているところであろう。現時点で補完医療・代替医療

領域で発表されている診療ガイドライン（英語）の状況を把握するためにPubMedで下記の検索を行ったところ17件が得られた（表5）。

English[la] AND (complementary medicine[majr] OR alternative medicine[majr]) AND practice guideline[pt]

さらに、これに“Evidence”を掛け合わせると、NIH Office of Alternative Medicineが1997年にMedical Care誌に発表した補完医療・代替医療のエビデンスのレビュー結果のみが得られる。この報告では、補完医療・代替医

表5 PubMedで検索された代替・統合医療領域の「ガイドライン」関連論文（2005年7月時点）

著者	タイトル	書誌事項
1: Royal College of Midwives.	Normal childbirth.	RCM Midwives. 2004 Aug; 7(8): 332.
2: Jankovic M, et al.	Psychosocial Issues in Pediatric Oncology. Nonconventional therapies in childhood cancer: guidelines for distinguishing non-harmful from harmful therapies: a report of the SIOP Working Committee on Psychosocial Issues in Pediatric Oncology.	Pediatr Blood Cancer. 2004 Jan; 42(1): 106-108.
3: American Association of Clinical Endocrinologists.	American Association of Clinical Endocrinologists medical guidelines for the clinical use of dietary supplements and nutraceuticals.	Endocr Pract. 2003 Sep-Oct; 9(5): 417-470.
4: IGDA Workgroup, WPA.	Idiographic (personalised) diagnostic formulation.	Br J Psychiatry Suppl. 2003 May; 45: s55-57.
5: American Diabetes Association.	Unproven therapies.	Diabetes Care. 2003 Jan; 26 Suppl 1: S142.
6: American Diabetes Association.	Unproven therapies.	Diabetes Care. 2000 Jan; 23 Suppl 1: S110.
7: American Holistic Nurses' Association.	Position on the role of nurses in the practice of complementary and alternative therapies.	Beginnings. 2001 Mar-Apr; 21(2): suppl 1-2.
8: American College of Obstetricians and Gynecologists Committee on Practice Bulletins-Gynecology.	ACOG Practice Bulletin. Clinical Management Guidelines for Obstetrician-Gynecologists. Use of botanicals for management of menopausal symptoms.	Obstet Gynecol. 2001 Jun; 97(6): suppl 1-11.
9: American Veterinary Medical Association.	AVMA Guidelines for Complementary and Alternative Veterinary Medicine.	J Am Vet Med Assoc. 2001 Jun 1; 218(11): 1731.
10: Fayeton S, van Wassenhoven M; Working Group 'Clinical Verification of Symptoms', European Committee for Homeopathy.	Clinical verification of symptom pictures of homeopathic medicines.	Br Homeopath J. 2001 Jan; 890(1): 29-32.
11: Saskatchewan Registered Nurses' Association.	Complementary and alternative therapies: role and responsibilities of the registered nurse.	Concern. 1999 Feb; 28(1): 16-18.
12: Firenzuoli F, Gori L.	Guidelines in phytotherapy. Phytotherapy in Italy. Associazione Nazionale Medici Fitoterapeuti (ANMFIT).	J Altern Complement Med. 1999 Apr; 5(2): 219-220.
13: Coates JR, Jobst KA.	Integrated healthcare: a way forward for the next five years? A discussion document from the Prince of Wales's Initiative on Integrated Medicine.	J Altern Complement Med. 1998 Summer; 4(2): 209-247.
14: Darley MA.	Complementary therapies: the position of the UKCC [practice guideline]	Complement Ther Nurs Midwifery. 1995 Aug; 1(4): 106-109.
15: Levin JS, et al.	Quantitative methods in research on complementary and alternative medicine. A methodological manifesto. NIH Office of Alternative Medicine.	Med Care. 1997 Nov; 35(11): 1079-1094.
16: American College of Obstetricians and Gynecologists.	ACOG committee opinion. Female genital mutilation. Number 151—January 1995. Committee on Gynecologic Practice. Committee on International Affairs.	Int J Gynaecol Obstet. 1995 May; 49(2): 209.
17: American Holistic Nurses' Association.	Code of ethics for holistic nurses.	J Holist Nurs. 1992 Sep; 10(3): 275-276.

療のエビデンスが関係者の予想に反して、良質のものが多かったことが述べられているが、それらを用いて診療ガイドラインの形にまとめることは目的とされていない。表の文献を概観すると、海外でも補完医療・代替医療領域でEBMの手法を用いた診療ガイドラインはこれからの課題であることが推測される。

今後、国内において補完医療・代替医療領域の診療ガイドラインの作成が進められる場合は、特に次の3点について十分な検討を要する。

- (1) 何の目的で
- (2) どういう立場の人間が
- (3) 誰を対象として 診療ガイドラインを作成するのか？

ガイドライン作成という大きな作業の足場となる上記の3点については、その動きを進めようとする関係者の間で十分な議論を尽くし、コンセンサスを形成しておく必要がある。その上に立って、既存のエビデンスを系統的に収集、評価し、本稿で述べてきたように、それらの長短、特性を理解して活用すること、そして医療者、患者の意思決定の支援であるという診療ガイドラインの趣旨を十分理解して、作成・利用・普及の仕組みを構築していくことが望まれる。

11. 「EBMの手法を用いた診療ガイドライン作成」が明らかにした課題

海外では診療ガイドラインで推奨を示すことに異議を唱えるグループも少なくない。その代表が英国のBMJ (British Medical Journal) group から出されている“Clinical Evidence” (<http://www.clinicalevidence.com/cweb/conditions/index.jsp>) であろう。その主張は、“We supply evidence, you make the decisions. The experience of the clinical practice guideline movement has shown that it is nearly impossible to make recommendations that are appropriate in every situation.”

