

# The Informed Prescriber



## 正しい治療と薬の情報



医薬品・治療研究会 Aug./Sep., 2007 Vol.22 No.8-9

*Critical Choices in Drugs and Therapeutic Alternatives*

### 提携誌

Adverse Drug Reaction Bulletin (英) / Arznei-telegramm (独)

Der Arzneimittelbrief (独) / Arzneiverordnung in der Praxis (独)

Australian Prescriber (豪) / Drug & Therapeutics Bulletin (英)

Drugs Bulletin(印) / Pharma-Kritik(瑞) / La Revue Prescrire(仏)

Drug Bulletins Review(WHO)

## GRADE システム: エビデンスから推奨へ

GRADE system: Going from evidence to recommendations

相原守夫、Gordon Guyatt, Yngve Falck-Ytter, 浜六郎

The Informed Prescriber(TIP)誌 2007; 22: 91-102

## GRADE\* システム:エビデンスから推奨へ GRADE System: Going from evidence to recommendations

相原守夫(1)\*, Gordon Guyatt(2)\*, Yngve Falck-Ytter(3)\* 浜六郎(4)\*

### 1. はじめに:

カナダのマクマスター大学の研究チームによってはじめて用いられた「根拠に基づく医療: EBM (evidence-based medicine)」という用語が登場してから、15年経過した今、その知名度はゆるぎないものになった。また、1970年代以来、さまざまな組織が、さまざまにエビデンスの質(レベル)と推奨の強さについてのグレーディングシステムを使ってきた<sup>1)</sup>。しかし、残念ながら、エビデンスレベルの質のグレード付け方法や、推奨レベルのグレード付けの方法は、それを作成した組織によって種々に異なっている。同じエビデンスでも、作成組織が異なれば、異なる質のレベルに分類されるため、当然ながら推奨グレードにも違いが出てくる。エビデンスの質レベルや、推奨グレードを表すにも、記号(A B C…), 番号(123, I II III…), その組み合わせ(Ia, Ib, IIa…), シンボルマークなどが用いられる。この状況は効果的な情報伝達を妨げ、混乱させている<sup>2-28)</sup>。EBMの輸入国である日本は、これらの多くを日本語化して用いるためにさらに混乱が増幅し<sup>29)</sup>、その状況は現在も変わっていない<sup>30)</sup>。診療ガイドラインに関しては、ある治療法に関する推奨度がどの程度信用性のあるエビデンスに裏づけられたものかを知る必要がある。しかし、従来のエビデンスのレベルは、ランダム化比較試験であるか、観察研究であるかどうかなど、研究デザインを主体とした分類で決められ、推奨の強さはエビデンスのレベルや数によって判断されるものが多かった。このような判定方法では、たとえそのランダム化比較試験で患者にとって最重要の結果が得られると

の証明がなされていない場合や、結果の信頼性に関わる重大な問題があっても、内的妥当性の評価などが不十分なこともあり本来推奨すべきでない治療法の推奨や、その逆の場合などが生じていた。つまり clinical decision making に際して、本来最も重要な医療の受給者である患者側の因子が十分に考慮されたものではなかったのである。

こうした推奨に関する誤分類を避けるためには、系統的で明示的なアプローチによってエビデンスの質を批判的に吟味する必要があり、医療の真の利益と害のバランスを考慮しなければいけない。また、ある療法の「推奨」に際しては、単純で、明瞭な推奨グレーディングシステムによって、臨床医が必要としている主要情報を伝えるとともに、その根拠となった情報を提供する必要がある。

患者中心の医療においては、エビデンスや推奨のレベルの判定には、上記のようなエビデンスの質の詳細な判定にとどまらず、どのようなエビデンスが考慮に値するか、どのようなアウトカムが重要なのか、介入行為の利益と害・負担バランスがどのようなものか、についても明確にすべきである。その場合、当然のことながら、エビデンスのレベルと推奨の強さは別々のものと考えるべきであ

り<sup>28)</sup>、個々の患者によって、あるいは異なる医療環境においては、推奨内容は同一とは限らない。さらに、患者のもつ合併症、問題とする医療技術以外に利用できる治療技術や検査などによって介入行為の安全度や害反応の予知能力についても自ら異なってくるはずである。したがって、医師の専門性や経験、利用可能技術、患者側因子、エビデンスの質の3要素を統合して判定することが、いわゆる患者中心のEBM, Patient-centered care である。

これら3要素を考慮に入れ、現状における国際的混乱を解消すべく登場したのが、Grading of Recommendations Assessment, Development and Evaluation (GRADE: 推奨度の検討・開発・評価) ワーキンググループによる GRADE システムである<sup>31)</sup>。

GRADE システムでは推奨度を「強い」「弱い」の2段階で表示するが、利用者(患者、臨床医、行政)にとっての意味を表1に示した。

海外では多くの学会や団体が診療ガイドラインの作成に際して、GRADE システムを採用している。その普及状況と比較すると、国内の現状は、いくつかの診療ガイドライン作成に際して、GRADE システムを利用したとされているものの、必ずしもそれらは適切な利用とは言えない。そこで我々は、GRADE システムの適正な理解・普及が必須と思ひ、ここにGRADE システムを紹介するもので

表1 ガイドライン利用者(患者・臨床医・行政)それぞれにとっての「強い推奨」「弱い推奨」の意味(訳註a)

<p><b>強い推奨(用いる場合)</b></p> <ul style="list-style-type: none"> <li>・患者に対して: この状況ではほぼ全員がその推奨に沿った診療を希望し、ほんの一部の人たちが受け入れないだけだろう。患者の価値観や好みに沿う意思決定を行うための型どおりの意思決定支援は必要でないことが多い。</li> <li>・臨床医に対して: ほとんどの患者がその介入を受け入れられるようにすべきである。ガイドラインでこの推奨をしている場合、この推奨が遵守されているかどうかは、その医療機関の機能評価の指標としうる。</li> <li>・行政に対して: ほとんどの状況で施策として採用することができる。訳註b</li> </ul> <p>例: 深部静脈血栓症の患者に対する肺塞栓予防のため抗凝固療法を早期に開始すること。市中肺炎に対する抗生物質の使用。手術後感染予防のための手術直前1回抗生物質使用。</p>
<p><b>弱い推奨(用いる場合)</b></p> <ul style="list-style-type: none"> <li>・患者に対して: この状況では、半数以上の患者は示唆された診療方法を希望するはずだが、その診療方法を希望しない患者も多い。</li> <li>・臨床医に対して: 患者の価値観や選好に沿う意思決定を行うための意思決定支援が有用であろう。エビデンスやエビデンスの要約を自分自身で再検討すること。</li> <li>・行政に対して: 行政的には、多くの利害関係者の参加のもとで、かなりの議論が必要であろう。</li> </ul> <p>例: 運動能力の低下した(上葉に優位の)重症肺気腫患者に対する肺容量減少手術。特発性静脈血栓塞栓症(VTE)の患者に対する継続的(漫然とした)抗凝固療法。</p>

参考: Am J Respir Crit Care Med 2006; 174: 605-614 (Table-1)

訳註 a: 上記はいずれも「推奨する場合」の強弱についてであるが、negative/against の場合には意味が異なる。

訳註 b: 行政施策上の重要度とは異なる。

\*GRADE: Grading of Recommendations Assessment, Development and Evaluation (推奨度の検討・開発・評価)

(1) 相原内科医院 (弘前市青山3-8-2)

(2) Department of Clinical Epidemiology and Biostatistics, HSC-2C12, McMaster University, 1200 Main St West, Hamilton, ON, Canada

(3) Case Western Reserve University, School of Medicine, VA Medical Center Cleveland, 10701 East Blvd. Cleveland, OH, 44106, USA

(4) NPO 法人医薬ビジランスセンター (大阪市天王寺区上汐3-2-17, 902)

\*全著者が、GRADE working group member である

ある。

本システムで多用されるいくつかの用語の意味は、末尾の“表7：用語集”を参考にさせていただきたい。

2. GRADE ワーキンググループの発足とその理念・基本構造

GRADE ワーキンググループは、医療技術の推奨評価において用いられてきた既存のグレーディングシステムの短所を改善したいと願う人々により、非公式の共同体として2000年に発足した。GRADEシステムは、このワーキンググループにより2004年にはじめて発表されたものであるが<sup>31)</sup>、これは、ワーキンググループの2年間にわたる反復した検討の成果によるものであった。すなわち、従来のエビデンスや推奨の判定としての代表的な6システム(ACCP, NHMRC, OCEBM, SIGN, USPSTF, UTTFCPS)を使って、12項目についての評価を比較したが、その結果は非常に一致度が低く、従来システムの欠点を指摘した<sup>42)</sup>。また、同グループは、おのおの独立した17人で、12件のエビデンス・プロフィールに関して、GRADEシステムのパイロット研究をおこなった。アウトカムの重要性に関する評価の一致度は高かったが、limitation, consistency, impreciseなどの項目の評価については、かなりばらつきがあった( $\kappa$ 係数=0~0.82)。しかし、不都合の多くはディスカッションで用意に解決するもので、明瞭、理解度の高さ、分別能から、わずかな改変をおこなうことで解消するとして、改変GRADEシステムを作成した<sup>43)</sup>。

この時点では、研究デザインの分類に関するエビデンスの質として、現在にはない“中(moderate)”というカテゴリーがあった(Quasi-randomized trialが該当)。また、グレードダウンの因子なども当然ながら現在とはやや異なったものであった。このような長期の反復検討を繰り返し、単純・明瞭化したGRADEシステムが誕生したわけである。その後、急速な勢いで、海外の多くの学会やグループのガイドライン作成に採用されてきて、現在も普及している<sup>32-50)</sup>。

最近、GRADE ワーキンググループは、GRADEシステムについて国際的なレベルでの理解度を深めるため、5シリーズの発表を行った<sup>48-52)</sup>。第一部：GRADEシステムの全体像、第二部：ガイドライン作成者のための推奨決定とエビデンスの質、第三部：推奨度の解釈、

