診療ガイドラインおよび Minds における医学文献の評価選定基準について

森 實 敏 夫*

神奈川歯科大学 内科

Morizane T. (Department of Medicine, Kanagawa Dental College, 82 Inaokacho, Yokosuka, Kanagawa 238-8580, Japan) Evaluation criteria for selecting clinical studies for developing clinical practice guidelines and the Medical Information Network Distribution Service (Minds). *Igaku Toshokan* 2004;51(1):51-56

Clinical practice guidelines have been developed by official organizations to help healthcare workers with clinical decision-making. The use of these guidelines is expected to improve patients' outcomes. Critical appraisal of the clinical studies used to form a knowledge base is a crucial step in the process of developing guidelines. The quality of evidence depends on the study design, with randomized controlled trials providing the most reliable evidence. A meta-analysis of randomized controlled trials can reduce the chance of a statistical error and provides better evidence than a single trial. Only randomized controlled trials can provide a definite answer as to whether a specific intervention can improve the outcomes with clinical relevance. Evidence based on non-randomized clinical trials is weaker than that provided by randomized controlled trials. A hierarchical evaluation system according to study design, with meta-analyses at the top and experts' opinions at the bottom, has been used to determine the strength of recommendations. However, even randomized controlled trials are not immune to various biases. A more precise and detailed evaluation method for selecting clinical studies that are meaningful, valid, reliable, and relevant is needed. The Japan Council for Quality Health Care has developed the Medical Information Network Distribution Service (Minds) for Japanese healthcare workers and the public. An evaluation system for this service is currently under development.

key words : Practice Guideline, Medical Information Network Distribution Service, Minds, Critical Appraisal, EBM, Evidence-based Medicine, Study Design

I. 診療ガイドラインとは

診療ガイドラインは通常学会・当局などの公的機関により組織された委員会によって作成される。目的は医師や他の医療ケア供給者の行う診療における意思決定や患者の意思決定を支援することである。たとえば英国の診療ガイドライン作成を行う中心的機関である NICE (National Institute for Clinical Excellence)¹⁾ では「診療ガイドラインは入手可能な最良のエビデンスを基盤として保健医療専門家による個々人のケアに関する推奨の記述である。診療ガイドラインは保健医療専門家の実践を援助はするがその知識や技術の代わりになるものではない。優れた診療ガイドラインは保健医療のプロセスを変えそのアウトカムを改善することができる」と述べている。

作成の過程は 1) 臨床上の問題点 (クリニカルクエ

スション) の特定と現在の診療の明確化 2) 包括的な文献検索 3) 論文の批判的吟味 4) エビデンスレベルの判定 5) クリニカルクエスションごとのまとめの作成 6) 勧告 (または推奨) 作成とその強さの設定である²⁾³⁾。このように診療ガイドラインは Evidence-based medicine (EBM) の手法に則ってこうすべきであるという勧告を作成し集めたものである。

診療ガイドラインの評価法として Shaneyfelt ら⁴⁾ のチェックリストや AGREE の方法⁵⁾ が考案されている。なお AGREE は診療ガイドライン評価の国際的統一基準をめざして 2001 年に The AGREE Collaboration⁶⁾ により作成された。AGREE のチェックリストの中には「ガイドライン作成グループは関連したすべての職業グループからの参加者を含んでいる」「患者の見方や嗜好が調査されている」といった項目が含まれている。わが国ではまだ実現していないがガイドライン作成委員会には患者・介護者の代表やコメディカルの代表、企業の代表も参加が求められることになる。英国の診療ガイドライン作成機関である NICE のガイドライン開発プロセス⁷⁾ の中の診療ガイドラインの基本原則にも同様のこと

*Toshio MORIZANE : 〒238-8580 神奈川県横須賀市稲岡町 82.
Tel.046-825-1500 Fax.046-822-8801
morizane@kdcnet.ac.jp

表 1. NICE の診療ガイドラインの基本原則

診療ガイドラインは：

- ・ 臨床ケアの質の改善を目的とすべきである
- ・ 治療法や管理法の臨床的有効性に取り組むべきである（すなわちそれらがどれだけ有効か）
- ・ 治療法や管理法の費用対効果に取り組むべきである（すなわちそれらがコストに比してどれだけ有効か）
- ・ 助言的であるべきである
- ・ 当該ガイドラインの影響を受ける可能性のあるすべての人を考慮に入れたプロセスによって開発すべきである（通常は保健医療専門家、患者とその介護者、保健医療サービス管理者、より広範な一般市民、政府、保健医療産業をふくめる）
- ・ 可能な限り最良のエビデンスおよび専門家のコンセンサスを基盤にすべきである
- ・ 患者、National Health Service (NHS)、NHS stakeholders が敬意を払うような方法によって開発すべきである
- ・ 患者および保健医療専門家の参加の下に開発すべきである
- ・ イングランドとウェールズの NHS において合理的に予想しうる臨床ケアを提示すべきである