であり、推奨を核心とする診療ガイドラインの問題点を指摘し、傾聴に値する。“Clinical Evidence”のような「診療エビデンス集」は判断を（診療ガイドライン以上に）個人に委ねているが、臨床医の立場からは得意な領域では十分活用できて、経験の限られているケース、専門外の課題では、客観的なエビデンスに加えて「結局、一般的にはどうすることが勧められるのか」という専門家の推奨が明示されたガイドラインが好まれる可能性も十分ある。そういった意味で、診療エビデンス集と診療ガイドラインは相補的に用いられるのが良いと思われる。

1999年以降、国内ではEBMによる診療ガイドライン作成がさまざまな学会を中心に行なわれ、その成果が発表された。その作業過程で多くの課題が明らかにされたが、その中で最も大きなものの一つは「日本人を対象としたレベルの高いエビデンスがきわめて乏しい」事実の確認であった。基礎的な医学研究が偏重され、臨床現場の意思決定を支援する臨床研究や疫学研究が立ち遅れていた日本の研究風土の特性^{35,36)}が、期せずして明らかにされたと言える。

特にインパクトファクターに象徴される研究業績の得点主義は、国内において臨床医を含む医学研究者の志向に少なくない影響を与えている。インパクトファクターによる学術誌やその研究領域のランク付けと、診療ガイドラインに引用される文献（臨床の問題解決に貢献する研究と考えられる）を多く掲載している学術誌の評価は一致していない³⁷⁾。EBMと診療ガイドラインの認知が高まると共に、実際の患者の役に立つエビデンスを作るといった問題意識と適切な研究方法論が普及することで、真に意味ある臨床研究が活性化されていくことが期待される。

診療ガイドラインは医療の質を向上させるための特効薬ではない。使い方を誤れば瞬時に両刃の剣に転じる。その意義、役割、そして限界を医療者と社会一般の人々がどのように共有していくべきか、開かれたディスカッションを積み重ねていく必要があろう。診療ガイドラインに関する世界的な動きとして2002年にはWHOを含む24カ国46組織よりなる“Guideline International Network (GIN)”が発足し、2004年には推奨度の決定に関するGRADEの提案などが発表された³⁸⁾。いずれも今後の診療ガイドライン問題を考える上で大きな意義を持つ取り組みである。ガイドラインへのアクセスビリティを高めるため、米国では国立ガイドライン・クリアリングハウス (<http://www.guideline.gov/>)、英国では健康情報電子ライブラリー (<http://www.nelh.nhs.uk/>) がインターネットを通じて診療ガイドラインを広く一般の人々へも公開している。国内でも現在、財団法人日本医療機能評価機構の「医療情報サービス事業」 (<http://www.ebm.jcqhcc.or.jp/>) 通称・“Minds” (Medical Information Network Distribution Service) の整備が進められている³⁹⁾。

これまで述べてきたように、診療ガイドラインは無批判に遵守されるべきものではなく、世界的な動向、基準に照らしたcriticsにさらされるものである。利用者は、診療ガイドラインの限界を知った上で、専門性と主体性を発揮して（使わないという選択肢も含めて）その情報を活用することが望まれている。

医療に対する関心の増大、社会における医療情報のイ

ンフストラクチャーの充実と共に、医療における患者と医師の関係の変化は加速されるであろう。EBM, そして診療ガイドラインを通じて、医療は社会における、そして社会に対するそのアカウンタビリティを問われていると言えよう。

