

益)、Harm (害) と、Burden (負担)、Cost (費用) のバランスを評価して、推奨の強さを決めるようになっている。この点は長所でもあるが、後述するように、問題を引き起こす可能性もあると考えられる。

#### D 考察

わが国では、福井・丹後ら<sup>1</sup>の厚生労働省研究班が作成した「診療ガイドラインの作成の手順 ver. 4.3」が標準的手法として用いられてきた。なお、2007年にその改訂版<sup>2</sup>に該当するものが発表されている。わが国で作成された診療ガイドラインはこれに基づいているものが多いが、エビデンスの質の分類も推奨度分類も、実際にはそれぞれ異なる方法が採用されている。その理由は、複合的と考えられるが、一番大きな理由は疾患特異性と考えられる。すなわち、分野によって、研究の数には大きな差があり、研究の質も異なる。一般的に、患者数が多い疾患は、論文数が多く、ランダム化比較試験も多く、サンプルサイズの大きな研究が多いが、患者数の少ない疾患は、ランダム化比較試験に乏しく、質の高い研究は少ない。また、薬物治療が主体の疾患と、外科的治療が主体の疾患でも論文の数、質が異なる。したがって、それぞれの対象疾患によって、エビデンスの質、推奨度の分類法が異なるのは当然ともいえる。

GRADE システムは大きな進歩とはいえ、未完成であり、BMJ に発表が予定されていた診断法などに関する論文もまだ出版されていない。今後の変更や改良の可能性が残されているのが現状であり、

現時点で取り入れるとすると、推奨度の分類だけになる可能性がある。しかし、推奨度の分類だけを取り入れても、推奨を支持するエビデンスの評価が、推奨ごとに、あるいは診療ガイドラインごとに異なっていれば、相互に直接比較することはできない。

American College of Chest Physicians (ACCP) は、2006年に、当時の会長と副会長名で、診療ガイドラインの作成手法として GRADE を採用することを発表している<sup>36</sup>。この声明の中で、ACCP の決定が、GRADE システムの保証をするものではないことも述べられている。この点は、他の学会・組織でも同様であり、それぞれの組織が、それぞれの組織の意思決定の手順を踏んで、GRADE システムを今後の診療ガイドライン作成手法としての採用を決定したことを発表している。

ACCP は GRADE システムのエビデンスの分類を、Low と Very low を1つにし、Low として、4段階から3段階に改変している<sup>37</sup>。同様の改変は、American Thoracic Society (ATS)<sup>38</sup>、UpToDate も行っている。腰痛に関する診療ガイドライン<sup>32</sup>は ACP と American Pain Society の腰痛診療ガイドライン作成委員会が作成したものであるが、当初は USPSTF の A, B, C, D, I の推奨度分類を用い、途中から GRADE システムを採用したため、主文は GRADE システムによる分類が抱えているが、文末に USPSTF の推奨度分類が併記されている。この診療ガイドラインでも、エビデンスの質の分類は High, Moderate, Low の3段階に分類し、Benefit が Risk を明らかに上回るか上回らない場合に Strong

recommendation、BenefitとRiskとBurdenが finely balanced の場合に Weak recommendation としている。さらに、正味のBenefitとHarmを決めるのに不十分 (insufficient) なエビデンスの場合、I recommendation としている。

したがって、エビデンスの質を High : A, Moderate: B, Low: C とし、強い推奨を1、弱い推奨を2とすると、1-A, 1-B, 1-C, 2-A, 2-B, 2-C に分類され、さらにすべき方向と、しない方向で別れるので、9種類、あるいは12種類の推奨の分類が形成される。

上記の、I recommendationはThe USPSTFの推奨度分類<sup>10</sup>のIに相当するもので、解説には、次のように書かれている：“The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.” すなわち、「USPSTFは現在のエビデンスはその医療のBenefitとHarmのバランスを評価するのに不十分であると結論付けた。エビデンスがないか、質が悪いか、矛盾しており、BenefitとHarmのバランスを決定することができない」としている。このような、推奨度分類を必要とする疾患もあることは事実であり、GRADEシステムが疾患特異性に合わせて改変されて利用されることを否定することはできないであろう。

また、腰痛のガイドライン<sup>32</sup>では、最初に効果の程度 (Magnitude) の定義をしている。これも、Small/slight, Moderate,