がうたわれている（表 1）。

II. 研究デザインによる医学文献の評価

ガイドライン作成の過程ではそれぞれの論文の研究デザインに基づいてエビデンスレベルを決めエビデンスレベルに応じて勧告の強さを決める事が行われている。広く用いられているのが表 2-A に示すものあるいはこれに準じたものである。NICE で用いられている分類もこれと同じである。わが国で厚生労働省厚生科学研究費を用いて作成されたさまざまな診療ガイドラインでもほぼ同様の分類が用いられた（表 2-B）。

臨床医学研究の評価はその研究結果が臨床の現場で用いられた場合に患者にベネフィットがもたらされるかどうかによって評価されるべきである。ベネフィットがあるということは患者にとって意味のあるアウトカムが改善することである。患者にとって意味のあるアウトカムとは 1) 症状の改善 2) 健康関連 QOL (Quality of

life 生活の質) の改善 3) 再発の抑制 4) 生存の延長 5) 発症の抑制 6) 感染の防御などである。検査成績など臨床パラメータの改善はその改善がこれらアウトカムの改善と関連していることが証明される必要がある。

多くの場合症例対照研究あるいはコホート研究によって危険因子や病因が明らかにされる。そして病因や危険因子を制御することすなわち介入によって患者にとって意味のあるアウトカムが改善するかどうか証明された時点で患者にベネフィットをもたらす方法が明らかにされる。これを確実に証明する研究デザインはランダム化比較試験である。言い換えると介入の効果を証明するゴールドスタンダードは正しく行なわれたランダム化比較試験である。したがってその結果が患者にベネフィットをもたらす医学研究の主なものランダム化比較試験といっても差し支えない。さらにランダム化比較試験は介入を加える群とそうでない対照群はランダム割り付けによって作成されるので介入以外の点で差のない 2 群を比較することが出来る。したがってアウトカムに統計学的に有意な差が認められた場合にはその差が介入によってもたらされたものであることが確実に証明される。このように正しく行なわれたランダム化比較試験の結果は妥当性が高いということと患者へベネフィットをもたらす方法を確実に証明できるということが上記のエビデンスレベルの決め方の根底にある。しかしランダム化比較試験にともなう主要なバイアスとして 1) 選択バイアス 2) 実行バイアス 3) 症例減少バイアス 4) 検出バイアスが存在しランダム化比較試験でもさまざまな質

表 2-B. 日本の診療ガイドラインで用いられてきたエビデンスレベルの分類

レベル	研究デザイン
I	システマティックレビュー/メタアナリシス
II	1つ以上のランダム化比較試験 (RCT)
III	非ランダム化比較試験 (CCT)
IV	分析疫学的研究 (コホート研究や症例対照研究)
V	記述的研究 (症例報告やケースシリーズ)
VI	患者データに基づかない専門委員会や専門家個人の意見

表 2-A. 研究デザインによるエビデンスレベルの分類*

レベル	研究デザイン
Ia	複数のランダム化比較試験のメタアナリシスによる
Ib	少なくとも 1つの、ランダム化比較試験による
IIa	少なくとも 1つの、よくデザインされた非ランダム化比較試験による
IIb	少なくとも 1つの、他のタイプがよくデザインされた準実験的研究による
III	比較研究や相関研究、症例対照研究など、よくデザインされた観察的研究による
IV	専門委員会の報告や意見、あるいは権威者の臨床経験

*Agency for Health Care Policy and Research (AHCPR), 現 Agency for Health Quality and Research (AHQR)による分類。コホート研究は III に分類。

さらに複数のランダム化比較試験をメタアナリシスによって統計学的に統合することによってより妥当性の高い結果を得ることが出来る。サンプルサイズが大きくなるため統計学的検出力が高くなりエフェクトサイズが改善さればらつきの小さい効果指標の値が得られる。