参 考 文 献

- 1) Institute of Medicine. Guidelines for clinical practice: from development to use. Washington DC, National Academy Press, 1992.
- 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. 平成10年度医療技術評価推進検討会報告書. わかりやすいEBM講座. 東京. 厚生科学研究所. 2000: 8-28.
- 3) 中山健夫. EBMを用いた診療ガイドライン: 作成・活用ガイド. 東京. 金原出版. 2004: 78-80.
- 4) 福井次矢, 丹後俊郎. 診療ガイドラインの作成の手順. EBMジャーナル. 2003; 4: 28-37.
- 5) 棚橋紀夫. EBMに基づく治療ガイドラインの作成と応用の問題点. 第44回日本神経学会総会プログラム・抄録集. 2003; 61.
- 6) J. Last. A Dictionary of Epidemiology (4th edition). Oxford University Press (Oxford) 2000.
- 7) Nakayama T, Budgell B, Tsutani K. Confusion about the concept of clinical practice guidelines in Japan: On the way to a social consensus. *Int J Qual Health Care* (in press).
- 8) Eddy DM. Clinical decision making: from theory to practice. Designing a practice policy. Standards, guidelines, and options. *JAMA*. 1990;263(22):3077, 3081, 3084.
- 9) 津谷喜一郎, 長澤道行. 医師と診療ガイドライン: “professional autonomy”の視点から. *日本医師会雑誌*. 2003; 129(11): 1793-1803.
- 10) Farquhar CM, Kofa EW, Slutsky JR. Clinicians' Attitudes to Clinical Practice Guidelines. *Med J Aust*. 2002;177(9):502-506.
- 11) 中山健夫, 福原俊一, 武田裕子ら. 一般臨床医のEBMと診療ガイドラインに対する意識について: プライマリケア医を中心とするメーリングリスト登録者を対象とした質問表調査. 厚生労働科学. EBMを指向した「診療ガイドライン」と医学データベースに利用される「構造化抄録」作成の方法論の開発とそれらの受容性に関する研究 (主任研究者・中山健夫). 2001年度報告書. p. 41-45.
- 12) Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993; 342(8883): 1317-1322.
- 13) Koyama H, Fukui T. A review of research on the clinical effectiveness of therapeutic practice guidelines, 1991-2000: the need for standardization of reporting style. *General Medicine* 2002; 1: 1-7.
- 14) Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA*. 1999; 281(20): 1900-1905.
- 15) 福井次矢, 中山健夫. 日本国内で作成された診療ガイドラインの質評価の試み. 厚生労働科学. EBMを指向した「診療ガイドライン」と医学データベースに利用される「構造化抄録」作成の方法論の開発とそれらの受容性に関する研究 (主任研究者・中山健夫). 2001年度報告書. p. 41-45.
- 16) Appraisal of Guidelines, Research, and Evaluation in Europe (AGREE) Collaborative Group. Guideline development in Europe. An international comparison. *Int J Technol Assess Health Care* 2000; 16(4): 1039-1049.
- 17) 長谷川友紀. 診療ガイドラインを取り巻く状況: AGREE Collaborationの動向. EBMジャーナル. 2003; 4(3): 294-297.
- 18) Shiffman RN, Shekelle P, Overhage JM, et al. Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. *Ann Intern Med* 2003; 139(6): 493-498.
- 19) 中山健夫. EBMを用いたガイドラインの作成と応用: EBMの考え方と問題点を踏まえて. *Progress in Medicine*. 2003; 23(12): 3143-3151.
- 20) Lenzer J. Alteplase for stroke: money and optimistic claims buttress the “brain attack” campaign. *BMJ* 2002; 324: 723-729.
- 21) Choudhry NK, Stelfox HT, Detsky AS. Relationships between authors of clinical practice guidelines and the pharmaceutical industry. *JAMA* 2002; 287(5): 612-617.
- 22) 鈴木博道. NICE ガイドライン開発への患者・介護者の参画. *あいみっく* 2004; 25: 10-14.
- 23) 宮本昭正監修. EBMに基づいた患者と医療スタッフのパートナーシップのための喘息診療ガイドライン 2004 (小児編). 東京. 協和企画. 2004年.
- 24) 宮本昭正監修. EBMに基づいた患者と医療スタッフのパートナーシップのための喘息診療ガイドライン 2004 (成人編). 東京. 協和企画. 2004年.
- 25) 「ぜんそく診療: 患者参加で指針作り, 分かりやすく図解」. 朝日新聞. 平成16年6月13日.
- 26) 稲葉一人. インフォームド・コンセントを充実させるためのガイドライン. 厚生労働科学. EBMを指向した「診療ガイドライン」と医学データベースに利用される「構造化抄録」作成の方法論の開発とそれらの受容性に関する研究 (主任研究者・中山健夫). 2002年度報告書. p. 50-58.
- 27) Haynes RB, Devereaux PJ, Guyatt GH. Clinical expertise in the era of evidence-based medicine and patient choice. *ACP J Club* 2002 Mar-Apr; 136(2): A11-14.
- 28) Haynes RB, Devereaux PJ, Guyatt GH. Physicians' and patients' choices in evidence based practice. *BMJ* 2002; 324(7350): 1350.
- 29) 砂原茂一. 抗炎症薬に関する有効報告: RCT と historical controlled study の比較. *日本内科学会雑誌*. 1970; 59: 588.
- 30) Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med* 1982 Feb; 72(2): 233-240.
- 31) Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000; 342(25): 1878-1886.
- 32) Concato J, Shah N, Horwitz RI. Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs. *N Engl J Med* 2000; 342(25): 1887-1892.

- 33) Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001; 286: 821–830.
- 34) Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? *N Engl J Med* 2000; 342(25): 1907–1909.
- 35) Fukui T, Rahman M. Contribution of research in basic and clinical sciences in Japan. *Intern Med* 2002; 41(8): 626–628.
- 36) Nakayama T, Yamazaki S. Percentages of reports of clinical trials, written in seven non-English languages, that have structured abstracts. *General Medicine* (in press).
- 37) Nakayama T, Fukui T, Fukuhara S, et al. Comparison between impact factors and citations in evidence-based practice guidelines. *JAMA* 2003; 290(6): 755–756.
- 38) Atkins D, Best D, Briss PA, et al. GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328(7454): 1490.
- 39) 特集 Minds 入門：インターネットによる診療ガイドライン活用方法. *医事新報*. 2004; 4184: 1–15.

ABSTRACT

Clinical Practice Guidelines Developed According to the Principles and Methodologies of Evidence-Based Medicine (EBM): Real Movements, Problems and Future Prospects in Japan

Takeo NAKAYAMA

Department of Health Informatics, Kyoto University School of Public Health

Evidence-based approach to the development of clinical practice guidelines has attracted significant attention from medical groups as well as from the public. According to the U.S. Institute of Medicine, clinical practice guidelines are defined as the systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances. However, the role and importance of clinical practice guidelines are not fully understood by stakeholders in Japan. This review aims to introduce real movements related with clinical practice guidelines in Japan, problems and future prospects of them.

Key words: clinical practice guidelines, evidence-based medicine, grade of recommendation, stakeholders, professional autonomy

TOPICS

エビデンスの質とお勧め度のグレーディング

Grading Quality of Evidence and Strength of Recommendations

GRADE Working Group

(BMJ 2004; 328(7454): 1490-7.)