表 2-a 研究デザインからみたエビデンスレベル

ランダム化比較試験 (RCT)	= 高
観察研究	= 低
その他(症例報告あるいはケースシリーズ)	= 非常に低

表 2-b エビデンスの質のグレーディング基準

<p><b>グレード・ダウン</b> 5項目</p> <p>RCTに限らず、良質な観察研究にも適用できる。その場合、エビデンスの質は、“低”から“非常に低”にグレードが低下する。</p>	<ul style="list-style-type: none"> <li>● RCT sの方法にかかわる欠点(limitations): RCTのデザインおよび実施結果に“深刻”な欠点がある(-1)*, “非常に深刻”な欠点がある(-2)**.</li> <li>● 結果が不一致(inconsistency)または異質性(heterogeneity)がある(サブ解析の問題も含め): (-1)*</li> <li>● 患者への適用上、エビデンスが間接的(indirect): 対象集団や介入方法、コントロール、アウトカムの直接性が、乏しい(-1)*, あるいは、極めて乏しい(-2)**</li> <li>● データが不精確(imprecise)、あるいは、まばら(sparse)である。(-1)*</li> <li>● 報告バイアスが存在する可能性が高い。(-1)*</li> </ul>
<p><b>グレード・アップ</b> 3項目</p> <p>良質な観察研究の出発点のエビデンスの質は“低”であるが、方法の適切さと結果によっては、“中”あるいは“高”の質までにグレードアップされることもありうる。</p>	<ul style="list-style-type: none"> <li>● 強い関連性(association)、あるいは大きな効果・影響(magnitude of effect)が示される(2つ以上の交絡因子の可能性のない複数の観察研究が一致している)。すなわち、相対リスク(RR)&gt;2 あるいはRR&lt;0.5。(+1)*; 妥当性が脅かされるおそれがなく、極めて強い関連性(もしくは効果・影響の程度が極めて大きい)がある。すなわち、RR&gt;5 あるいはRR&lt;0.2。(+2)**</li> <li>● 用量-反応関係(勾配)がある。(+1)*</li> <li>● 交絡因子の可能性があるが、そのすべてが結果に対して、効果・影響を減少させる方向に働いたと考えられる場合。(+1)*</li> </ul>

カッコ内の数値は、エビデンスの質の変化を示している  
 (-1)(+1)\* = 1段階グレードがアップダウン(例: 高から中へ、あるいは、低から中へ)  
 (-2)(+2)\*\* = 2段階グレードがアップダウン(例: 高から低へ、あるいは、低から高へ)

表 2-c 最終的なエビデンスの質と、その意味

高=さらなる研究を実施しても、推定効果への確信は変わることはほとんどない(very unlikely).
中=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)、結果としてその推定が変わるかもしれない(may).
低=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(very likely)、結果としてその推定が変わる可能性がある(likely).
非常に低=どの推定効果も非常に不確かなものである(very uncertain).

第4部：診断検査と戦略におけるGRADE、第5部：医療資源とGRADE、(仮称)である。

GRADEシステムでは、種々のレベルのアウトカム(エンドポイント)ごとに、“利益”と、“downsides(害, 負担, コスト)”, などの関連情報を収集して、エビデンスの“質”の評価と推奨度を決定する。

“エビデンスの質”とは、“推定結果に対する確信(confidence)がもてる程度”と定義されている。具体的手続きとしては、研究デザインによる、高、低、最低の3分類(表2-a)を出発点として、研究デザインの厳密さ、実施された方法や、結果のデータによって、ランクを昇降させ(表2-b)、最終的には4分類(高、中、低、最低)に評価分類する(表2-c)。したがって、研究デザインではランダム化比較試験で高いレベルに分類されて

も、最低ランクに分類されることがあり、症例対照研究のような低いレベルでも最終的には高いレベルに分類されることがありうる。つまり、エビデンスの質の判定に種々の因子を取り入れ、研究デザインとエビデンスの質が分離されているのである。

また、“推奨の強さ(推奨度)”は後述するように、“利益”と“downsides(害, 負担, コスト)”とのバランスの程度を考慮して決定し、推奨度のレベルは、“強い”と“弱い”の2種類である(表1)。ただし、推奨する方向(positive)と、推奨しない方向(negative/against)別に、2種類のレベルの推奨がある。すなわち、GRADEシステムは中間層の推奨基準を排除したものであり、日本のガイドラインの多くで採用されてきた、A, B, C, D分類の「C: 行うよう勧めるだけの根拠が明確でない」、さらに“C”を、

表3 GRADE エビデンス・プロフィールの例



CQ: 膵臓癌患者で、幽門残存膵十二指腸切除術(PPPD)を標準 Whipple 膵十二指腸切除と比較する患者あるいは集団: 手術可能な膵臓癌あるいは乳頭部癌  
 セッティング: 入院患者  
 研究デザイン: ランダム化比較試験

論文: Karanicolas PJ, et al: Systematic review and meta-analysis of pylorus-preserving versus standard Whipple pancreaticoduodenectomy for pancreatic or periampullary cancer. Annals of Surgical Oncology 2007;14:1825-1834

アウトカム	質の評価(Quality assessment)						研究結果のまとめ (Summary of findings)			
	研究数 参加数	研究の欠点	一致性	直接性	精確性	報告バイ アス	RR <sup>1</sup> (95% CI)	最良平均値 (SWPD 群)	ARR <sup>2</sup> (95% CI)	エビデ ンスの質
5年死亡率	3 229	深刻な欠点 (-1) <sup>3</sup>	重要な問題 なし	問題 なし	精確	ありそうで ない	RR 0.98 (0.87 - 1.11)	82.50%	-0.02 -0.12 ~ 0.08	+++ 中
周術期 死亡率	6 490	深刻な欠点 (-1) <sup>3</sup>	重要な問題 なし	問題 なし	不精確 (-1) <sup>4</sup>	ありそうで ない	RR 0.40 (0.14 - 1.13)	4.90%	-0.02 -0.05 ~ 0.01	++ 低
輸血 (単位)	5 320	深刻な欠点 (-1) <sup>3</sup>	重要な問題 なし	問題 なし	不精確 (-1)	ありそうで ない	-	-2.45	VMD: -0.66 -1.06 ~ -0.25	+++ 中
胆汁漏出	3 268	深刻な欠点 (-1) <sup>3</sup>	重要な問題 なし	問題 なし	不精確 (-1) <sup>4</sup>	ありそうで ない	RR 4.77 (0.23 - 97.96)	0.00%	0.02 -0.02 ~ 0.05	++ 低
入院期間 (日)	5 446	深刻な欠点 (-1) <sup>3</sup>	重要な問題 なし	問題 なし	不精確 (-1) <sup>4</sup>	ありそうで ない	-	19.17%	VMD: -1.45 -3.28 ~ 0.38	++ 低
胃内容 排泄遅延	5 442	深刻な欠点 (-1) <sup>3</sup>	説明不能な 不一致(異 質性) (-1) <sup>5</sup>	問題 なし	不精確 (-1) <sup>4</sup>	ありそうで ない	RR 1.52 (0.74 - 3.14)	25.50%	0.11 -0.08 ~ 0.29	+ 非常に低

注: 1. RR= 相対危険(ランダム効果モデル)

2. ARR: 絶対リスク減少率, WMD= 重み付け平均差

3. 全研究にわたり割付の隠蔽が不確実, 1件の研究でのみ患者盲検化, アウトカム評価のブラインド化なし, 3研究で脱落率(>20%), 1研究で  
intention-to-treat 解析なし

4. 信頼区間の間に1があるため, どちらの手術方法の利益(あるいは害)が大きいとは不明

5. 異質性: I<sup>2</sup>=72.6%, p=.006

分割した「C1: 行うことを考慮しても良いが, 十分な科学的根拠はない」や「C2: 科学的根拠がないので, 勧められない」とは決定的に異なっている。従来のような, 「科学的根拠はないのに行っても良い」というような曖昧さが徹底的に排除されているのが特徴といえる。

(1) 国際的な広い範囲のガイドライン作成担当者によって作成された, (2) エビデンスの質と推奨度を明確に分離したこと, (3) 代替治療戦略のアウトカムの重要性を明示的に評価すること, (4) 明示的, 包括的基準でエビデンスの質をグレードイングした, (5) エビデンスから推奨へのプロセスを明瞭化した, (6) 患者の価値観や好みを明瞭に取り入れる, (7) 臨床医, 患者, および政策立案者のために, 「強い」推奨, 「弱い」推奨を明確, かつ実践的な解釈とした, (8) 系統的なレビュー, 健康技術評価, およびガイドラインの作成に役立つ, というものであり, これらの一部は以前発表されている<sup>31)</sup>が, 最近 update された FAQ での比較表(後述)も参考になる。

以下に, GRADE システムによるエビデンスの質の評価や推奨度の決定プロセス, さらには診療ガイドライン作成への利用などについて, FAQ および, 関連論文<sup>31,44-50)</sup>を参考にして紹介する。

### 3. GRADE システムにおけるエビデンスの質

GRADE システムでは, 種々のレベルのアウトカム(エンドポイント)ごとに, 収集されたエビデンスの“質”の評価を行う。エビデンスの質とは, “推定効果(estimate of effect)に対する確信(confidence)のもてる程度”と定義される。

関心のあるアウトカムにおける推定効果の確信を明瞭かつ単純に評価するために, GRADE システムでは, 最初に, 研究デザインによってエビデンスの質を分類し, ランダム化比較試験は“高(high)”, 観察研究は“低(low)”, その他のエビデンスを“非常に低(very low)”とする。(表2a)

観察研究によるエビデンスは, 一般的に, ランダム化比較試験によるエビデンスよりも, その質は低くなる。その顕著な例としては, ホルモン補充療法による冠動脈心疾患のリスク評価がある。観察研究ではリスク低減効果が認められたが, その後のランダム化比較試験ではリスク低下は認められず, むしろリスクが増大するというものだった<sup>53)</sup>。

エキスパートの意見をエビデンスのカテゴリに分類する際には, やや混乱がありうる。もし, それがエキスパートにより, すべてのエビデンスを検討して質が高いか低いかを評価し“判定”されたものなら, エクスパートの“判定”は適

切(高い, あるいは低い)としてよい。

しかし, エクスパートの意見が, エクスパートの報告や彼らの経験という意味ならば, それらは, 症例報告や, 対照を置かない臨床観察と同様(あるいはそれ以下の), “非常に低”のレベルとすべきである。

さらに, 研究デザインとは別に, エビデンスの質は, 多くの因子で変化することがある。それらは, グレードダウンの5項目と, グレードアップの3項目である(表2b)。

出発点では, 研究デザイン評価として, “moderate(中)”というカテゴリーはなく, また“非常に低”の研究に対しては, グレードダウン・アップの評価はせず, 最終的な判定として, 4分類(高, 中, 低, 非常に低)に判定される(表2c)。各カテゴリーの意味は, 高=効果がどの程度確実なのかという推定が, さらに研究を重ねても変わる可能性が少ないもの, 中=さらに研究を重ねた場合, 効果がどの程度確実なのかという推定に重大な影響を与え, 推定が変わる可能性のあるもの, 低=さらに研究を重ねた場合, 効果がどの程度確実なのかという推定に重大な影響を与え, 推定が変わる可能性が高いもの, 非常に低=推定される効果がまったく不確実なもの, である(表2c)。