Large/substantialの3段階分類であるが、たとえば、Large/substantialのSize of Effectの定義は、“Pain scales: Mean ·20-point improvement on a 100-point VAS or equivalent Back-specific functional status: Mean ·20-point improvement on the ODI, ·5 points on the RDQ, or equivalent All outcomes: SMD ·0.8” となっている。このように、その疾患で臨床的に意義があると考えられる効果を事前に定義し、それに基づいてエビデンスの評価を行う手法が、今後他の疾患でも採用される可能性がある。この効果の大きさに関しては、GRADEでは、エビデンスの質をグレードアップする要素として、リスク比 $>2$ あるいは $<0.5$ の場合は、+1、それぞれ $>5$ あるいは $<0.2$ の場合は、+2とスコアを増加させることが提案されている。疾患によって、これらの閾値を変える必要があったとしてもそれは当然といえる。

さて、現時点では、診療ガイドライン作成法としてはGRADEシステムとUSPSTFのシステムの2つが最も有力な方法といえる。USPSTFのシステムでは、エビデンスのグレーディングはCertainty (確実性) のレベルとして、High, Moderate, Lowの3段階に分類している。ただし、その分類の基準はGRADEシステムと同じではない。表1, 2にUSPSTFの分類を、表2, 3にGRADEシステムに基づく分類を示す。

さて、GRADEシステムに関して抱く疑問の中で、最大のものは、エビデンスの質と推奨の強さが分離されているという点である。確かに、これら2つの概念が分離されている方が、ガイドラインの作成は

容易になるであろう。BenefitがHarmを明らかに上回っている場合に、強い推奨はとなるが、エビデンスの質がLowまたはVery lowであっても、強い推奨とすることが可能である。この場合、Benefitを受ける患者の割合が、Harmをこうむる患者の割合より、相対的にはるかに高いとしても、それを支持するエビデンスの質が悪い場合に、強い推奨とするには、別の根拠が必要になる。たとえば、作成委員会のコンセンサスの形成手順を別途定めて、それにしたがって決定するような方法が必要になる。この点については、GRADEシステムは明らかにしていない。

古くからルチンに行われている介入で、特に確固たるエビデンスはないようなものを強い推奨とする場合を想定するのであれば、理解できるが、エビデンスの質の悪いものを強く推奨できるということを許容すれば、“根拠に基づく診療ガイドライン”とは言いにくいのではないだろうか。

また、BenefitとHarmは、適切な研究で証明されないと、BenefitがあるともHarmがあるとも言えないのではないか。すなわち、臨床研究が一定の条件を満たした患者で、ある介入でアウトカムが改善することを証明した場合に、Benefitがあるといえるのであって、エビデンスの質が低くても、Benefitが確実にあると判定して、強い推奨ができるとは考えにくい。

BenefitとHarm、あるいはHarm, Burden, Costの3つをあわせて、Downsides、のバランスを評価して推奨を決めるということについては、USPSTFのシステムもGRADEシステムも同じである。いずれも

Certainty（確実性）という言葉でこのバランスの程度を表している。BenefitとHarmは率の値、あるいはリスク比やNNT（Number needed to treat）やNNH（Number needed to harm）といった効果指標の値としてその程度が表される。しかし、Benefitを証明する研究と、Harmを証明する研究では、後者が数少なく（sparse data）、普通起きる率も低いので、サンプルサイズが大きくないと評価が困難なことが多い。したがって、多くの場合、Benefitについては、Certaintyが高くHarmについては、Certaintyが低くなる。したがって、両者のバランスというよりも、BenefitとHarmをそれぞれ個別に評価して、それぞれのCertaintyに基づく判断が望ましいのではないか。

また、悪性腫瘍のように生存がアウトカムとして重要な疾患の場合であれば、副作用による死亡がある一定の割合で起きるような治療法でも容認されうるが、高血圧や糖尿病のような疾患の場合には、低率であっても死亡のような重篤な害が発生するような治療はBenefitが大きくても推奨はできないであろう。このように、BenefitとHarmの評価には疾患特異性があり、また今後の課題であるが、患者の価値観をどのように取り入れるかも課題のひとつである。

推奨度の決定におけるもう一つの疑問は、個々の介入に関して、GRADEシステムにしたがって、推奨度を決めた場合、従来の介入をふくめた他の介入との比較が行われずに、強い推奨となるものが出てくることである。たとえば、1-Aの推奨の治療法が複数提示された場合、診療ガイ