訳：津谷喜一郎¹⁾ 中山 健夫²⁾ 島村 治子¹⁾

臨床ガイドラインの質は、そのベースとなったエビデンスと用いられた「判断」(judgment)によって決まってくる。GRADEのアプローチは、ガイドラインのユーザーが、ガイドラインのお勧め度の背後にある「判断」の評価を容易にすることを目的とする。

要旨 (Summary)

臨床ガイドラインやその他種々の「お勧め」(recommendation)を利用するユーザーは、提示されているお勧めがどれだけ信頼をおけるものなのかを知る必要がある。システムティックで明示的な手法を用いれば、判断(judgment)の誤りは少なくなり、コミュニケーションも円滑になる。そこでわれわれは、エビデンスの質およびお勧め度を評価するシステムを開発した。これは、種々の介入や状況に適応できる網羅的なシステムである。ここでは、ガイドラインユーザーの視点から、われわれの提唱するアプローチについて概説する。

お勧め度(strength of recommendation)に関する判断には、利益と害のバランス、エビデンスの質(quality of evidence)、状況に対応したエビデンスの適用、およびベースラインリスクに関する確実性(certainty)についての検討が必要である。また、コスト(リソースの活用)についての考慮も重要であ

る。エビデンスの質とお勧め度を評価するシステムに整合性がなければ、批判的吟味や情報のコミュニケーションを促すというその潜在的機能が損なわれてしまう。われわれの提唱するシステムは、このような複雑な判断プロセスを支援するものであり、簡潔さと、すべての重要事項に関する網羅的で理解しやすい考察へのニーズとをバランスよく取り込んだシステムとなっている。

はじめに (Introduction)

エビデンスやお勧めの判断においては、さまざまな事項を検討しなければならない。たとえば、中等度うつ病の治療薬として、選択的セロトニン再取り込み阻害薬(SSRI)と三環系抗うつ薬の選択について考えてみよう。臨床医は、どのアウトカムを考慮し、それぞれのアウトカムについてどのエビデンスを取り入れ、各エビデンスの質をどのように評価し、三環系抗うつ薬と比較した場合にSSRIには害より利益が

¹⁾ 東京大学大学院薬学系研究科医薬経済学

²⁾ 京都大学大学院医学研究科社会健康医学専攻系医療システム情報学

This article is reproduced with permission from the BMJ Publishing Group.

多いといえるかどうかについて判断しなければならない。また、リソースは限られており、一度 SSRI に支出された費用はもう戻ってこないのに、健康上得られる利益の増分が追加的コストに見合うかどうかについても判断する必要がある。

しかし、個々の臨床医や患者が何の支援もなく臨床上的判断を行うことは現実的ではない。通常、臨床医や患者は、判断のための資料として診療ガイドライン (clinical practice guidelines) を用いる。ここに記載されているお勧めは、利用可能なエビデンスにアクセスでき、臨床上的問題や研究手法を熟知し、お勧め事項の考案に十分な時間を割くことのできる専門家集団によって作成されている。

システムティックに作成されたガイドラインのユーザーは、そこに提示されているエビデンスやお勧めにどの程度の確信 (confidence) をおいたらよいかを知る必要がある。そこでわれわれは、ガイドラインの確信の基礎となる要素、また診療ガイドラインや医療に関わる種々のお勧めのなかに明示的 (explicitly) あるいは黙示的 (implicitly) に取り入れられている複雑な判断をシステムティックに行うアプローチについて解説する。説明を簡潔にするために、すべてのニュアンスを議論したり、このアプローチを適用するうえでガイドライン作成班が必要とするような詳細なガイダンスについてはここでは取り上げない。このような情報は著者から入手されたい。

www.GradeWorking-Group.org

システムティックで明示的な判断手法を用いてエビデンスの質を評価することにより、誤りを防ぎ、判断内容についての批判的吟味を促し、情報のコミュニケーションを向上させることができる。1970年代以来、多くの組織がさまざまなシステムを用いてエビデンスの質 (レベル) (quality of evidence (level)) やお勧め度 (strength of recommendation) をグレーディングしてきた^{1~28)}。しかし残念ながら、組織によって用いるシステムは異なっている。同じエビデンスやお勧めでも、システムによって「II-2, B」, 「C+, 1」, 「有力なエビデンス」 (strong evidence), 「強く勧められる」 (strongly recommended), などさまざまなグレーディングがなされている。これは混乱した状況であり、有効な情報のコミュニケーションを妨げている。

GRADE ワーキンググループは、既存のグレーディングシステムの問題点についての対策を考える非公式の共同グループとして発足した。表1は、上述の問題点、およびGRADEシステムによるその解決策をまとめたものである。GRADEシステムでは、従来のシステムと比較してより一貫性のある (consistent) 判断が可能となり、そのような判断のコミュニケーションにより、ヘルスケアにおいて、よりよいインフォームド・チョイス (informed choice) を支持することができる。Box 1 は、問題 (problem) の優先順位付け (prioritising) から実施の評価に至るまで、ガイドラインの作成と実施に関わるステップを示したものである。ここでは特にエビデンスの質とお勧め度のグレーディングに重点を置いている。

定義 (Definitions)

エビデンスの質とお勧め度について以下のように定義した。エビデンスの質とは、ある推定効果 (estimate of effect) が正しい (correct) かについて、どの程度の確信 (confidence) をおけるかを示すものである。またお勧め度とは、お勧めにしたがうことによる利益が害を上回ることにについて、どの程度の確信をおけるかを示すものである。

エビデンスの質の判断には、重要なアウトカムについて、個々の研究結果の妥当性を評価する必要がある^{26, 29~32)}。われわれの提唱するアプローチでは、このような判断プロセスに沿って、以下の事項について逐次的に判断する。

- 個々の重要なアウトカムについての、研究全般におけるエビデンスの質
- 意思決定に関わる重大なアウトカムの特定
- 重大なアウトカムに関する全体的なエビデンスの質
- 利益と害のバランス
- お勧め度