### 3 - (a) グレードダウンの5因子:

エビデンスの質の低下の代表的な因子として、研究の欠点 (study limitations)、結果の不一致 (inconsistency) の問題、エビデンスの直接性の問題 (indirectness)、データの不精確さ (imprecision)、報告バイアス (出版バイアスを含む) がある。

#### (1) 研究の欠点 (study limitations):

研究計画 (デザイン) そのものと、実施した研究全般についての欠点 (問題点) のことであり、ランダム化比較試験については、割付の隠蔽がないこと、バイアスの影響を受け易い主観的なアウトカムを用いた研究、ブラインド化されていないもの、フォローアップでの大きな脱落、ITT (intention-to-treat) 解析を遵守していないもの、利益が見込まれたとして試験を早期中止した場合、などである。早期中止については、今後、群逐次手法を用いた研究が増える可能もあり注意が必要である。研究上の欠点 (limitation) の程度は、2種類で、深刻か、非常に深刻か、である。

例えば、血栓症を合併したヘパリン誘発性血小板減少症 (HIT) の治療でダナパロイドナトリウムの利益があると示唆するランダム化比較試験がある。しかし、その試験はブラインド化されたものではなく、また血栓塞栓症の解消判定は、主治医による主観的な判定であり、エビデンスの質を低下させる欠陥があるものだった<sup>54)</sup>。

また、最近報告された膵臓癌例での外科治療関連論文 (幽門輪保存法と標準 Whipple 膵頭十二指腸切除術) で<sup>47)</sup>、6件のRCTを統合したメタアナリシス研究 (574例) では、割付遮蔽の欠如、患者やアウトカム評価者のブラインド化の欠如、フォローアップの脱落などの欠点があり、5年死亡率、輸血量という各々の重要なアウトカムについてのエビデンスの質は、“高 (high)”ではなく“中 (moderate)”であった。さらには、周術期死亡率、胆汁漏出、入院期間というアウトカムについての評価は、“低 (low)”であり、後述のように、胃内容排出遅延については、“非常に低 (very low)”であった。(表3)

#### (2) 結果が不一致 (inconsistency of results)

いくつかの試験結果で得られた結果が相互に一致せず (異質性 heterogeneity,

あるいは variability 多様性)、その理由が説明できない場合には、結果の不一致があるとする。多様性 Variability は、集団の違い (例えば、薬剤によっては、より病気の重い対象集団に投与された場合では相対的効果が大きくなる)、介入の違い (例えば、高用量の薬剤はより効果が大きくなりやすい);アウトカムの相違 (例えば、治療効果が経過とともに減弱する)、などでもおこりうる。もし、結果が一致せず (異質性あり)、研究者が納得できる説明ができないなら、エビデンスの質は低下することになる。

前述のメタアナリシス報告例では、胃内容排泄遅延 (delayed gastric emptying) をアウトカムとした幽門輪保存術の標準 Whipple 膵頭十二指腸切除術との比較では、方法論上の深刻な欠点によって低下して“中”のレベルとなったエビデンスの質を、この異質性の存在のためにさらに“低”まで低下させることになった。メタアナリシスにおける研究間の異質性の存在は、 $I^2$  値<sup>67)</sup> 72.6%,  $p=0.06$  から判定された。

#### (3) エビデンスが間接的 (indirect):

ガイドライン作成者はエビデンスの直接性 (directness) に関して2種類の問題に直面する。ひとつは、2種の薬剤 (AとB) のうちひとつを考慮する場合で、AとBを比較したランダム試験はないが、Aとプラセボ、Bとプラセボ、のランダム化比較試験がある場合である。このような場合、AとBの効果の程度を間接的に比較することになり、直接比較よりもエビデンスの質は低下する。第2のタイプは、対象患者、介入、比較対照、アウトカムについての相違であり、臨床試験の測定対象が、どの程度、関心の対象に似ているかである。もし目の前の患者が、研究内容と比較して、より高齢、重症、合併症が多く、異なる場合には直接性は不確実なものになり、間接的 (indirect) と評価される。

また、薬剤が、研究に使用されたものと同じクラスに属してはいるが異なる場合、そのエビデンスの直接性については不確かと判定されるであろう (例; 気管支喘息での吸入ステロイド療法に関する似たようなランダム化試験がなされていても、薬剤の種類、用量の違い (大量療法) や、使用回数の違いなど)。

さらに、代理アウトカムを用いる研究は、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

はない (例; 喘息の悪化や死亡率が最も重要であるときに、代理アウトカムとして一秒量 (FEV1) しか測定されていない場合など)。

#### (4) 不精確 (imprecision):

標本サイズが小さい場合やイベント数が少ない研究は、信頼区間が広くなり、利益についても、downsides (害、負担、コスト) についても、その結果を良質のエビデンスとして扱うのは躊躇するすなわち、その結果は不確実とすべきである。また、研究が1件しかない場合にも、不精確と判定し、エビデンスの質は低下すべきである。

前述の例では、幽門輪保存の代替手術の多くのアウトカムでは、著明効果と無効の両者の存在や、あるアウトカムでは両方向での重要な相違が存在している。また、前例の胃内容排泄遅延のアウトカムにおいては、相対危険 = 1.52 (95% 信頼区間 0.74-3.12) であり、最終的には、エビデンスの質は“非常に低い”までに低下することになっている (表3)。

#### (5) 報告バイアス (reporting bias):

たとえば検討対象としている薬剤に有利な結果が得られれば出版されやすく、有益性が認められなかったり害がみとめられた場合には出版されがたい。これは出版バイアスと呼ばれ、結果の報告バイアスの最たるものである。また、たとえば、研究結果が出版されたとしても、アウトカムのうち検討対象の薬剤に都合のよいものだけが報告され、都合の悪いアウトカムの結果が報告されないことがしばしばある。極端な場合は、総死亡が増加傾向を示しているのに、ある疾患の罹患率が低下したことのみで有効としているような例がある。

このように報告バイアス (出版バイアスなど) の可能性が高い場合も、グレードダウンになる因子の例としてあげられている、痔核の患者でのフラボノイドについての14試験では、明らかに大きな利益がみられたが、研究に登録された患者はわずか1432人だった<sup>55)</sup>。これらの試験のほとんどにスポンサーが強く関与しているため、利益を全く示さない未発表の試験が存在するのではないかとという疑問がわく。

Cochrane レビューにおいてさえも、同定されたランダム化比較試験の約半数が患者にとって重要と思われるエンドポイント (アウトカム) の効果推定に貢献

できていない。またメタアナリシスで同定されたランダム化比較試験のうち実際に効果推定に貢献できた試験の占める割合と推定効果には逆相関が見られたことから、この割合の低いメタアナリシスは真実の効果を過大評価している可能性がある。これを outcome reporting bias(アウトカム報告バイアス)と言うが<sup>66)</sup>、今まで注目されてきた study reporting bias(研究の報告バイアス/出版バイアスなど)に加えて、このアウトカム報告バイアスにも留意すべきである。

以上のような研究の欠点(limitations)が多ければ多いほど、エビデンスの質は低いものになる。すなわち、膵臓がん例(表3)では、5つのランダム化比較試験が利用できたとしても、各アウトカムによっては、最終的なエビデンスの質は“中”、ないし、“非常に低”にまで低下するものであった。

### 3-(b) グレードアップの3因子:

良質であっても観察研究は一般にエビデンスの質は“低”であるが、稀ながら、エビデンスの質がグレードアップし、“中”、“あるいは”高”のエビデンスと評価されることがありうる。主に3つの因子がある(表2-b)。すなわち、(1)効果の程度(magnitude of effect)が大きい、(2)交絡因子のために効果・影響が過小評価になっていると考えられる、(3)用量-反応関係(勾配)の存在である。

#### (1) 効果の程度(magnitude of effect)が大きい

治療効果や危険度などエフェクト・サイズ(effect size)が大きい場合である。相対リスク(RR) >2, あるいは RR < 0.5 の場合には効果・影響の幅が大きく、グレードアップとなる。また、相対危険(RR) > 5 あるいは (RR) < 0.2 の場合には、効果・影響が非常に大きいと考えるのが妥当であろうと考えられている。

例えば、観察研究のメタアナリシスで<sup>66)</sup>、自転車用ヘルメットはかなりクラッシュにかかわる自転車乗りの頭部外傷の危険を減少させると報告されており(OR 0.31, 95% CI 0.26-0.37)、この大きな効果はエビデンスの質を“中”にグレードアップさせる。また、心弁膜置換術実施患者でのワーファリン予防治療効果を評価した観察研究でのメタアナリシスで<sup>67)</sup>、血栓塞栓症の相対危険(RR)は 0.17 (95% CI 0.13-0.24) で、この非常

に大きな効果は、エビデンスの質を“高”のレベルを示唆するものである。

#### (2) 交絡因子のため、効果・影響が過小評価になっている

交絡因子のため、結果データで示された効果・影響が真の効果・影響よりも弱められていると考えられる場合である。例えば、もし、より重病の患者だけが実験的な介入治療を受けていて、それでも実験群でより良い結果ということであれば、それは実際の介入治療効果は得られた結果以上の効果を示唆するものであるということである。

#### (3) 用量-反応関係(勾配)の存在

例えば、ワーファリンによる抗凝固療法をうけている患者で、INRのレベル上昇と出血リスクの増加の間には用量反応勾配があることが観察研究で判明しており、治療域以上の抗凝固療法は出血リスクを増加させるという確信になっている。

### 3-(c) 全体的エビデンスの質とアウトカム:

以上のような因子を考慮して、最終的に各アウトカムについてのエビデンスの質を決定し、4つのカテゴリーのどれかに判定する。

各アウトカムについてのエビデンスのレベルを評価し、推奨について検討することになるが、ここで、重要なことは、どのようなアウトカムが患者にとって重大(critical)なのかということである。結果が利益であれ、害であれ、複数の研究で結果が同じならば、問題はないが、重要ないくつかのアウトカムに関してエビデンスの質が異なっている場合には判断が面倒となる。

その際の原則は、それぞれのアウトカムについてエビデンスの質を決定したうえで、“全体的なエビデンスの質”は、“重大(critical)”なアウトカムの“最も低いエビデンスの質”を採用することである。

すなわちアウトカムが重大(critical)なのか、重要だが重大とはいえない(important but not critical)のかを考慮すべきであり、エビデンスの質が“低”のアウトカムが、重要で重大(critical)なら、全体的エビデンスは、“低”と判定する。