症例対照研究は後ろ向き研究であり、ある疾患/病態といったアウトカムを持つ群とそれを持たない群で背景因子をマッチングさせ過去にさかのぼって危険因子への曝露の割合を比較する研究デザインである。コホート研

```

graph TD
    Q1["Q1: 曝露群と非曝露群の比較が行われているか [注1]"] -- No --> NC["No → 非比較研究 [注2]  
(Non-comparative study)"]
    Q1 -- Yes --> Q2["Q2: 曝露とアウトカムが同じ群で同時に測定されているか"]
    Q2 -- Yes --> CS["Yes → 横断研究 [注3]  
(Cross-sectional study)"]
    Q2 -- No --> Q3_1["Q3-1: 2群以上の群を比較しているか"]
    Q3_1 -- No --> Q3_2["Q3-2: 介入の前, 間, 後で複数の測定がなされているか"]
    Q3_1 -- Yes --> Q4_1["Q4-1: 研究者が曝露を割り付けているか"]
    Q3_2 -- No --> BAS["No → 前後比較研究  
(Before-after study)"]
    Q3_2 -- Yes --> TS["Yes → 時系列研究  
(Time-series study)"]
    Q4_1 -- No --> Q4_2["Q4-2: 曝露はランダム割り付けか"]
    Q4_1 -- Yes --> Q4_3["Q4-3: 曝露は群レベルでの割り付けか"]
    Q4_2 -- No --> CCT["No → 非ランダム化試験 (個人または群)  
(Controlled clinical trial, CCT)"]
    Q4_2 -- Yes --> Q4_3
    Q4_3 -- No --> RCT["No → ランダム化比較試験  
(Randomized controlled trial, RCT)"]
    Q4_3 -- Yes --> CRCT["群ランダム化比較試験  
(Cluster randomized controlled trial)"]
    CRCT --> Q5["Q5: 群を規定しているのは"]
    Q5 -- アウトカム --> CCS["症例対照研究  
(Case-control study)"]
    Q5 -- 曝露 --> Q6["Q6: コホートデザインか"]
    Q6 -- No --> NRC["No → 同時対照のある非コホート研究(対照のある時系列研究など)  
(Non-randomized trial with concurrent control, Time-series study with control, etc.)"]
    Q6 -- Yes --> Q7["Q7: 前向きか"]
    Q7 -- No --> RCS["No → 後ろ向きコホート研究  
(Retrospective cohort study)"]
    Q7 -- Yes --> PCS["Yes → 前向きコホート研究  
(Prospective cohort study)"]
  
```

【注3】この横断研究は危険因子とアウトカムとの関連を分析して、因果関係などを明らかにすることを目的とする研究のことである。単に、陽性率などを記述するだけの研究は非比較研究に含める。

究は前向き研究であり危険因子に曝露した群と非曝露群（コホート）を経過観察し疾患/病態といったアウトカムの発生率を比較する研究デザインである。いずれも危険因子への曝露とアウトカムの間に相関が認められれば危険因子が同定されることになる。症例対照研究はすでに過去に起こったことを後で調査することになるので危険因子への曝露の観察が不完全になりやすいことと背景因子をマッチングさせた対照群の設定が難しくエビデンスレベルは前向き研究であるコホート研究よりは低いとされる。これら2つは介入が行なわれることは無いので観察研究とも呼ばれ主に危険因子の同定に用いられるが介入の効果を証明することに用いられることもある。

また診断に関する研究は横断研究として行なわれることが多くある時点でアウトカムと予知因子がほぼ同時に測定され疾患群でその因子の陽性率＝感度が高く対照群でその因子の陰性率＝特異度が高いことを証明する。この場合には疾患群の同定にはその時点での診断のゴールドスタンダードが用いられる。

このように医学文献の評価では研究デザインの同定が必ず必要であるがそのアルゴリズムを図1に示す⁸⁾。ここで解説していないデザインについては文献を参照されたい⁹⁾。

Ⅲ. より詳細な評価法

より詳細なエビデンスレベルの評価法¹⁰⁾も可能である。臨床医学論文を評価するためのスコアシステムやチェックリストは現在まで少なくとも26個考案されている¹¹⁾。これらの評価法は互いに同じチェック項目を含んでいるがそれぞれ独自の項目も含んでおり評価の結果はかならずしも同じにはならない。それぞれの評価法を用いて質が高い論文と低い論文に分類しそれぞれでメタアナリシスを行った結果統合された効果指標の値と分類された論文の質の関係が同じではなく逆になるものも示されている¹²⁾。すなわちこれらの評価法は同じ結果をもたらさない。Delphi法を用いたコンセンサスに基づき作成されたチェックリストとしてVerhagenら¹³⁾のものがある。これは次のようなものでランダム化比較試験に適用されるが臨床疫学的見地の共通項とも言えるであろう。