これらの事項についてどれだけ正確な判断が下せるかは、臨床上的「クエスチョン」がどれだけ明確に定義されているか、影響を受ける人々にとって重要であると考えられるアウトカムがすべて考慮されているかどうかにかかっている。「クエスチョン」で

表1 GRADE システムと他のシステムの比較

要 因	他のシステム	GRADE	GRADEシステムを用いることのメリット*
定 義	エビデンスの質(レベル)とお勧め度に関する黙示的(implicit)な定義	明示的(explicit)な定義	各グレードの意味およびグレード決定に考慮すべき事項が明確にされている
判 断	重要なアウトカム,各アウトカムのエビデンスの質,全体的なエビデンスの質,利益と害のバランス,および健康上得られる利益増分の価値についての黙示的判断	逐次的(sequential)な明示的判断	判断プロセスが明確であり,黙示的な判断プロセスにありがちなエラーやバイアスのリスクを減らすことができる
エビデンスの質を決める「鍵」となる要因	アウトカムについてエビデンスの質が個別に評価されていない。エビデンスの質が研究デザインのみで判断されている場合が多い	研究デザイン,研究の質,一貫性,直接性,についてシステムティックかつ明示的に検討して,エビデンスの質を判断している	左記の要因が適切に検討されるように配慮されている
エビデンスの質に影響するその他の要因	明示的には考慮されていない	不正確もしくはまばらなデータ,報告バイアス,関連の強さ,用量反応曲線,ありそうな交絡因子についても明示的に考慮	その他の要因についても考慮されるように配慮されている
全体的なエビデンスの質	利益に関するエビデンスの質に基づいて黙示的に判断されている	意思決定に重大な影響力をもつアウトカムに関してはもっとも質の低いエビデンスも踏まえて判断されている	重大なアウトカムに関するエビデンスが欠如している場合でも全体的なエビデンスの質について誤った評価がされにくいように配慮されている
アウトカムの相対的重要性	黙示的に考慮されている	重大な(critical)アウトカム,重大とまではいかないが重要な(important)アウトカム,重要性がなく無視できる(ignored)アウトカム,について明示的に判断されている	全体的なエビデンスの質とお勧め度を評価する際に,個々のアウトカムが適切に考慮されるよう配慮されている
健康上の利益と害のバランス	明示的に考慮されていない	重要な利益と害の得失,それらのエビデンスの質,特定の状況におけるエビデンスの適用,ベースラインリスクの確実度について明示的に考慮されている	利益と害についてより明白な判断が下せるように配慮されている
健康上得られる利益増分がコストに見合うかどうか	明示的に考慮されていない	正味の(net)健康上の利益について考慮された後に,明示的に考慮されている	正味の健康上の利益について明白な判断が下せるように配慮されている
エビデンスと研究結果の要旨	記載に一貫性がない	質の評価と研究結果の要旨を含む,GRADE提唱の一貫したエビデンス・プロフィール	すべての研究班メンバーが同じ情報に基づいて判断を下し,またその情報が他者にもアクセスできるものであるよう配慮されている
用 途	複数の機関に活用されることはまれであり,経験的評価もほとんどされていない	幅広い領域の機関が国際的に共同して作成および評価をしている	過去の実施状況に基づいて,システムの実用性,信頼性を高め,適用可能性を広げる

* GRADEシステム以外のアプローチには,概してこれらのメリットが含まれていない。ただし,ここに掲げるメリットのうちいくつかを組み入れられているシステムもある。

Box 1: ガイドライン作成のための逐次的プロセス

はじめのステップ

1. プロセスの確立

問題の優先順位付け, 研究班メンバー選定, 利害衝突の表明, グループプロセスへの合意など

準備的ステップ

2. システマティック・レビュー

最初のステップは, すべての重要なアウトカムについて, 入手可能な最良のエビデンスを特定し批判的に吟味する, もしくはシステマティック・レビューを行う。

3. 重要なアウトカムについてエビデンスのプロフィールをまとめる

サブレビューもしくはリスクグループごとにシステマティック・レビューの結果に基づいたプロフィールが必要であり, そのプロフィールにはエビデンスの質の評価と研究結果の要旨についての記述が含まれていなければならない。

エビデンスの質とお勧め度のグレーディング

4. 各アウトカムに関するエビデンスの質

エビデンス・プロフィールに記載されている情報および表2に示される基準に基づいて判断する。

5. アウトカムの相対的重要性

エビデンスのプロフィールには重要なアウトカムのみが記載されていなければならない。記載に加えられたアウトカムは分類上意思決定において重大もしくは(重大とまではいかないが)重要とされるものでなくてはならない。

6. 全体的なエビデンスの質

アウトカム全般に関する全体的なエビデンスの質は, 重大なアウトカムに関するエビデンスのうち, もっとも質の低いエビデンスにわたって判断するべきである。

7. 利益と害のバランス

利益と害のバランスは, 重要な健康上の利益や害を基準とし, 「正味の(net)利益あり」, 「利害得失あり」, 「不確かな利害得失あり」, 「正味の利益なし」に分類する。

8. 正味の利益とコストとのバランス

健康上得られる利益増分は追加的コストに見合うものといえるか。リソースは常に限られているため, お勧めを作成する際にはコスト(リソースの活用)について検討することが重要となる。

9. お勧め度

お勧めの内容はお勧め度(お勧めにしたがうことによる利益は害を上回る, ということについてどの程度の確信をおけるかを示す)を反映していなくてはならない。

つぎのステップ

10. 実施と評価

変化に対する障害に取り組むための効果的な戦略の実行, 実施状況の評価および継続的な更新など。

は, 比較の対象となっている選択肢(例: SSRIと三環系抗うつ薬), 対象患者(例: 中等度うつ病の成人患者), 想定している状況(例: 英国におけるプライマリ・ケア)が明確になっていなければならない。