本膵臓癌例に例えると、推奨を作成する者が、仮に胃内容排泄遅延が重大(critical)なアウトカムと判断するなら、

全体的エビデンスの質は“非常に低”となるだろう。胃内容排泄遅延は重要だが非常に重大とはいえない(not critical)ものならば、5年死亡率に関するエビデンスの“中”の質があるにもかかわらず、(明らかに重大(critical)な周術期死亡率についての結果を基にして)、全体的エビデンスの質は“低”であろう。

忙しい臨床医、忙しい患者や行政担当者には簡潔、明瞭、容易に理解できるエビデンス要約が必要であり、GRADEシステムではこのような支援のためのツールを準備している。これは、visual basicを基本としたspread sheetで、一連のCQから、アウトカムの相対的なランキング、エビデンスの階層的判定、推定効果の算出などを支援する簡便なツールである。(2007.8月現在β版で、未公開)。

また、GRADEシステムでは、ガイドラインなどでも、推奨とエビデンスの質として、一見して内容が理解できるように、オリジナルのシンボル・記号や推奨度ナンバーリング(1あるいは2)の利用も許容している。(図1)また、医療の最前線においては、最適な医療行為としては、推奨度にもっとも興味があることから、本systemでは、第一に“推奨度”を、次に“エビデンスの質”を表示する(例:強い推奨/エビデンスの質は“高(high)”)。

### 4. GRADEシステムにおけるアウトカムの3段階評価:

どのようなガイドラインや推奨の作成においても、介入(治療や診断など)に関しての疑問(clinical question: CQ)を設定し定式化するが、それは4つの因子からなりたっている。すなわち、Patient, Intervention, Comparison, Outcomeである。前述の例では、膵臓癌で手術した患者で、幽門輪残存の改変手術は、標準の広範囲切除術と比較し、短期・長期の死亡率、輸血、胆汁漏出、入院期間、胃内容排泄遅延についての効果・影響である。

アウトカムを抽出する際、患者にとっての各アウトカムの重要性を相対的に判断し、重大(critical)、重要だが重大ではない(important but not critical)、重要ではない(not important)、の3段階とする。各研究での著者が主要アウトカムとしているものや、他のアウトカム、さらには自分がなんらかの決定をおこなう上で重要かもしれないと思う(利益と害、

図1 エビデンスの質と推奨度の表現の例

エビデンスの質	シンボル	文字
高	⊕⊕⊕⊕	A
中	⊕⊕⊕○	B
低	⊕⊕○○	C
非常に低	⊕○○○	D

推奨度	シンボル	番号
強い推奨(介入の実行あるいは反対)	↑↑ OR ↓↓	1
弱い推奨(介入を実行あるいは反対)	↑? OR ↓?	2

注:GRADEでは、特に上記のいずれを使用することを規制しているものではないが、文字や番号にこだわるならば、推奨度には番号を、エビデンスの質の判定表示には文字を使用するよう薦めている。

関連したものならコストまで含め)他のアウトカムも追加して記載する。全アウトカムのうち重大・重要なものをエビデンス・プロフィール評価の対象とする。アウトカムが重大か、重大ではないが重要か、重要でないかの決定は、ひとつの価値観であり、その際には、提案する推奨が影響をおよぼす人の価値観をできるだけ考慮すべきである。

図2は階層構造を示すが、褥瘡を有する老人病患者での経腸栄養効果についてのエビデンスを調査する委員会によって選ばれたアウトカムの重要性の確立することへのアプローチ例の一つである。ガイドライン作成者はスケールの上部(7~9)に相応するものを重大なアウトカム、中段スケール(4~6)を重要だが重大なほどではない、低スケール(1~3)は臨床研究で時々測定される代理アウトカムで、患者には直接的な重要性はないものとすべきである。ガイドライン

パネルは、この例が示すような明白なアプローチの種類を求めて努力すべきである。

5. GRADE システムにおける推奨度

本システムでの推奨度とは、“介入による望ましい効果(利益)が、望ましくない効果など、downsides(害・負担・コスト)を上回るか(positive)、あるいはその逆(negative)かについて、どの程度確信できるかを示すもの”と定義される。望ましい効果とは、死亡率の低下、特定疾患による死亡率低下、罹患率低下、QOL改善、治療の負担軽減(治療を受ける必要性、血液検査やモニターのための通院の不便性など)、医療費の減少であり、望ましくない結果とは死亡率上昇、罹患率上昇、あるいはQOLへの悪影響、害反応症状や検査異常などである。本システムでは、推奨度の分類は、強(strong)、弱(weak)の2種類であ

るが、推奨の方向性として、介入を実施する(positive/recommend)、しない(negative/against)の2種類があり、結果としての推奨表現としては4種類となる(表4)。推奨度の分類が2種類で少ないのは、臨床医、患者、政策担当者の理解を容易にするためである。

この2分類表示により、臨床医、患者、さらには行政担当者にとっても、推奨度の理解が容易になるであろう。GRADEワーキンググループでは、あえて各立場の違いによる推奨度の意味を具体的に説明している。(表1)

また、推奨度を表現する用語としては、“強・弱”にこだわる必要はなく、例えば、“推奨する(recommend)”や“勧める(suggest)”,あるいは、“~すべき(should)”や“~してもよい(may, might)”など、実際にはさまざまな表現が用いられている。ただ、これまでのGRADE会議から、従来記載されていた、do it, probably do it, probably don't it, don't itは、誤解が生じるやすく好ましくないという意見から現在の表現になっているが、これらの用語を禁止しているものではない。

一方、従来日本のガイドラインの多くで採用されてきた、A, B, C, D分類は上記の4カテゴリーとは異なる。特に、カテゴリーC, Dが問題であり、“C”は、「行うよう勧めるだけの根拠が明確でない」、Dは「[行わないよう勧められる]と定義されている(国内4分類)。また、さらに“C”を、C1[行うことを考慮しても良いが、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」に分ける国内5分類もある。GRADEシステムでは、この点が決定的に異なる。つまり、GRADEシステムでは、用いるべきでない方向にも強い推奨と、弱い推奨の2段階があり、従来のような、科学的根拠はないのに「行っても良い」というような曖昧さを徹底的に排除したことが特徴といえる。

推奨度の強さは、その定義からも明らかかなように質の高いエビデンスは強い推奨につながるものが一般である。しかし、推奨度は臨床に直結したものであり、エビデンスの質の他の因子も考慮する必要がある、そのための4つのキーとなる因子がある。(表5)

- 5 - (a) 推奨度決定の4つの主要因子
- (1) エビデンスの質
- 推奨度の決定として重要な因子がエビ

図2 患者の重要度に従ったアウトカムの階層表示例 (褥瘡を有する老人病患者での経腸栄養の効果)<sup>49)</sup>

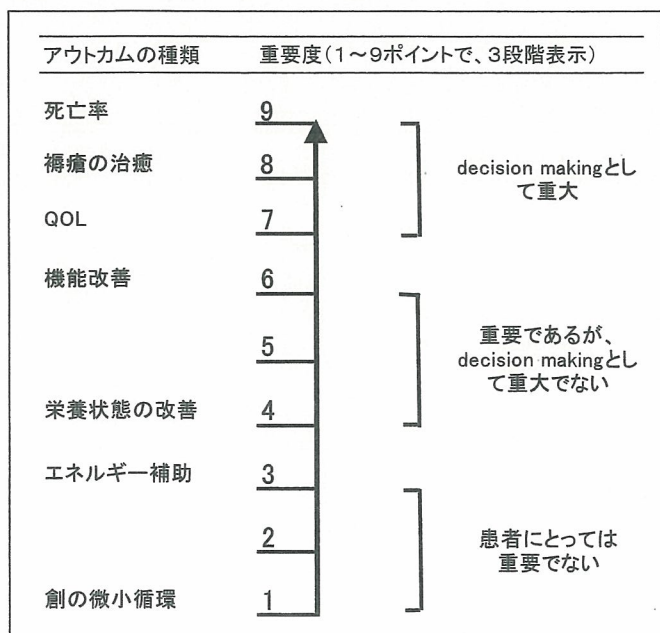


表4 推奨度：2分類

1. 強い推奨	ある介入がによる利益が明らかに (clearly) 害・負担・コストをしのぐ、あるいは明らかにその逆である場合。
(表現例)	"we recommend for ~" or "recommend against ~" "clinicians should ~" or "clinicians should not ~" ~を行うよう強く推奨する。あるいは、~を行わないよう強く推奨する。
2. 弱い推奨	(1) 利益/downsides (害, 負担, コスト) のバランス評価の確実性が低い (less certain) 場合, (2) "低い質" のエビデンスの場合, あるいは, (3) 利益と不利な面が比較的接近していることを示唆する "高い質" のエビデンスがある場合。
(表現例)	"we suggest ~", or "suggest not ~" "clinicians might ~" or "clinicians might not ~" ~を行うほうがよいかもしれない (どちらかといえば~を行うのがよいと思われる)。あるいは ~は行わないほうがよいかもしれない (どちらかといえば、~は行わないほうがよいと思われる)。

参考: ACPJC 144 (2006), Am J Respir Crit Care Med 2006: 174

デンスの質である。もし、ある介入による利益や害の程度が不確実であるならば、強い推奨をすることは問題である。

(2) 望ましい結果と望ましくない影響とのバランス

利益と downsides (害, 負担, コスト) とのバランスである。利益が downsides を明らかに上回る場合や、逆に downsides が利益を明らかに上回る場合にはその推奨の確信は、強い (strong) ものになる。利益と downsides が拮抗している場合 (trade-offs の関係にある場合) や、利益と downsides, あるいは両者のいずれかの程度に不確かさがある場合は、推奨度は、弱い (weak) ものになる。例えば、心筋梗塞でのアスピリンや、低リスク心房細動患者でのワーファリンなどである。

(3) 価値観と好み

若年者と高齢者では、延命に対する価値観は同じではない。したがって、治療による利益と害・負担に対する評価は異なりうる。また、特発性深部静脈血栓症 (DVT) の40歳男性患者で、1年間用量調整ワーファリン治療を受けている例を

考えてみよう。もしその患者が標準強度のワーファリン治療を継続するならば、DVT再発は約10%/年の減少となることが推定される。毎日のワーファリン錠服用やビタミンK摂取量を一定にすること、抗凝固治療の強度モニターの血液検査、という治療による必然的な負担と、小~大出血のリスク上昇を比較して、DVT再発を強く嫌う患者は、それらの欠点はワーファリン服用に値するものと思うかもしれないし、患者によっては、その利益はリスクに見合ったものでないと考える人もいるだろう<sup>46)</sup>。