ドラインの利用者は選択に迷うであろう。もし、2つの治療法を直接比較したランダム化比較試験があって、差が無いことが同等性試験や非劣性試験で証明されているような場合であれば、結果として大きな問題は生じないと考えられるが、もし、優劣が証明されている場合には、それを前提に推奨を決定する必要があると思われる。したがって、エビデンスの評価から、BenefitとHarmの評価、そして、推奨の決定に至るプロセスに、他の介入との比較のステップが必要ではないだろうか。

推奨の強さを2段階に単純化することも利点がある一方、不都合もある。たとえば、1-Aと1-Cあるいは1-Bを同じ強い推奨であっても、利用者が同じ強さと受け取るかどうか疑問である。利用者の傾向としては、同じ強い推奨であっても、1-Aと1-Cであれば、前者を実施する割合が多くなるのではないかと。結果として、2段階の推奨度が6段階になる可能性があり、たとえば、A, B, C, D, E, Fの6段階に分けて、A, B, Cは強い推奨、D, E, Fは弱い推奨というのと同じになってしまう可能性が考えられる。

わが国では、USPSTFに準じた、A, B, C, DあるいはA, B, C, D, I、最近では、CをC1とC2の2つに分けた、A, B, C1, C2, Dの推奨度分類が用いられることが多いが<sup>2</sup>、診療ガイドラインの利用者はA, Bに分類された推奨の実施には詳細な記述を読まなくても賛同するが、Cにどのように対処するかが大きな関心事で、Cの推奨がよく読まれることになる。そのような要請にこたえる形で、C1, C2の亜分類が作られた。USPSTFの分類を元にするこれらの分

類と、GRADEシステムの分類のどちらが、利用者にとって、理解しやすいかは、今後検討が必要であろう。

GRADEシステムを含むすべての診療ガイドライン作成システムには、内在する共通の問題がある。それは、さまざまな段階で判定 (Judgment) が必要になるということである。判定は、評価者個人個人で異なる可能性が高い。たとえば、GRADEシステムにおいて、エビデンスの評価のスタートは研究デザインの判定である。このステップで、論文に” randomized” と書いてあれば、ランダム化比較試験と判定するのか、中身まで吟味するのか、” randomized” と書かれていても、実際にどのように実行したのか記述がない場合は、どうするのかといった問題が発生し、評価者がどの段階でどのような判定を下すかは個々に異なる可能性がある。McAlisterら<sup>28</sup>の論文はこのような変動の可能性を示唆しているのかもしれない。

GRADEシステムを用いる場合でも、たとえば、ランダム化比較試験と判定した後、研究の質を評価する際に、Study limitationsやConsistencyの判定などいずれの判定も評価者によって異なる可能性がある。GRADEシステムのすぐれている点は、評価の項目や手順を明示的に提案したことにあるのは、間違いないが、それでも変動する人間の判定という要素は排除するのは困難である。

それでは、判定をできるだけ、再現性高く、適切に行うにはどうすべきなのか？Treadwellらの提案が注目に値する。かれらは、USPSTFとGRADEのシステムを取

り上げ、評価した上で、エビデンスの質の、判定の変動による問題をできるだけ小さくする方法を新たに提案している。また、評価すべきエビデンスの属性として、stability(安定性)とstrength(強さ)という概念を提示している。定量的に「どの程度有効か？」に答える概念として、stabilityをあて、効果指標の大きさとそのばらつきで表される概念としている。定性的に、「有効かどうか？」に答える概念として、strengthをあて、効果指標の値の大きさだけでなく、方向性によって表される概念としている。これら2つの概念は独立しているものではない。

エビデンスの属性として、Meaningfulness, Relevance, Validity, Reliabilityの4つの概念があることが報告されており、医学文献の評価に用いることを提案した<sup>40,41</sup>。Relevanceは臨床的意義のことで、GRADEシステムではDirectnessの一部に対応する。Validityは妥当性であるが、同じく、Study designとStudy qualityにほぼ対応し、Directnessにも一部対応する。Reliabilityは同じく、Consistencyにほぼ対応する。Treadwellらの提案からは、エビデンスの属性として、未知あるいは異なる概念が存在する可能性を示唆するようにも思える。

GRADEシステムでは、アウトカムのDirectnessについては、1から9までの整数でスコア化し、Critical, Important but not-critical, Not-importantの3段階に分類することも提案されている<sup>8</sup>。この点は、本斑会議のテーマに直接関わる問題である。アウトカムの重要性につい