**重要なアウトカムについてのエビデンスの質
(Quality of evidence for each important outcome)**

入手可能なエビデンスのシステマティック・レビューがこれらの判断をガイドする。レビューを行う者は, 研究デザイン, 研究の質, 一貫性, 直接性, という4つの要素を考慮しなければならない。

(1) 研究デザイン(Study design)

研究デザインとは基本的な研究のデザインのことであり, 広義の分類での観察研究およびランダム化試験のことである。このことは, 論理的(logical)かつ実証的(empirical)なエビデンスがこれを支持している³³⁻³⁶。

一般に, 観察研究から得られる結果はランダム化試験のそれと類似しているが, 必ずしもいつも同じ結果が得られるとは限らない。その顕著な例として, ホルモン補充療法による冠動脈性心疾患のリスクについて, 観察研究ではリスク低減効果が認められたが, 続いて実施されたランダム化試験ではリスク低下が認められず, かえってリスクが増大するという結果が得られた^{37, 38}。残念ながら, 観察研究と同様の結果がランダム化試験からも得られるかどうかは予測できない。お勧め内容の考案において質の高いランダム化試験からの結果が得られている場合には, 非ランダム化試験から得られた異なる結果に固執する者はほとんどいないだろう。

その一方で, ランダム化試験がいつも実施できるわけではなく, 観察研究から得られるエビデンスの方が優れている場合もある。まれな有害作用(adverse effect)の事例などがこれに当たる。さらに, ランダム化試験から得られた結果が適用できない場合もある。たとえば, 現実に結果を適用する集団に比べ, 試験参加者が厳選された意識の高い集団である場合である。それゆえ研究デザインの適切性(appropriateness)とともに, 研究の質(quality), 研究結果

の一貫性(consistency), エビデンスの直接性(directness)の考慮が必須である。たとえば, 手術中の死亡や大腸内視鏡施行時の穿孔などのような, 治療に関連した合併症の発生率のエビデンスは, デザインの良いケースシリーズ(case series)から得られたものの方が, ランダム化試験の結果よりも直接的な関連性があると考えられる。同様に, 偽陽性のスクリーニング結果によって生じたりコール率や処置(マンモグラフィ(乳房のX線検査)後の生検実施率など)に関しては, コホート研究から質の高いエビデンスが得られる。

(2) 研究の質(Study quality)

研究の質とは, 詳細な研究手法とその実行についてである。レビュアーは, 個々の重要なアウトカムについて, 適切な基準を用いて研究の質を評価しなくてはならない^{26, 29-32}。ランダム化試験についていえば, 割り付けの隠蔽(concealment), ブラインド化, フォローアップ, が適切であるかどうかを基準とする。また, 質の評価を下げる場合には, レビュアーはその理由を明確にする必要がある。たとえば, 痛みに対するある介入の効果について, 患者および医師のブラインド化が不十分であればそのエビデンスは低く評価され, このことは重大な限界であるとみなされる, という説明がなされるだろう。

(3) 一貫性(Consistency)

一貫性とは, 一連の研究から得られる効果の推定結果(estimate of effect)が類似していることを意味する。もし結果に説明のつかない重大な矛盾点が存在すれば, そのアウトカムに関する効果推定の確信度は低くなる。効果の方向性の違いや効果の違いの程度, 違いの有意性は重要な矛盾点が存在するかどうかの決定を導く(若干恣意的となるのは免れないが)。矛盾に対して納得できる説明がなされる場合には, サブグループごとに効果の大きさを推定すべきである。たとえば, 頸動脈内膜切除術の効果が狭窄の重症度によって異なる場合には, サブグループごとに効果推定をすべきである。

(4) 直接性(Directness)

直接性とは, 研究の試験参加者, 介入, アウトカ

ム指標が, ガイドラインを適用する状況にどれだけ類似しているかである。たとえば, 対象患者の年齢や重症度が参加者よりも高かったり多くの合併症がある場合には, エビデンスの直接性は不確かなものとなるだろう³⁹。重要な不確実性があるかどうかを決定するには, まず推定効果に重大な差異を生じさせる妥当な理由があるかどうかを考えてみるとよい。多くの介入は, ほとんどの患者群で似たような相対的効果(relative effect)をもつため, エビデンスの直接性を判断する際に厳しすぎる基準を設定してはならない。ただし, 行動介入のように文化的な違いが大きな意味合いを持つような療法においては, 厳しい基準の設定が適切な場合もある。

また, 勧められる医薬品が, 同じクラスに属してはいるが研究に用いられた医薬品とは異なるというような場合, そのエビデンスの直接性についてレビュアーから不確かであるという指摘を受けるかもしれない。他の介入でも同様の問題が生じうる。たとえば, 研究のように集中的でないカウンセリング介入や, 研究では用いられていない代替的な外科治療法に, 研究結果を一般化することはできるだろうか。このような判断を下すのは非常に困難となりうる⁴⁰。そのため, 研究者は自らが導いた結論についての論拠(rationale)を説明する必要がある。

その一方で, 一般的に, 代理のアウトカム(surrogate outcome)を用いる研究は, 人々にとって重要なアウトカムを用いる研究と比較すると, エビデンスの直接性が十分でない。そのため, 代理のアウトカムについてのエビデンスの直接性を検討する場合は, より厳しい基準を設けるのが賢明である。代理のアウトカムに基づいた間接的エビデンスが, 後の試験で誤りであったことが明らかになった例がいくつかある。心筋梗塞患者に対して死亡の代理として不整脈の抑制を用いた研究⁴¹, 冠動脈性心疾患の代理としてリポ蛋白の変化を用いた研究³⁷, 閉経後の女性に対して骨折の減少の代理として骨密度を用いた研究⁴², などである。