1) 治療割り付け

1. ランダム化されているか
2. コンシールメントが行なわれているか
- 2) 最も重要な予後因子について群間に差が無い
- 3) 適格例の基準が決められているか
- 4) アウトカムの検査者は目隠しされているか

表 3. 検索式の構成

項目		疾患/病態	予知因子（介入/曝露）	対照	アウトカム	研究デザイン
実際の語句		肝硬変において	β ブロッカー投与	EVL	出血	ランダム化比較試験
文献検索	電子データベース	liver cirrhosis [mh]	adrenergic beta-antagonists [mh]	endoscopy, gastrointestinal [mh] AND (ligation [tw] OR ligate [tw] OR banding [tw] OR band [tw])	bleeding OR rupture	randomized controlled trial [pt]
	マニュアル検索	内科学消化器病学 肝臓病学 内視鏡に関する医学雑誌 学会プロシーディングス書籍	内科学消化器病学 肝臓病学内視鏡に関する医学雑誌 学会プロシーディングス書籍			

*この例で実際に PubMed を検索する場合には次のような検索式を用いることができる。
liver cirrhosis [mh] AND adrenergic beta-antagonists [mh] AND endoscopy, gastrointestinal [mh] AND (ligation [tw] OR ligate [tw] OR banding [tw] OR band [tw]) AND (bleeding OR rupture) AND randomized controlled trial [pt]

- 5) ケアの供給者は目隠しされているか
- 6) 患者は目隠しされているか
- 7) 一次エンドポイントの点評価値とばらつきの指標が示されているか
- 8) 治療企図分析が行われているか

IV. Minds における医学文献評価選定基準

現在(財)日本医療機能評価機構¹⁴⁾では Web 上で検索可能な診療ガイドライン情報の提供とともに医学文献の内容を構造化抄録として Web 上に検索できる形で提供する医療情報サービス事業 Medical Information Network Distribution Service 通称 Minds と呼ばれる事業が進行中である。著者は医学文献評価選定委員会委員長を勤めさせていただいているがここで用いる予定の医学文献選定法について紹介する。従来の医学文献の評価法をさらに掘り下げて客観的で信頼できる評価法を追究しようとしている。なおすべての過程が確定しているわけではないので将来変更される可能性があることをお断りしておく。

文献検索はクリニカルクエスチョンからスタートする。検索式は5つの項目から構成される。1) 疾患/病態 2) 予知因子(介入/危険因子) 3) 対照 4) アウトカム 5) 研究デザインであるが検索結果に応じていずれかの項目を削除したり OR で語句を追加したりしながら柔軟に対応する。(表3)。

検索結果からタイトルとアブストラクトに基づき次の選定基準で選定を行う。

- 1) 言語：英語または日本語。
- 2) 人を対象とした研究。
- 3) 1) 疾患・病態 2) 予知因子(介入曝露) 3) 対照 4) アウトカム 5) 研究デザインの各項目がクリニカルクエスチョンと一致している。
- 4) 必要に応じて委員会が指定した語句を含むあるいは委員会が指定した条件を満たす
- 5) さらに ACP Journal Club¹⁵⁾の採択基準を参考に研究目的によって分類しそれぞれについて以下に述べる基準を満たすものを選択する。
 - 1) 予防治療に関する研究 (Prevention or treatment)
 - A) ランダム割付が行なわれている
 - *なおランダム化比較試験以外でも委員会で適格と判断したものは採用する。
 - 2) 診断に関する研究 (Diagnosis)
 - A) 対象疾患だけでなく対照疾患を含む
 - B) sensitivity, specificity, likelihood ratio, positive rate, positivity, odds ratio などの語句を含む

- 3) 予後に関する研究 (Prognosis)
 - A) コホート研究(なお曝露による群分けをしていない研究も含める)
- 4) 病因に関する研究 (Etiology)
 - A) 曝露と推定されるアウトカムの関係を探求している
 - B) アウトカムに対するリスクがある明確に定義された群で前向きにデータ収集が行なわれている：次の順で望ましい。ランダム化比較試験, 準ランダム化比較試験, 非ランダム化比較試験, 症例ごとにマッチングが行われたか比較する群を生成するのに統計学的調整が行なわれたコホート研究, ネスティッド症例対照研究(なお症例対照研究も研究デザインの分類は行う。まれな疾患や病態がアウトカムの場合には症例対照研究も採用する)
- 5) 医療の質の向上あるいは持続的教育に関する研究 (Quality improvement or continuing education)
 - A) ランダム割付が行なわれている
- 6) 医療保健プログラムあるいは介入の経済に関する研究 (Economics of health care programs or interventions)
 - A) 日本を対象にしている
 - B) 研究対象の経済的な問題は実際の患者における選択肢の比較に基づいている
- 7) システマティックレビューまたはメタアナリシス (Systematic review article or meta-analysis)
 - A) 特に条件を設定せずすべて採用する。