て、患者や介護者の価値観を取り入れることは、非常に重要である。1人の患者の経験や意見に基づいて決めるのではなく、患者や介護者を対象にした、何らかの調査研究に基づいて決める必要があるのではないかと考える。

推奨の強さの決定に影響する因子については、著者の私見では、その介入により、「得られるものが負担、害を上回ることをどれだけ確実に証明しているか」であり、その証明に影響する因子は、1) アウトカム、2) 効果の大きさ、3) ベースラインリスク、4) 外的妥当性、5) エビデンスの強固さ：妥当性(研究デザイン、方法)、実行コンプライアンス、確実性(効果指標の信頼区間)、6) 害の種類、頻度と程度、7) 負担(費やされる時間と身体的・精神的負担)、8) 費用(下流も含めた円で表される額) 9) 他の介入との比較、があげられる(図2)。

最後に、ACCPがGRADEシステム採用決定の記事を掲載した同じ号に、その解説の論文だけでなく、“資源の分配” Resource allocation”の問題に関するGuyattら<sup>42</sup>の論文が掲載されており、推奨度を決める際に、その診療ガイドラインの主要な推奨については、下流も含めた費用を考慮することが望ましいと述べられている。資源の分配とは、限りある医療費をだれに分配するのか、予後の期待できない、高齢者に多くの費用を使うのと、これらからの人生がまだ長い、子供に使うのとどちらが社会全体として、公平・公正なのかといった問題に答えることである。今後、診療ガイドラインの作成に医療経済学の専門家の参加を求め

る必要が出てくるであろうし、現状で必要なマンパワーがあるのか、今後の課題であろう。

## E 結論

診療ガイドラインの作成のためのGRADEシステムについて考察した。

- 1) GRADEシステムは診療ガイドライン作成法を明示的に示しているが、主観的な判定を除去することはできない。
- 2) エビデンスの質と推奨の強さのグレーディングを分けている点が、長所であるが、短所でもある。

- 3) 疾患特異性にどの程度対応できるかは今後の実績を見る必要がある。

## F.健康危険情報

なし。

## G.研究発表

1. 論文発表  
なし
2. 学会発表  
なし。

## H.知的財産権の出願・登録状況

なし。

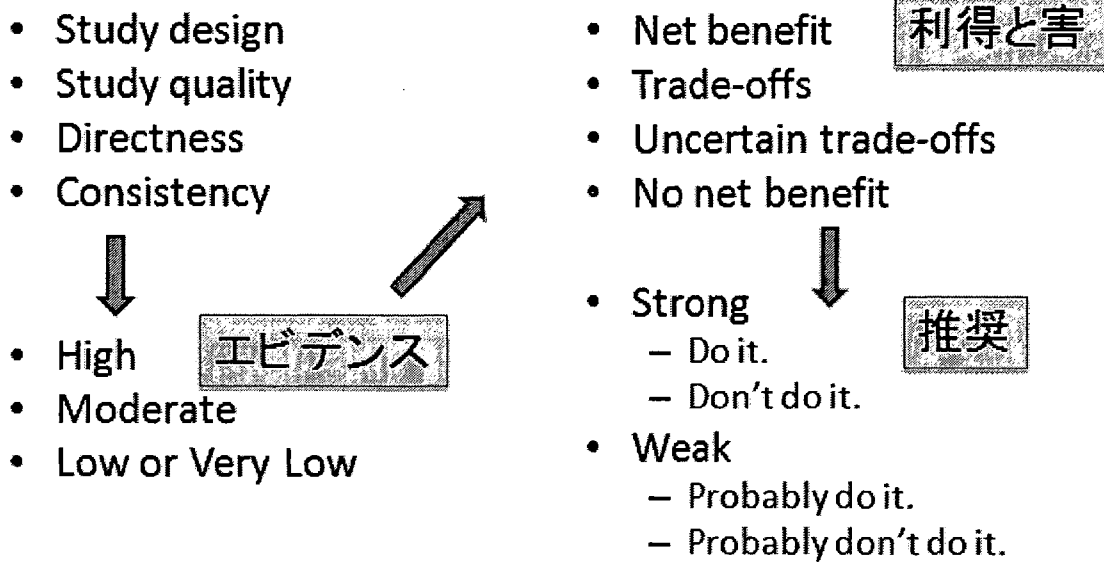


図 1. GRADE システムの概略。文献 6 に基づく。

表 1. U. S. Preventive Services Task Force (USPSTF)による推奨の分類。(文献 10 より)。

Grade	Definition	Suggestions for Practice
A	The USPSTF recommends the service. There is high certainty that the net benefit is substantial.	Offer or provide this service.
B	The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial.	Offer or provide this service.
C	The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small.	Offer or provide this service only if other considerations support offering or providing the service in an individual patient.
D	The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits.	Discourage the use of this service.
I	The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.	Read the clinical considerations section of USPSTF Recommendation Statement. If the service is offered, patients should understand the uncertainty about the balance of benefits and harms.