適格な(appropriate)治療によるヘルスアウトカムの向上や, 偽陽性の結果から生じる害の軽減などのような, 正確な診断によって左右される可能性のある重要なアウトカムの代理のアウトカムとしては, 診断検査の正確さを挙げるができる。診断の正

Box 2: エビデンスのグレードを決めるための基準

エビデンスの種類

- ランダム化試験 高
- 観察研究 低
- その他のエビデンス ... 非常に低

以下の場合にはグレードを下げる

- ・研究の質に関わる深刻 (serious) (-1), もしくは非常に深刻 (-2) な限界
- ・重要な矛盾 (-1)
- ・直接性 (directness) に関する若干の (-1) もしくは重大な (-2) 不確実性
- ・不精確 (imprecise) もしくはまばらなデータ (-1)
- ・報告バイアスが存在する可能性が高い (-1)

以下の場合にはグレードを上げる

- ・関連性 (association) を示す有力なエビデンス。有意な相対リスク >2 (<0.5)。2件以上の研究から得られた一貫性のある (consistent) エビデンスに基づいており、ありそうな交絡因子がない (+1)⁴⁶⁾
- ・関連性を示す非常に有力なエビデンス。有意な相対リスク >5 (<0.2)。直接的なエビデンスに基づいており、妥当性を脅かすような深刻な状況がない (+1)
- ・用量反応勾配のエビデンス (+1)
- ・すべてのありそうな交絡因子が実際に作用したとすれば効果を減少させる方向にはたらいたと考えられる (+1)

確さに関する研究デザインを検討する際には、通常のものとは異なる基準を適用しなければならない。その一方で、エビデンスの直接性は、正しく分類されること (真の陽性または陰性) もしくは誤って分類されること (偽陽性または偽陰性) が、それによって生じた重大な結果とどれほどの関連性を有するといえるのかに基づいて検討される。たとえば、急性尿路結石の疑いのある患者の診断においては、経静脈性尿路造影法よりも造影剤不使用のヘリカルCTの方が、偽陰性の結果が少ないということを示すエビデンスが数々の優れたデザインの研究から一貫して得られている⁴³⁾。しかし、このエビデンスが健康への重要な影響を意味するかどうかについては大きな不確実性がある⁴⁴⁾。そのため、お勧めの作成においてはこのエビデンスの質は低いとみなされる。

また、介入を直接比較することができず、異なる研究間での比較が必要な場合にも間接的エビデンス

の問題が生じる。たとえば、SSRIとプラセボ薬、そして三環系抗うつ薬とプラセボ薬を比較するRCTはあっても、SSRIと三環系抗うつ薬を直接比較した試験がない場合である。研究が違えばさまざまな状況が異なっており、それらが研究結果に影響を及ぼすことから、間接比較は直接比較よりも常に不確実性をはらんでいる⁴⁵⁾。

4つの要素の結合 (Combining the Four Components)

個々の主要なアウトカムにおけるエビデンスの質は、上記の4つの要素、すなわち研究デザイン、研究の質、一貫性、直接性、をそれぞれ考慮した後に決定される。われわれの提唱するアプローチでは、まず研究デザインをRCTと観察研究 (コホート研究、ケースコントロール研究、断続的な時系列研究 interrupted time series analysis, コントロールされた前後比較研究) に分類する。つぎに、それらの研究に深刻な限界があるか、結果に重要な矛盾があるか、エビデンスの直接性に関する不確実性が十分に説明されているかを考慮するとよい (Box 2 参照)。エビデンスの質のグレーディングは以下のように定める。

「エビデンスの質」のグレード

- 高 (High)** 推定効果への確信は、さらなる研究を重ねても変わることはない。
- 中 (Moderate)** ... さらなる研究が推定効果への確信に重要なインパクトを持つ可能性があり (likely), その結果、推定効果が変わるかもしれない (may)。
- 低 (Low)** さらなる研究が推定効果への確信に重要なインパクトを持つ可能性が高く (very likely), その結果、推定効果が変わる可能性がある (likely)。
- 非常に低 (Very low)** 推定効果は非常に不確かである。

研究の質における限界、結果における重要な矛盾、あるいはエビデンスの直接性に関する不確実性は、エビデンスのレベルを下げる要因となる。たとえば、利用可能なすべての研究に深刻な限界がある場合、グレードは一段階下がり、その限界が非常に深刻であれば、グレードは二段階下がる。致命的な欠陥のある研究は除外されるかもしれない。

その他にも、データが不精確 (imprecise) もしくは

Box 3: 不精確もしくはまばらなデータ

不精確(imprecise)もしくはまばらな(sparse)データを定義するための経験的な基盤はないが、以下の2つに定義されうる。

- ・データはまばら(sparse)であるとは、研究結果に示される事象(event)もしくは所見(observation)の数が少なく、それらが情報を提供しない(uninformative)場合。
- ・データは不精確(imprecise)であるとは、評価結果が重要な害もしくは重要な利益のいずれかに合致するものではあるが、その信頼区間の幅が広い場合。

これらの異なった定義のどちらを採用するかによって導かれる判断も異なってくる。これらの定義を調和するのは可能ではないかもしれないが、データの不精確さもしくはまばらさによりエビデンスの質のグレードを考慮する場合には、つぎのガイダンスを示すこととする。