[注]

・臨床予測ガイドに関する研究 (Clinical prediction guides) については Minds abstract の作成は行わない。文献の存在を紹介するのみにとどめる。

・鑑別診断に関する研究 (Differential diagnosis) は診断に関する研究と同様に取り扱う。

さらに Meaningful すなわち意味があるかどうかすなわち Understandable かどうかで評価したうえで全文を取り寄せる。全文から EBM データテーブル構造化抄録 (Minds abstract) を作成し EBM レビュアーと疾患専門レビュアーのチェックを受け最終的に医学文献評価選

定委員会でチェック，確定した上で Web に掲載される。委員会では 1) Validity (妥当性) 2) Reliability (信頼性) 3) Relevance (臨床的意義) の 3 つの項目について半定量的に評価を行う計画である。Meaningfulness (意味) も含めこれら 4 つの概念は互いに関連する部分もあるが独立した概念である。簡単に解説すると次のようになる。

Meaningfulness は意味を持つかどうかという概念である。言い換えると評価者が理解できるかどうか問うことを問う。理解できるかどうかは知識経験能力によって異なるので一定の水準を越えた評価者が評価する必要がある。

Validity はそれが測定しようと意図していることを測定しているかどうかを問う概念である。例えばプラセボ対照ランダム化比較試験で癌患者の生存率が改善することを証明しようとする場合に腫瘍縮小効果があることだけを証明してもそれは意図していることを測定していることにはならないので Validity は低い。

Reliability は再現性を問うものである。同じ研究を繰り返した場合同じ結果が得られた場合 Reliability は高い。

Relevance は臨床との関わり の深さであり臨床的意義のことである。患者にベネフィットがあることを直接説明しているかどうかが問題となる。一方でそのベネフィットの大きさも相対的に評価すべきであり特に介入については従来の介入と比較して相対的にその臨床的意義を評価すべきであろう。ベネフィットの大きさはさまざまな効果指標を用いて表すことが出来るが絶対値だけを問題とするのではなく相対的に評価されるべきである。

V. 最後に

医学論文の評価にはさまざまな困難が付きまとう。論文にすべてが記述されているわけではなく記述されていなくても実行されていたり逆に記述されていても実際には不十分にしか実行されていない場合もありうる。研究デザイン以上に細部の評価をすることは多大な労力と時間を要するので研究デザインに基づいた評価を行ないそれにしがって勧告の強さを決める方法は実用性を考えると受け入れざるを得ない。しかし同じランダム化比較試験であってもさまざまな質のものが存在するのも事実であり今後ここで述べたようなより詳細で精度の高い評

価法が普及する可能性がある。

参考文献

- 1) National Institute for Clinical Excellence[internet]. <http://www.nice.org.uk/> [accessed 2004-2-19]
- 2) Cook DJ, Greengold NL, Ellrodt AG(et al). The relation between systematic reviews and practice guidelines. *Ann Intern Med* 1997;127(3):210-6.
- 3) 福井次矢. EBM に基づいた診療ガイドラインの作り方. *EBM ジャーナル* 2000;1(4):438-43.
- 4) 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):1900-5.
- 5) The AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care* 2003; 12(1):18-23.
- 6) The AGREE Collaboration [internet] <http://www.agreecollaboration.org/>
- 7) NICE ガイドライン作成マニュアル [internet] <http://www.nice.org.uk/pdf/guidelinedevelopmentnccsfinal.pdf> [accessed 2004-2-19]
- 8) Žaza S, Wright-De Aguerro LK, Briss PA, Truman BI, Hopkins DP, Hennessy MH(et al). Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. Task Force on Community Preventive Services. *Am J Prev Med* 2000;18 (1suppl): 44-74.
- 9) 森實敏夫. 臨床医のための EBM アップグレード. 東京: 医学書院;2002.
- 10) Downs SH ,Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998; 52(6):377-84.
- 11) Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, Lau J. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002; 287(22):2973-82.
- 12) Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; 282(11):1054-60.
- 13) Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM(et al). The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol* 1998; 51(12):1235-41.
- 14) (財)日本医療機能評価機構 [internet] <http://jcqhc.or.jp/> [accessed 2004-2-19]
- 15) ACP Journal Club [internet] http://www.acpj.org/shared/purpose_and_procedure.htm [accessed 2004-2-19]