表2. U.S. Preventive Services Task Force Levels of Certainty Regarding Net Benefit (文献10より)。

Level of Certainty	Description
High	<p>The available evidence usually includes consistent results from well-designed, well-conducted studies in representative primary care populations. These studies assess the effects of the preventive service on health outcomes. This conclusion is therefore unlikely to be strongly affected by the results of future studies.</p>
Moderate	<p>The available evidence is sufficient to determine the effects of the preventive service on health outcomes, but confidence in the estimate is constrained by such factors as:</p> <ul style="list-style-type: none"> <li>the number, size, or quality of individual studies</li> <li>inconsistency of findings across individual studies</li> <li>limited generalizability of findings to routine primary care practice</li> <li>lack of coherence in the chain of evidence</li> </ul> <p>As more information becomes available, the magnitude or direction of the observed effect could change, and this change may be large enough to alter the conclusion.</p>
Low	<p>The available evidence is insufficient to assess effects on health outcomes. Evidence is insufficient because of:</p> <ul style="list-style-type: none"> <li>the limited number or size of studies</li> <li>important flaws in study design or methods</li> <li>inconsistency of findings across individual studies</li> <li>gaps in the chain of evidence</li> <li>findings not generalizable to routine primary care practice</li> <li>lack of information on important health outcomes</li> </ul> <p>More information may allow estimation of effects on health outcomes.</p>

表 3. GRADE システムによる推奨の分類 (ACCP の例。文献 37 より。)

Balance of Benefits to Risks and Burdens				
Strength of Evidence	Benefits Outweigh Risks/Burdens	Risks/Burdens Outweigh Benefits	Evenly Balanced	Uncertain
High	1A	1A	2A	
Moderate	1B	1B	2B	
Low or very low	1C	1C	2C	2C

\*1A = strong recommendation; 1B = strong recommendation; 1C = strong recommendation; 2A = weak recommendation; 2B = weak recommendation; 2C = weak recommendation.

表 4. Quality of Evidence Scale (文献 33 より)

High	RCTs without important limitations or overwhelming evidence from observational studies*
Moderate	RCTs with important limitations (inconsistent results, methodologic flaws, indirect, or imprecise) or exceptionally strong evidence from observational studies*
Low or very low	Observational studies or case series

得られるものが負担、害を上回ることをとればだけ確実に証明しているか  
 - Benefits > Harms, Burdens, Costs

- アウトカム ⇔ 疾患特異性、患者さんの視点
- 効果の大きさ ⇔ 効果指標の大きさ
- ベースラインリスク ⇔ 疾患特異性
- 外的妥当性(一般化可能性)
- エビデンスの強固さ
  - 妥当性 ⇔ 研究デザイン, 方法
  - 実行コンプライアンス
  - 確実性 ⇔ 効果指標の信頼区間、研究間のばらつき、P値
- 害 → 種類と頻度と程度
- 負担 → 費やされる時間と身体的・精神的負担
- 費用: 下流も含める → 円で表示可能
- 他の介入・治療法との比較 ⇔ 疾患特異性