- ・研究が1件しかない場合は、データの不精確さもしくはまばらさを判断する閾値を低く設定するべきである。サンプルサイズが小さい(あるいは事象が少ない)単一の研究において、信頼区間の幅が広く、害と利益の双方に可能性が及んでいる場合、そのようなデータは不精確もしくはまばらであるとみなされるべきである。
- ・信頼区間の幅が広いために、その他のアウトカムとは無関係に、内容的に矛盾するお勧めに合致してしまう場合には、データは不精確もしくはまばらであるとみなされるべきである。

まばらな(sparse)場合(**Box 3** 参照)や報告バイアスのリスクが高い場合などもエビデンスの質が下がる。

エビデンスの質を上げると考えられる要因としては次のような事項が挙げられる。

- ・非常に強い関連性 (very strong association, 例: SSRI と比較して、三環系抗うつ薬の中毒による死亡リスクは50倍である, 表2参照), あるいは強い関連性 (strong association, 例: 二輪車運転時, ヘルメット不使用者の頭部外傷リスクは使用者の3倍である⁴⁷⁾。
- ・用量反応勾配のエビデンス
- ・もっともな (plausible) 説明のできない交絡があり, そのすべてが観察効果を小さくする要因となりえる。(例: 営利病院と非営利病院における死亡

率を比較する研究で, 結果に影響を及ぼすかもしれないもっともな説明因子が調整されていない場合, 観察効果は過小評価されていると考えられる⁴⁸⁾。このとき, 死亡リスクは営利病院の方が高いというエビデンスはより説得力があることになる)。

これらの評価は累積的に行われる。たとえば, RCTに深刻な限界があり, かつエビデンスの直接性も不確かである場合, エビデンスの等級は「高」から「低」まで下がることになる。

害と利益に関するエビデンスの質の判断にも同様のルールが適用されるべきである。ありそうな (plausible) 重要な害については, それがありそうなことを示唆する間接的エビデンスを考慮したうえで, エビデンスサマリーに加えるべきである。たとえば, 黒色腫(melanoma)のスクリーニング検査に関する不安が懸念されているが, 直接的なエビデンスがない場合には, 他のタイプのスクリーニング検査に関する研究のエビデンスを考慮することが適当であろう。

重要なアウトカムに関する複数の研究を通してのエビデンスの質は, コクラン・レビューのようなシステマティック・レビューを用いて判断することが可能であり, そうすべきである。ただし, 全体としてのエビデンスの質, 利害得失(trade-off), お勧めに関する判断を下すには, 概してレビューから得られる結果以上の情報が必要である。

全体的なエビデンスの質 (Overall quality of evidence)

既存のシステムでは, 全体的なエビデンスの質の判断が, 介入から得られる利益に関するエビデンスの質を基準としていることが多い。しかし, 有害作用のリスクが重大な意味を持ち, かつリスクに関するエビデンスが利益のエビデンスよりも弱い場合, 害のリスクに関する不確かさを無視することには問題がある。エビデンスの質がどんなに低くとも, そのアウトカムが意思決定に不可欠な場合には, 全体的なエビデンスの質の等級づけ(rating)の基礎とすべきであると提案する。

表2 プライマリ・ケアにおける中程度のうつ病の治療薬としての選択的セロトニン再取り込み阻害薬 (SSRI) と三環系抗うつ薬の比較試験の質の評価

研究の件数	質の評価 (quality assessment)				
	デザイン (design)	質 (quality)	一貫性 (consistency)	直接性 (directness)	影響を及ぼす その他の要因*
うつ病の重症度 (ハミルトンうつ病評価尺度を用いて 4-12 週間後に測定)					
シタロプラム (8)					
フルオキセチン (38)					
フルボキサミン (25)					
ネファゾドン (2)	RCT	深刻な 限界なし	重要な 矛盾なし	直接性に関する 一定の不確かさ (アウトカム指標) [†]	なし
パロキセチン (18)					
セルトラリン (4)					
ベンラファキシン (4)					
一過性の副作用による治療の中止					
シタロプラム (8)					
フルオキセチン (50)					
フルボキサミン (27)					
ネファゾドン (4)	RCT	深刻な 限界なし	重要な 矛盾なし	直接的	なし
パロキセチン (23)					
セルトラリン (6)					
ベンラファキシン (5)					
中毒による死亡 [§]					
イギリス国家統計局 (1)	観察研究より 得られたデータ	深刻な 限界あり [‡]	研究は 1件のみ	直接的	非常に強い関連性

WMD = weighted mean difference (加重平均差), RRR = relative risk reduction (相対的リスク減少)

* 不精確もしくはまばらなデータ, 強いもしくは非常に強い関連性, 報告バイアスのリスク高, 用量反応勾配のエビデンス, 結果に影響を与える未調整の交絡。

[†] 試験期間が短かったため, アウトカムの直接性が不確かであった。

[‡] 低リスクの人がそうでない人よりも SSRI の投与を受けていた可能性が高く, 抗うつ薬を変えることによって自殺企図を回避できたかどうかは定かではない。

[§] 中毒による死亡に関するベースラインリスクが不確かである。

重大(critical)とまではいかないが重要な(important)アウトカムは, エビデンス・プロフィールに加え, 健康上の利益と害のバランスについての判断に際して考慮すべきである。しかし, これは全体としてのエビデンスの質の評価の考慮には入れてはならない。アウトカムが重大か, 重大ではないが重要か, それとも重要ではないかの決定は, ひとつの価

値判断である。その際には, 提案するお勧めが影響を及ぼす人の価値観をできる限り考慮するべきである。

何を重大とするかという決定は困難なものとなりうる。有害なアウトカムのもっともらしさ(plausibility)は, それが重大であるならば評価に影響を及ぼす。もっともらしさのない推定的な害を示唆する弱