(4) コスト

一過性脳虚血発作患者では、アスピリン投与よりクロピドグレル投与などがコストはかなり高く、より高いコストは介入の強い推奨を低下させることになる。

また、医療の利益得失を比較する際、アウトカムの一つとしてコストを考慮すべきである。しかし、費用は他のアウトカムより、時間、地理的区域、および含意の上ではるかに多彩である。特許期限が切れると薬剤コストは急に下がる傾向があり、同じ薬剤に対する費用も国が異なり、規制が異なれば著しく異なっ

てくる。さらに、医療資源もかなりばらつきが大きく、例えば、同じ高価な薬剤の年間処方費用は、米国の独身看護師1人の給料、ポーランドの看護師6人分、および中国の看護師30人分の給料を支払うことに相当している。コストの影響を強くうける推奨は、医療資源の拡大に応じて、また、時間の経過により変化しやすいものである<sup>50,58)</sup>。

5-(b) 推奨度のグレーディング分類一覧

推奨度の表示と、エビデンスの質の組み合わせを一覧表示すると、表6のようになる。

弱い推奨/エビデンスの質が“高”、強い推奨/エビデンスの質が“非常に低”、などがあるが、各々の意味をよく理解していただきたい。

なお、「強い推奨度」の行政的意味は、適応とされる患者に対する介入として採用すべき行政施策であることを示しているだけであり、行政施策上の重要度とは異なるので、注意が必要である。

6. GRADE システムの意義と国内外の格差

GRADE システムの認識・理解、あるいはその採用は国内外で全く異なっている。GRADE システムはそもそも、医療技術の“推奨”に関して、「つくる」、「つかう」、「つたえる」、いずれの側にも対応したものである。

6-(a) GRADE システムの現状と今後

本GRADEシステムは、従来の診療ガイドラインの推奨度とは全く異なっている。従来の診療ガイドラインでは、研究デザインに基づくエビデンスレベルと数を主体とし、それにせいぜい結果のばらつきや臨床の有効性の大きさなどを加味して推奨度が決定されていた。し

表5 推奨度に影響する4主要因子、強い推奨、弱い推奨の例

因子の種類	強い推奨 (例)	弱い推奨 (例)
エビデンスの質	多くの質の高いランダム化比較試験が気管支喘息での吸入ステロイドの利益を証明している (訳註 a)	気胸に対する胸膜癒着の効用を検討しているものは症例集積研究が一つあるだけである
望ましい効果と downsides のバランス	心筋梗塞でのアスピリンは死亡率を減少させ、毒性や不便さ、コストも少ない	低リスクの心房細動でのワーファリンはわずかに脳卒中減少をもたすが、出血リスクの上昇や、かなりの不便さがある
価値感や好みにおける多様性 (一様 v s 多様)	悪性リンパ腫の若年患者では、その多くが、治療に伴う毒性より、化学療法による寿命延長効果の方に高い価値をおくだろう	高齢の悪性リンパ腫患者では、治療に伴う毒性よりも、化学療法による寿命延長効果の方に高い価値をおくことはできないかもしれない
介入が医療資源の賢明な使用か否か	一過性脳虚血発作の患者での脳卒中予防としてのアスピリンは、低コストである	一過性脳虚血発作の患者での脳卒中予防としてのクロピドグレルやジピリダモール/アスピリンはコストが高い

訳註 a: フルチカゾンに関しては、害の可能性の大きさから、該当しないとの指摘がある<sup>69,70)</sup>

かし、GRADE システムは、エビデンスの質をアウトカムごとに、システムティックに評価し、アウトカムの重要性、患者因子(価値観、負担、コストなど)を考慮して、その作成プロセスも分かりやすくした“推奨度”の判定作成方法である。さらに、GRADE ワーキンググループは、常にシステムのさらなる単純化、明瞭化をめざし、年間数回の会合を継続している。GRADE システムの自己評価や問題抽出にとどまらず、他のシステムとの協調性、コクランライブラリーの要約の取り込みソフトの開発、GRADE システムでの用語の整理、診断研究や経済評価のシステムの発表を予定している。未だいくつかの問題はあるものの、その単純性、明瞭性は海外での採用状況を見ることでも容易に理解でき、web サイトでのFAQ解説、さらには今後、エビデンスの質の評価支援ツールの公開が準備されている。  
(<http://www.gradeworkinggroup.org>)

6- (b) 海外における GRADE システム利用の急速な普及

国外では2005年より、American Col-

lege of Chest Physicians, ACCP<sup>45)</sup>, American Thoracic Society<sup>46)</sup>, UpToDate<sup>33)</sup>, WHO<sup>34)</sup>, Ontario Ministry of Health and Long-Term Care (Canada)<sup>35)</sup>, Endocrine Society Clinical Guideline (USA)<sup>36)</sup>, Kidney Disease: Improving Global Outcomes<sup>37)</sup>, Evidence-based Urology (EBUro)の一つである American Urological Association<sup>39)</sup>, U.S. Preventive Services Task Force (USPSTF)<sup>40)</sup>, Agency for Healthcare Research and Quality (AHRQ)<sup>41)</sup> など約30の医療機関や団体が、GRADE システムを、オリジナルあるいは、改変システムとして採用している(2007年8月時点)。いずれにも共通していることは、GRADE システムの利便性や客観性、明瞭性などを評価したもので、あらゆる診療分野で利用され、急速な普及をみており、臨床現場での推奨度の評価や診療ガイドライン作成、予防医学の領域での適用も報告されている<sup>39-41)</sup>。さらに、この度の Guyatt らによる GRADE についての5部シリーズの発表でさらに加速されるだろう<sup>48-52)</sup>。

6- (c) 国内における GRADE システムの普及の遅れ

一方、国内では、GRADE システムに関しては、2005年4月に厚生労働科学研究業績として、津谷らの概略、論文翻訳<sup>59-61)</sup>があり、「日本においても、このGRADE system が診療ガイドライン作成やシステムティック・レビュー、さらには、医薬品の適応外使用や未承認医薬品のエビデンスを理解するうえで、積極的に使用されることが期待される」と記載している。

しかしながら、GRADE システムの根幹ともいえるべき、「エビデンスの質の吟味・評価」の強調が不十分のまま、“推奨度4分類”として紹介された形となった。実際、最近作成された日本の診療ガイドラインの一部には、GRADE システムを参考・参照したと記載されているものがあるが、そのほとんどはGRADE システムを適切に利用したとはいえない。例えば、「小児急性中耳炎診療ガイドライン」や「褥瘡局所治療ガイドライン」でも同様であるが<sup>62,63)</sup>、エビデンスの質の吟味・評価はなく、むしろ従来からの研究デザインを主体としたエビデンスの

表6 推奨のグレーディング

推奨グレード	利益/リスク・負担	推奨を支持するエビデンスの質	意味
強い推奨/"高"の質のエビデンス	利益が明らかに害・負担をしのぐ、あるいはその逆。(clearly)	良好に実施された複数のランダム化比較試験からの一定したエビデンス、あるいはバイアスのない複数の観察研究での例外的に強固なエビデンス	推奨はあらゆる環境において、全患者に適用できる(can)。さらなる研究を実施しても、推定効果への確信は変わることはほとんどない(very unlikely)。
強い推奨/"中"の質のエビデンス	利益が明らかに害・負担をしのぐ、あるいはその逆。(clearly)	重大な制約(一貫性がない結果、研究方法の欠陥、直接性がない、不精確、のいずれか)のある複数のランダム化比較試験のエビデンス、あるいはバイアスのない複数の観察研究での通常ではありえない強いエビデンス。	推奨はあらゆる状況において、全患者に適用できる(can)。さらなる研究が(もし実施されるなら)、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)、結果としてその推定が変わるかもしれない(may)。
強い推奨/"低"の質のエビデンス	利益が明らかに害・負担をしのぐ、あるいはその逆。(clearly)	複数の観察研究、あるいは重大な欠陥か直接性のないエビデンスの複数のランダム化比較試験による、少なくともひとつの重大なアウトカムに対するエビデンス	推奨は、より質の高いエビデンスが利用できるようになると変わるかもしれない(may)。さらなる研究が(もし実施されるなら)、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)、結果としてその推定が変わる可能性がある(likely)。
強い推奨/"非常に低"の質のエビデンス(非常に稀)	便益が明らかに害・負担をしのぐ、あるいはその逆。(clearly)	システムティックでない臨床観察による少なくともひとつの重大なアウトカムに対するエビデンス、あるいは重大な直接性のないエビデンス	推奨は、より質の高いエビデンスが利用できるようになると変わるかもしれない(may)。少なくともひとつの重大なアウトカムに対してのどの推定効果も非常に不確かなものである。(very uncertain)。
弱い推奨/"高"の質のエビデンス	便益が害・負担に近い。(closely)	良好に実施された複数のランダム化比較試験からの一定したエビデンス、あるいはバイアスのない複数の観察研究での例外的に強固なエビデンス	最適な行為は、環境、患者、あるいは社会的価値によって異なるかもしれない(may)。さらなる研究を実施しても、推定効果への確信は変わることはほとんどない。(very unlikely)。
弱い推奨/"中"の質のエビデンス	便益が害・負担に近い(closely)	重大な制約(一貫性がない結果、研究方法の欠陥、直接性がない、あるいは不精確のいずれか)のある複数のランダム化比較試験のエビデンス、あるいはバイアスのない複数の観察研究での通常ではみられない強いエビデンス	ある環境の下で、ある患者では別の治療が良い可能性がある(likely)。さらなる研究が(もし実施されるなら)、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)、結果としてその推定が変わるかもしれない(may)。
弱い推奨/"低"の質のエビデンス	便益、害、負担の推定が不確実(uncertainty)、便益が害・負担に近いかもしれない。(may be closely)	複数の観察研究、あるいは重大な欠陥か直接性のないエビデンスの複数のランダム化比較試験による、少なくともひとつの重大なアウトカムに対するエビデンス	他の治療も同様に妥当なものかもしれない(may)。さらなる研究は、推定効果への確信に重要なインパクトを持つ高い可能性があり(very likely)、結果としてその推定は変わる可能性がある(likely)。
弱い推奨/"非常に低"の質のエビデンス	便益、害、負担の推定が非常に不確実(major uncertainty)、便益が害・負担とバランスがあるかどうかわからない。(may or may not be closely)	システムティックでない臨床観察による少なくともひとつの重大なアウトカムに対するエビデンス、あるいは重大な直接性のないエビデンス	他の治療も同様に妥当なものかもしれない(may)。少なくともひとつの重大なアウトカムに対してのどのような推定効果も非常に不確かなものである(very uncertain)。

引用: Am J Respir Crit Care Med 2006; 174: 605-614 (表6)