アセスメントの対象

アセスメントの対象を構成する項目  
 • Relevance  
 • Validity  
 • Reliability

得られるものにみあうか

他の介入との比較

図 2. 推奨の強さの決定に影響する因子 (私見)。

## 文献

1. 福井次矢, 丹後俊郎: 診療ガイドラインの作成の手順 ver. 4.3。2001年。
2. 福井次矢、吉田雅博、山口直人(編集): Minds 診療ガイドライン作成の手引き。2007年、医学書院。
3. 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;281:1900-5.
4. Shiffman RN, Shekelle P, Overhage JM, Slutsky J, Grimshaw J, Deshpande AM: Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. Ann Intern Med 2003;139:493-8.
5. Guirguis-Blake J, Calonge N, Miller T, Siu A, Teutsch S, Whitlock E: Current processes of the U.S. Preventive Services Task Force: Refining evidence-based recommendation development. Ann Intern Med 2007;147:117-122. PMID: 17576998
6. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, Hill S, Jaeschke R, Leng G, Liberati A, Magrini N, Mason J, Middleton P, Mrukowicz J, O'Connell D, Oxman AD, Phillips B, Schunemann HJ, Edejer TT, Varonen H, Vist GE, Williams JW Jr, Zaza S: Grading quality of evidence and strength of recommendations. BMJ 2004;328:1490.
7. 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。薬理と治療 2005;33:1241-1254.
8. 相原守夫、Guyatt G、Falck-Ytter Yngve、浜六郎: GRADE システム: エビデンスから推奨へ。The Informed Prescriber 2007;22:91-102.
9. 相原守夫、Guyatt G、Falck-Ytter Yngve: GRADE システムと SoF。Therapeutic Research 2007;28:2335-2354.
10. Barton MB, Miller T, Wolff T, Petitti D, LeFevre M, Sawaya G, Yawn B, Guirguis-Blake J, Calonge N, Harris R; U.S. Preventive Task Force: How to read the new recommendation statement: methods update from the U.S. Preventive Services Task Force. Ann Intern Med 2007;147:123-7. PMID 17576997
11. 相原内科医院のウェブ (<http://homepage3.nifty.com/aihara/>)。
12. Oxman AD, Fretheim A, Schunemann HJ: Improving the use of research evidence in guideline development:

- introduction. Health Res Policy Syst 2006;4:12. ID:17116254
13. Schunemann HJ, Fretheim A, Oxman AD: Improving the use of research evidence in guideline development: 1. Guidelines for guidelines. Health Res Policy Syst 2006;4:13. ID:17118181
14. Oxman AD, Schunemann HJ, Fretheim A: Improving the use of research evidence in guideline development: 2. Priority setting. Health Res Policy Syst 2006;4:14. ID:17134481
15. Fretheim A, Schunemann HJ, Oxman AD: Improving the use of research evidence in guideline development: 3. Group composition and consultation process. Health Res Policy Syst 2006;4:15. ID:17134482
16. Fretheim A, Schunemann HJ, Oxman AD: Improving the use of research evidence in guideline development: 5. Group processes. Health Res Policy Syst 2006;4:17. ID:17140442
17. Schunemann HJ, Oxman AD, Fretheim A: Improving the use of research evidence in guideline development: 6. Determining which outcomes are important. Health Res Policy Syst 2006;4:18. ID:17140444
18. Oxman AD, Schunemann HJ, Fretheim A: Improving the use of research evidence in guideline development: 7. Deciding what evidence to include. Health Res Policy Syst 2006;4:19. ID:17140445
19. Oxman AD, Schunemann HJ, Fretheim A: Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence. Health Res Policy Syst 2006;4:20. ID:17147809
20. Schunemann HJ, Fretheim A, Oxman AD: Improving the use of research evidence in guideline development: 9. Grading evidence and recommendations. Health Res Policy Syst 2006;4:21. ID:17147810
21. Schunemann HJ, Fretheim A, Oxman AD: Improving the use of research evidence in guideline development: 10. Integrating values and consumer involvement. Health Res Policy Syst 2006;4:22. ID:17147811
22. Edejer TT: Improving the use of research evidence in guideline development: 11. Incorporating considerations of cost-effectiveness, affordability and resource implications. Health Res Policy Syst 2006;4:23. ID:17147813
23. Oxman AD, Schunemann HJ, Fretheim A: Improving the use of research evidence in guideline development: 12. Incorporating considerations of equity.