表7 用語集

用語	意味・意義
absolute risk reduction (ARR)	絶対リスク減少率: 対照群でのイベント発生率(CER)と治療群でのイベント発生率(EER)との差。すなわち, $ARR=CER-EER$
association(関連)	厳密には関連の強さ(強固性)。特に観察研究で因果関係の強さを補強する因子。GRADE システムでは, グレードアップ評価因子のひとつ。相対危険の程度により, 強い関連性( $RR>2, <0.5$ ), 非常に強い関連性( $RR>5, <0.2$ )を判定する。large magnitude of effectとも記載される。
bias(バイアス)	研究の繰り返しによっても生じる推定値と真の値との差(系統的な誤差のこと)。対象者の選択やデータ収集・分析, などに際して, 方法が不適切な場合に生じ, 結果を誤らせる重要な要因となる。
burdens(負担)	負担/重荷: 患者あるいは介護人(家族など)が, 推奨を硬く守るようなことが嫌かもしれない要求で, 回復には, より長期間かかるかもしれない治療の選択, より頻繁な検査を受けなければならない, などである。
confidence interval (CI: 信頼区間)	真の値が存在すると期待される研究結果の範囲を示す。利益と害の療法の情報を提供する区間推定ともいう。
continuous data (連続データ)	一定の間隔で並ぶ連続した数値で, 血圧, 体重, 血球数など。数値変数 numeric variable, 間隔変数 interval variableともよばれる。
direction of confounder(交絡因子が影響する方向)	観察研究でのグレードアップ評価因子の一つ。結論に示された効果が, 交絡因子のために真の効果よりも弱められていると考えられる場合。
directness(直接性)	EBMのPICOに相当するもので, 研究の試験参加者, 介入, アウトカム指標が問題としている介入の状況と一致しているかである。確実な, あるいは非常に確実な問題があるかで, 1~2ポイントの減点となる
dose-response gradient(用量-反応関係/勾配)	グレードアップの評価因子の一つで, 用量-反応関係(用量-反応勾配)があること(曝露量が多くなればなるほど, 反応あるいは相対危険が大きくなる)
downsides(マイナス面/短所)	害(リスク), 負担(burdens), コストなど
early stopping(早期中止)	ランダム化比較試験の中間解析で, 有意な利益が得られたとして試験を早期に終了すること(また逆に害がある可能性が高くなったとして中断すること)がある。これは, 計画した患者数より少ない患者数, 短い観察期間となり, 介入による利益がある試験の場合にはグレードダウンとなりうる。害のための早期終了はこの限りではない。
funnel plot(漏斗図)	出版バイアスによって結果が歪められていないかどうかを検討するために, メタアナリシスに組み入れられた試験を図示する方法。
Group Sequential Method (群逐次手法)	研究期間中にデータが徐々に蓄積されていくような実験や調査において, 途中段階までに集まったデータをとりまとめて中間解析(Interim Analysis)を実施し, その結果をもとに実験や調査を早期に終了したり, 研究デザインを修正したりする手法で, 研究企画時に決定しておく必要がある。Early stopping参照。
Hazard ratio(ハザード比)	これまでに発生していない疾患がある短い時間間隔中に起こる速度を「罹患率(incidence rate あるいは morbidity)」という。同様に死亡がある短い時間間隔中に起こる速度を「死亡率(death rate) あるいは mortality)」という。いずれもこれまで発生していない事象が発生する速度であることから, これらをまとめてハザードといい, これらを2群間で比をとったものがハザード比である。当然HRが1ならば, 群間に差はなく, 1点だけを見ると, 相対危険と同じである。
heterogeneity(異質性)	異質性は「不一致(inconsistency)」と同様のカテゴリーである。患者間あるいは研究間の相違を意味し, データ蓄積の信頼性の低さ, または不適切性につながる。(引用文献25)では, $p<0.1$ , あるいはI <sup>2</sup> 値>20%と規定している)
I <sup>2</sup> 値	メタアナリシスなどでの異質性の検定法のひとつで, Cochranの $\chi^2$ 二乗(Q)テストを用い, 結果の変動の比率を算出する。具体的には(df=自由度: 研究数-1), $I^2=100\% \times (Q-df)/Q$ 。I <sup>2</sup> >20-25%のことが多いが厳密なカットオフ値はなく, また単独で利用されることは少なく, p値などとともに, 総合的に判断する。
imprecise(不精確)	信頼区間が広く, RRあるいはORが1をはさんでおり, 重要な害と重要な利益のどちらにも推定できる場合である。
inconsistency(不一致)	複数の研究で推定効果・影響(の方向, 程度)が一定していない(再現性が乏しい), あるいは異質性(heterogeneity)があること。
limitations(欠点)	ランダム化比較試験では, エビデンスの質のグレードダウンにつながるもので, 割付の隠蔽やブラインド化あるいはフォローアップが不十分, 両群間に重要な背景因子の偏りがある, 利益があるとしての早期中止, ITT解析でないものなど, 結果の解釈に偏りをもたらすもの
observational study(観察研究)	ある曝露要因の健康状態に対する関連を調べるための, 意図的介入をしない(非実験的)方法による研究。コホート研究, 症例対照研究, 横断研究が代表的な観察研究の方法である。実験的研究/介入研究に対する用語
outcome reporting bias (アウトカム報告バイアス)	報告バイアスの最も典型的な例。研究者(あるいは研究対象とした介入)にとって都合のよい結果(一部の病気や検査値の改善)は報告されやすく, より重要だが都合の悪い結果(死亡率の上昇や, 重篤な害など)は報告されにくい。その最も著しい例が, 研究報告バイアス(出版バイアス)である。
outcome(アウトカム/結果)	GRADEシステムでは, アウトカムの重要度を3段階で判定する:(1)重大, (2)重要だが重大でない, (3)重要でない, (GRADE profilerでは全体では, 1~9ポイントをあて, 最終的にはそれを3つのカテゴリーとする)
overall quality of evidence(エビデンスの質の総合評価)	各アウトカムについて, 研究デザイン, 欠点(limitations), 一致度(consistency), 直接性(directness), 精確性(precision), 報告バイアス, の5主要因子, グレードアップの3因子をチェックして, 重大/重要なアウトカムについてのエビデンスの質を評価する。

表7 用語集(つづき)

用語	意味・意義
quality of evidence (エビデンスの質)	利益と、downsides(害、負担、コスト)についての推定効果が正しいことにどの程度の確信(confidence)がもてるかを示すもの。GRADEシステムでは、4段階(high, moderate, low, very low)で評価する。
recommendation (推奨)	利益がdownsides(害・負担・コスト)を上回るか、その逆に、どの程度確信できるかを示すもの。推奨度は2種類であり、その方向性も2種類である。推奨の強弱を判定するために4つの主要因子の考慮が必要である
relative risk (RR: 相対リスク)	risk ratioともいう。治療群イベント発生率(EER)の対照群イベント発生率(CER)に対する比率。 RR=EER/CER
sparse(まばら)	一つの研究におけるアウトカム(結果)に関係した事象(イベント)の件数が少ない場合、その事象に関する情報が提供されない場合をいう。あるいは、ある結果に関する研究が1件しかない場合なども「まばら」とすべきであり、エビデンスの質を低下させることになる。
standard mean differences(連続データでの標準化平均差)	測定尺度の単位が異なる場合、効果の大きさの要約推定値を求めるため、共通尺度として各研究での両群の平均値の差を標準偏差で割った値などを計算し、重み付きの和を計算するなどの統合方法をとる
study design(研究デザイン)	基本的な研究デザインのことで、GRADEシステムでは、ランダム化比較試験、観察研究および、その他の3種類に分類される。
study quality(研究の質)	当初、各アウトカムにおけるエビデンスの質の評価の主要因子の一つとして記載されたが、「質」を評価する因子は種々あるため、近年の報告では、欠点(limitation)の有無とその程度によって、エビデンスの質が詳細に評価されるようになっている。
trade-offs(トレードオフ)	利益を大きくしようとするのとそれに伴って害も大きくなり、害を小さくしようとする、利益も小さくなるという関係。
validity(妥当性)	測定しようとした変数を正しく測定できている度合い、あるいは介入で達成しようとしたことが達成されている度合いをいう。
weighted mean difference (WMD: 重み付け平均差)	アウトカムが二分的(死亡や心筋梗塞など)ではなく、連続的(症状スコアや身長など)な場合に用いられるエフェクト・サイズの尺度の一つ。研究対象の群間におけるアウトカムの平均差は、異なるサンプルサイズと精度を計算に入れて重みを付けられる。WMDは絶対値であるので、元のアウトカムの尺度の単位をもつことが通常である。

レベル分類と推奨度分類と言わざるを得ない。

患者側因子を中心としたガイドラインの必要性は、国内でも従来から、診療ガイドライン作成の手順(GLGL) ver4.3 2001(福井・丹後)<sup>61</sup>で強調されていたが、実際に作成された国内診療ガイドラインの多くは、アウトカム(エンドポイント)の重要度や、患者因子を反映したものではない。アウトカム(エンドポイント)の重要度に関する検討はほとんどなく、また、患者参画での診療ガイドラインはわずかに3件と報告されている<sup>62</sup>。患者の語り、患者の価値観、あるいは、「診療ガイドライン作成過程への患者・支援者参画のためのガイドライン」、「Patient Involvement Guidelines (PIGL)」の作成<sup>64</sup>、患者の「ナラティブ(語り)」のデータベースとしてのDIPE x (Database of Individual Patient Experiences)-Japan

DIPEx-JAPAN <http://www.tip.gr.jp/dipex-j/>、などの準備が数年前より始まっている。また、NPO法人COML、アラジーボット、VOL-NEXT、あけぼの会、子宮筋腫・内膜症体験者の会・たんぼぼ、など診療ガイドライン作成班との共同・関連報告が増えつつある<sup>65</sup>。しかし、患者会にも、製薬企業の援助を受け

ている組織が少なくないため、そうした患者組織の代表がガイドラインの作成に参加して、はたして純粋に患者のための適切なガイドラインづくりに資することになるのだろうか。

#### 7. GRADEシステムの「問題点」

臨床決断として重要な推奨度の判定は非常に複雑であるが、従来のガイドライン作りにおける推奨度評価システムと比較して、GRADEシステムは単純性、明瞭性、アウトカムの重要性評価など、多くの利点が報告されている。

しかし、注意しておかなければならないのは、エビデンスの質評価におけるスコアリングが困難な例外的な場合にも機械的に分類する危険性であろう。各研究の質の評価が既存の基準で評価できない事例、あるいは評価者間で大きく分かれる例はありうるであろう。そうした場合には、単純なスコアリングによる判定にならないようにすべきであり、そのような事例の扱いを徹底的に議論し、どのように一般化できるかについて検討がなされる必要がある。また、今後登場する予定になっているGRADEソフトにしても、あくまでも評価支援のひとつのツールと考えておくべきである。

国内において、本システムを採用・利

用する上では、国内の多くの診療ガイドラインが作成基準も含め統一されておらず、非常に多くの支障があり、またGRADEシステムと国内ガイドラインとの比較・変換という点でも容易なものではない。

さらには、診断研究についてのGRADEシステムが近々報告される予定であるが、同じく最近発表された国内の診断研究に関するエビデンス分類<sup>66</sup>などと比較検討する必要もあるかもしれない。

いずれにしても、エビデンスの質の評価や患者因子を考慮した2段階推奨度判定というGRADEシステムの採用検討のためにも、早急に本システムについての誤解を解消すべきと思われる。近々、国内の診療ガイドライン作成手順の改訂版が出る予定であるが、国際的な流れを考慮した、より単純・明瞭で、医療の最前線での利用性が高いものとして、統一性のあるシステムとの協調を考慮したものにすべきと思われる。

#### 8. 最後に

本内容は、国内診療ガイドライン作成などへのGRADEシステム採用を強要するものではなく、純粋に本システムの適切な理解と普及を目的としたものであ

る。なお、さらなる理解のため、現FAQ和訳を、GRADE working group HP、および下記サイトで関連情報を公開紹介する。

<http://homepage3.nifty.com/aihara/>

参考文献

- 1) Guyatt G, et al: Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992; 268: 2420-5.
- 2) Canadian Task Force on the Periodic Health Examination. *CMAJ* 1979; 121: 1193-254.
- 3) National Guideline Clearinghouse: [www.guidelines.gov/browse/browse.aspx](http://www.guidelines.gov/browse/browse.aspx)
- 4) Center for evidence-based medicine. Levels of evidence and grades of recommendation: <http://www.cebm.net/index.aspx?o=1025>
- 5) Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1986; 89 (suppl. 2): 2-3S.
- 6) Cook DJ, Guyatt GH, et al. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1992; 102 (suppl. 4): 305-11S.
- 7) Hadorn DC, Baker D. Development of the AHCPR-sponsored heart failure guideline. *Jt Comm J Qual Improv* 1994; 20: 539-54.
- 8) Gyorkos TW et al. An approach to the development of practice guidelines for community health interventions. *Can J Public Health* 1994; 85 (suppl. 1): S8-13.
- 9) Guyatt GH, Sackett DL et al. Users' guide to the medical literature IX: a method for grading health care recommendations. *JAMA* 1995; 274: 1800-4.
- 10) Eccles M, et al. North of England evidence based guidelines development project: methods of guideline development. *BMJ* 1996; 312:760-2.
- 11) US Preventive Services Task Force. Guide to clinical preventive services. 2nd ed. Baltimore: Williams and Wilkins, 1996:xxxix-lv.
- 12) McGettigan P. et al. The effects of information framing on the practices of physicians. *J Gen Intern Med* 1999; 14: 633-42.
- 13) Guyatt GH, et al. Practitioners of evidence based care. Not all clinicians need to appraise evidence from scratch but all need some skills. *BMJ*. 2000; 320: 954-5.
- 14) Roman SH, Silberzweig SB, Siu AL. Grading the evidence for diabetes performance measures. *Eff Clin Pract* 2000; 3: 85-91.
- 15) Woloshin S. Arguing about grades. *Eff Clin Pract* 2000; 3: 94-5.
- 16) Edwards A. et al. The effectiveness of one-to-one risk communication interventions in health care: a systematic review. *Med Decis Making* 2000; 20: 290-7.
- 17) Guyatt G, et al. Grades of recommendation for antithrombotic agents. *Chest* 2001; 119: 3S-7S. [www.chestjournal.org/content/vol119/1\\_suppl/](http://www.chestjournal.org/content/vol119/1_suppl/)
- 18) Guyatt G, et al. Grades of recommendation for antithrombotic agents. *Chest* 2001; 119: 3-7S.
- 19) Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001; 323: 334-6.
- 20) Fuster V, et al.: ACC/AHA/ESC Guidelines for the Management of Patients with Atrial Fibrillation. *Circulation*. 2001; 104: 2118-50.
- 21) 診療ガイドラインの作成の手順 ver. 4.3 (福井・丹後 2001) <http://www.niph.go.jp/glg/4.3rev.htm>
- 22) Brown JP, Josse RG; Scientific Advisory Council of the Osteoporosis Society of Canada. 2002 clinical practice guidelines

- for the diagnosis and management of osteoporosis in Canada. *CMAJ* 2002; 167 (suppl. 10): s1-34.
- 23) Wright PJ, et al. Managing acute renal colic across the primary-secondary care interface: a pathway of care based on evidence and consensus. *BMJ*. 2002; 325: 1408-12.
- 24) Upshur RE: Are all evidence-based practices alike? Problems in the ranking of evidence. *CMAJ*. 2003; 169(7): 672-3.
- 25) Schunemann HJ, Best D, Vist G, Oxman AD, for the GRADE Working Group: Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. *CMAJ* 2003; 169 (7): 677-680.
- 26) Upshur REG: Are all evidence-based practices alike? Problems in the ranking of evidence. *JAMC* 2003; 169: 672-673.
- 27) DISCERN: Health information on the Internet. Available: <http://www.discrim-genetics.org/index.php>
- 28) Kunz R: "What's in the black box?". *Chest* 2006, 129: 7-10.
- 29) 津谷喜一郎他:「エビデンス」と「お勧め度」厚生労働省特定疾患調査研究「特定疾患に関する緊急研究 EBM 導入研究班」(主任 福井次矢)平成10年度報告書, 1999, p49-62.
- 30) 中山健夫: EBMの手法を用いた診療ガイドライン:日本における取り組み・課題と展望. *日本補完代替医療学会誌*, 2005; 2: 113-25.
- 31) Atkins D, Best D, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004; 328: 1490-5.
- 32) Petrisor BA, et al. Grading the evidence: levels of evidence and grades of recommendation. *Injury*. 2006; 37: 321-7.
- 33) UpToDate. [http://www.uptodate.com/service/editorial\\_policy.asp](http://www.uptodate.com/service/editorial_policy.asp)
- 34) WHO: Rapid Advice Guidelines on pharmacological management of humans infected with avian influenza A (H5N1) virus. [http://www.who.int/medicines/publications/WHO\\_PSM\\_PAR\\_2006.6.pdf](http://www.who.int/medicines/publications/WHO_PSM_PAR_2006.6.pdf)
- 35) Ontario Health Technology Advisory Committee (OHTAC): Gastric electrical stimulation. Health technology policy assessment. OHTAC Recommendation 2006: [http://www.health.gov.on.ca/english/providers/program/ohtac/tech/reviews/pdf/rev\\_ges\\_081806.pdf](http://www.health.gov.on.ca/english/providers/program/ohtac/tech/reviews/pdf/rev_ges_081806.pdf)
- 36) Wierman ME, Basson R, Davis SR, et al. Androgen therapy in women: an Endocrine Society Clinical Practice guideline. *J Clin Endocrinol Metab*. 2006 Oct; 91(10):3697-10. Epub 2006 Oct 3.
- 37) Uhlig K, Macleod A, et al: Grading evidence and recommendations for clinical practice guidelines in nephrology. A position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney International* 2006; 70: 2058-2065. <http://www.nature.com/ki/journal/v70/n12/abs/5001875a.html>
- 38) Goldsmith M et al: Synthesizing quantitative and qualitative research in evidence-based patient information. *J Epidemiol Community Health*. 2007 Mar; 61(3):262-70.
- 39) Dahm P. et al: Evidence-based clinical practice guidelines for prostate cancer: the need for a unified approach. *Current Opinion in Urology* 2007; 17: 200-207.
- 40) Guirguis-Blake J. et al: Current Processes of the U.S. Preventive Services Task Force: Refining Evidence-Based Recommendation Development. *Ann Intern Med*, Jul 2007; 147: 117 - 122. <http://www.annals.org/cgi/reprint/147/2/117.pdf>
- 41) AHRQ: Value of the Periodic Health Evaluation: Evidence Report/Technology

- Assessment, No. 136. 2007. <http://www.ahrq.gov/downloads/pub/evidence/pdf/phe/phe.pdf>
- 42) Atkins D, Briss PA, et al. Systems for grading the quality of evidence and the strength of recommendations II: Pilot study of a new system. *BMH Health Services Research* 2005; 5:25 doi: 10.1186/1472-6963-5-25.
- 43) Atkins D, Eccles M, et al. Systems for grading the quality of evidence and the strength of recommendations I: Critical appraisal of existing approaches The GRADE Working Group. *BMH Health Services Research* 2004; 4 :38 doi:10.1186/1472-6963-4-38.
- 44) Guyatt G, Vist G, et al: An emerging consensus on grading recommendations? *ACP J Club*. 2006; 140: A11-12.
- 45) Guyatt G, Gutterman D, et al: Grading strength of recommendations and quality of evidence in clinical guidelines: Report from an American College of Chest Physicians Task Force. *Chest* 2006; 129: 174-187.
- 46) Schunemann HJ, Guyatt G. et al. An official ATS statement: Grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med*. 2006 ;174(5):605-14.
- 47) Karanicolas PJ, et al: Systematic review and meta-Analysis of pylorus-preserving versus standard Whipple pancreaticoduodenectomy for pancreatic or perianipulmonary cancer. *Ann Surg Oncol*. 2007; 14:1825-34.
- 48) Guyatt G, Oxman AD, Falck-Ytter Y, et al: GRADE: An emerging consensus on rating quality of evidence and strength of recommendation. Series-1/5 *BMJ* submitted.
- 49) Guyatt G, Oxman AD, Falck-Ytter Y et al. What is 'quality of evidence' and why is it important to clinicians? *BMJ* submitted.
- 50) Guyatt G, Oxman AD, Falck-Ytter Y, et al, Going from evidence to recommendations. *BMJ* submitted.
- 51) Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Williams J, et al. GRADEing the quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* submitted.
- 52) Guyatt G, Oxman AD, et al : Grading recommendations: Incorporating considerations of resources use. *BMJ* submitted.
- 53) Rossouw JE et al (Writing Group for the Women's Health Initiative Investigators). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA*. 2002 Jul 17;288(3):321-33.
- 54) Chong BH et al (Australian HIT Study Group). Prospective randomised open-label comparison of danaparoid with dextran 70 in the treatment of heparin-induced thrombocytopenia with thrombosis: a clinical outcome study. *Thromb Haemost*. 2001 Nov;86(5):1170-5.
- 55) Alonso-Coello P, Zhou Q, Martinez-Zapata MJ, Mills E, Heels-Ansell D, Johanson JF, Guyatt G. Meta-analysis of flavonoids for the treatment of haemorrhoids. *Br J Surg*. 2006 Aug;93(8):909-20.
- 56) Thompson DC, Rivara FP, Thompson R. Helmets for preventing head and facial injuries in bicyclists. *Cochrane Database Syst Rev*. 2000; (2):CD001855.
- 57) Cannegieter SC, Rosendaal FR, Briet E. Thromboembolic and bleeding complications in patients with mechanical heart valve prostheses. *Circulation*. 1994 Feb; 89(2):635-41
- 58) Guyatt G et al. Addressing resource allocation issues in recommendations

- from clinical practice guideline panels: suggestions from an American College of Chest Physicians Task Force. *Chest* 2006; 129(1): 182-7.
- 59) 津谷喜一郎, 辻香織: GRADEによるエビデンスの質とお勧め度の強さ、「根拠に基づく診療ガイドライン」の適切な作成・利用・普及に向けた基盤整備に関する研究: 患者・医療消費者の参加推進に向けて(中山班, 平成16年度 総括・分担研究報告書)p6-9.
- 60) 津谷喜一郎, 中山健夫, 島村治子. エビデンスの質とお勧め度のグレーディング. 薬理と臨床, 2005; 33:1241-54.
- 61) 正木朋也, 津谷喜一郎: エビデンスに基づく医療(EBM)の承諾と方向性: 保健医療評価に果たすコクラン共同計画の役割と未来. 日本評価研究 2006; 6: 3-20.
- 62) 小児急性中耳炎診療ガイドライン(2006) 編集: 日本耳鼻科学会, 日本小児耳鼻咽喉科学会, 日本耳鼻咽喉科感染症研究会(金原出版 2006年11月)
- 63) 科学的根拠に基づく褥瘡局所治療ガイドライン(2007)日本褥瘡学会, 日本褥瘡学会編集(株 昭林社出版, 2005年9月)
- 64) 栗山真理子: 「診療ガイドライン作成過程への患者・支援者参画のためのガイドライン」(PIPG)作成に至るまでの動機と経緯について. 厚生労働科学研究費補助金 医薬安全・医療技術評価総合研究事業, 「根拠に基づく診療ガイドライン」の適切な作成・利用・普及に向けた基盤整備に関する研究: 患者・医療消費者の参加推進に向けて(中山班, 平成18年報告書). P126-128.
- 65) 中山健夫: 「根拠に基づく診療ガイドライン」の適切な作成・利用・普及に向けた基盤整備に関する研究: 患者・医療消費者の参加推進に向けて(平成18年報告書). P126-128.
- 66) Furukawa TA, Guyatt GH. Et al. Association between unreported outcomes and effect size estimates in Cochrane meta-analyses. *JAMA* 2007; 297: 468-70.
- 67) Higgins JP et al, Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557-560 (6 September), doi:10.1136/bmj.327.7414.557
- 68) 森實敏夫. 国民の視点に基づく適切な診断過程の確立に関する研究. 「診断に関する研究のエビデンスレベル分類法」. 厚生科学研究費補助金(医療技術評価総合研究事業)平成18年度総括・分担研究報告書. 2007, P14-23
- 69) 浜六郎, 喘息吸入剤フルチカゾン(フルタイド)による急性副腎不全, *TIP「正しい治療と薬の情報」* 19(3):25-29, 2004
- 70) 浜六郎, フルチカゾンは常用量でも副腎不全を起しうる, *TIP「正しい治療と薬の情報」* 19(5):48-51, 2004

## CAPSULE

### インターフェロンによる血栓性血小板減少性紫斑病

第1例. C型肝炎を有する28才男性が, SCベグインターフェロン $\alpha$  150  $\mu$ g/週とリバビリンによる治療を開始した. 3週間後, 細小血管性溶血性貧血, 褐色尿, 血小板減少症, 腎機能低下で来院した. Hb 8.8g/dl, 血小板数3.4万/mm<sup>3</sup>, LDH 2320 U/L, クレアチニン2.1 mg/dl, T.Bil 1.8mg/dlで, 血栓性血小板減少性紫斑病(TTP)と診断し, 血漿交換およびプレドニゾンによる治療を開始して, TTPは完全寛解に入った. 第2例. 多発性硬化症を有し, 9年間にわたってSCインターフェロン $\beta$ -1b 80万U/48hr投与を受けている50才

男性が, 全身倦怠感, 頭痛, 一時的な下痢, 腹痛および進行性の脱力で来院した. 微小血管症性貧血, 腎機能低下および血小板減少が見られた. Hb 5.6g/dl, 血小板数4.7万/mm<sup>3</sup>, LDH 1,090 U/L, クレアチニン2.6mg/dl, T.Bil 2.1 mg/dlであった. インターフェロン $\beta$ -1bを中止したが一般状態は悪化し, 視力低下, 特発性の皮下出血が出現, TTPと診断して, 早めの血漿交換を行った. 20日間の治療の後, TTPは完全寛解に入った. 第3例. 慢性C型肝炎を有し, 5ヶ月前からSCベグインターフェロン $\alpha$  150  $\mu$ /weekとリバビリンによる治療を受けている41才男性が, 頭痛, 全身倦怠感, 黒色便, 結膜出血, 血小板減少および微小血管症性貧血で来院した. Hb 4.4g/dl, 血小板数0.4万/mm<sup>3</sup>, LDH 771 U/L, T.Bil 3.3mg/dlであった. TTPと診断され, ベグインターフェロン $\alpha$ とリバビリンを中止, プレドニゾンと, 連日の血漿交換を開始した. 13回の血漿交換の後, 完全寛解に入った.

Serrano A et al. *Medicina Clinica* 128: 276, 2007

### クエチアピン(セロクエル)による糖尿病性ケトアシドーシスと高トリグリセリド血症

双極性障害を有する34才女性が, 2ヶ月前からクエチアピンとバルプロ酸Naを服用していたが, 急激な嘔吐と腹痛で入院した. 検査で, ケトン尿, 低Na血症および糖尿病性ケトアシドーシスを認めた. HCO<sub>3</sub><sup>-</sup> 9.7mEq/L, 血糖493mg/dl, トリグリセリド740mg/dl, base excess - 21.6, PH 7.12であった. 糖尿病ケトアシドーシスに対する標準的治療を行ったが, 代謝異常が是正された後も, 激しい腹痛が持続し, 急性腹症の徴候を示し, 腹部CTは, 急性膵炎を示唆した. 保存的治療で症状は軽減し, 高トリグリセリド血症に対して, フェノフィブラートの内服を開始した. 退院後, トリグリセリドは改善し, 血糖はインスリンでコントロールされている.

Dissanayake S et al. *Diabetic Med.* 24 (Suppl.1) :107 abstr.P269,2007

### タモキシフェン(ノルバデックス)による子宮内膜癌

高血圧症と糖尿病を有している71才女性が, 10年前から乳癌に対してタモキシフェン10mgを1日2回内服していたが, 臍出血で来院した. 超音波検査で, 子宮内膜が厚さ21mmで, 異常な形状を示した. 子宮は, 妊娠3ヶ月の大きさに腫大していた. パパニコロー抹標本で, 異形の腺管細胞を認め, 子宮内膜癌を示した. 開腹して, 子宮全摘と大網の部分切除を行った. 腫瘍の組織学的分析では, 腺管状乳頭状パターンの高分化腺癌を示した. 免疫組織化

学染色では, プロゲステロン受容体に強陽性を示したが, エストロゲン受容体には陰性であった.

Wu C-J et al. *Taiwan. J. Obs t. & Gynecol.* 46: 88,2007

### デフェロキサミン(デスフェール)の子宮内曝露による鉄欠乏性貧血

サラセミアを有する女性が, デフェロキサミンのSC投与を週5回受けていて, 妊娠1ヶ月前に中止した. 妊娠6ヶ月目に, デフェロキサミン治療を再開し, 出産まで2ヶ月間続け, 妊娠35週目に, 帝王切開で正常新生児を出産した. 臍帯血の検査では, Hb 14.5g/dl, Ht 54.5%であった. 新生児の血液検査で, フェリチン42ng/ml, 血清鉄/総鉄結合能53/222  $\mu$ g/dlであった. 生後3ヶ月目, 鉄補充食を受けていたにも関わらず, フェリチン14 ng/ml, 血清鉄/総鉄結合能80/372  $\mu$ g/dlであった. この母親は, 後に双子を出産したが, この時は, デフェロキサミンの子宮内曝露はなかった. 生下時, 2人の新生児のフェリチン・レベルは227および203ng/mlで, 血清鉄/総鉄結合能30/182および27/203  $\mu$ g/dlであった. 2回の妊娠の間の, 鉄貯蔵の状態とフェリチン値の違いは, 妊娠末期のデフェロキサミン投与の有無に関係していると考えられる.

Pearson HA et al. *J. Pediat. Hematology/Oncology* 29: 160, 2007

### エポエチン $\beta$ (エポジン)による赤芽球癆

RA反応(-)の関節リウマチを有する79才男性が, 慢性腎不全に由来する貧血に対して, SCエポエチン $\beta$  4000 IU週3回の治療を開始した. 2ヶ月経って, Hbレベルが上昇したため, エポエチン $\beta$  4000 IU週1回に減らしたところ, 11週間後, Hbレベルが8.5g/dlに低下し, エポエチン $\beta$  5000 IU週3回に増量したが, 臨床的に改善が見られなかった. その後, 輸血を受け, エポエチン $\beta$  10000 IU週3回に増量したが, Hbレベルは6.7g/dlに下がって輸血を受けた. 網状赤血球は見られず, 骨髓穿刺で, 赤芽球系が選択的に少なく赤芽球癆が疑われ, エポエチン $\beta$ を中止した. 抗エリスロポエチン抗体が強陽性を示し, プレドニゾン内服を開始したが, Hbレベルは依然として低く, 2週間毎に輸血を行った. 2ヶ月経って, 重篤な出血傾向が出現, APTTおよびプロトロンビン時間の延長, 血小板2.6万/mm<sup>3</sup>, D-dimerレベル高値(>1005), フィブリノゲン低値(<50ng/dl)と, DICが明らかであった. ヘパリン投与にも関わらず, 改善が見られず, 心血管の合併症を来して死亡した.

Manenti L et al. *Nephrology Dialysis Transplantation* 22: 1465, 2007