- Health Res Policy Syst 2006;4:24.  
ID:17147814
24. Schunemann HJ, Fretheim A, Oxman AD: Improving the use of research evidence in guideline development: 13. Applicability, transferability and adaptation. Health Res Policy Syst 2006;4:25. ID:17156457
25. Oxman AD, Schunemann HJ, Fretheim A: Improving the use of research evidence in guideline development: 14. Reporting guidelines. Health Res Policy Syst 2006;4:26. ID:17156458
26. Fretheim A, Schunemann HJ, Oxman AD: Improving the use of research evidence in guideline development: 15. Disseminating and implementing guidelines. Health Res Policy Syst 2006;4:27. ID:17156459
27. Atkins D, Briss PA, Eccles M, Flottorp S, Guyatt GH, Harbour RT, Hill S, Jaeschke R, Liberati A, Magrini N, Mason J, O'Connell D, Oxman AD, Phillips B, Schunemann H, Edejer TT, Vist GE, Williams JW Jr: Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. BMC Health Serv Res 2005;5:25. PMID:15788089
28. McAlister FA, van Diepen S, Padwal RS, Johnson JA, Majumdar SR: How evidence-based are the recommendations in evidence-based guidelines? PLoS Medicine 2007;4:e250.
29. Vigna-Taglianti F, Vineis P, Liberati A, Faggiano F: Quality of systematic reviews used in guidelines for oncology practice. Ann Oncol 2006;17:691-701. PMID: 16461333
30. 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;12:18-23.
31. Wierman ME, Basson R, Davis SR, Khosla S, Miller KK, Rosner W, Santoro N: Androgen therapy in women: an Endocrine Society Clinical Practice guideline. J Clin Endocrinol Metab 2006;91:3697-710. ID:17018650
32. Chou R, Qaseem A, Snow V, Casey D, Cross JT Jr, Shekelle P, Owens DK; Clinical Efficacy Assessment Subcommittee of the American College of Physicians; American College of Physicians; American Pain Society Low Back Pain Guidelines Panel: Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. Ann Intern Med. 2007;147:478-491. PMID: 17909209
33. McCrory DC, Lewis SZ, Heitzer J,

- Colice G, Alberts WM; American College of Chest Physicians: Methodology for lung cancer evidence review and guideline development: ACCP evidence-based clinical practice guidelines (2nd Edition). *Chest* 2007;132:23S-28S. PMID: 17873158
34. Ries AL, Bauldoff GS, Carlin BW, Casaburi R, Emery CF, Mahler DA, Make B, Rochester CL, Zuwallack R, Herrerias C: Pulmonary Rehabilitation: Joint ACCP/AACVPR Evidence-Based Clinical Practice Guidelines. *Chest* 2007;131:4S-42S. PMID: 17494825
35. Schünemann HJ, Hill SR, Kakad M, Bellamy R, Uyeki TM, Hayden FG, Yazdanpanah Y, Beigel J, Chotpitayasunondh T, Del Mar C, Farrar J, Tran TH, Ozbay B, Sugaya N, Fukuda K, Shindo N, Stockman L, Vist GE, Croisier A, Nagjdaliyev A, Roth C, Thomson G, Zucker H, Oxman AD; WHO Rapid Advice Guideline Panel on Avian Influenza: WHO Rapid Advice Guidelines for pharmacological management of sporadic human infection with avian influenza A (H5N1) virus. *Lancet Infect Dis* 2007;7:21-31. PMID: 17182341
36. Baumann MH, Gutterman DD: American College of Chest Physicians evidence-based guidelines--the next generation: considering resource use and evolution to a single grading system. *Chest* 2006;129:10-2. PMID:16424406
37. Guyatt G, Gutterman D, Baumann MH, Addrizzo-Harris D, Hylek EM, Phillips B, Raskob G, Lewis SZ, Schunemann H: Grading strength of recommendations and quality of evidence in clinical guidelines: report from an american college of chest physicians task force. *Chest* 2006;129:174-81. PM ID:16424429
38. Schunemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, Ernst A, Fahy BF, Gould MK, Horan KL, Krishnan JA, Manthous CA, Maurer JR, McNicholas WT, Oxman AD, Rubenfeld G, Turino GM, Guyatt G: 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;174:605-14. PMID:16931644
39. Treadwell J, Tregear SJ, Reston JT, Turkelson CM: A system for rating the stability and strength of medical evidence. *BMC Medical Research Methodology* 2006;6:52-71.
40. 森實敏夫、福岡敏雄、中山健夫、山口直比古、宮木幸一、星佳芳：EBM 実践のための医学文献評価選定マニュアル。2004年、ライフサイエンス出版、東京。
41. 森實敏夫：診療ガイドラインおよび Minds における医学関連文献の評価選定

基準について。医学図書館  
2004;51:51-56.

42. Guyatt G, Baumann M, Pauker S,  
Halperin J, Maurer J, Owens DK,  
Tosteson AN, Carlin B, Gutterman D,  
Prins M, Lewis SZ, Schunemann H:

Addressing resource allocation issues  
in recommendations from clinical  
practice guideline panels: suggestions  
from an American College of Chest  
Physicians task force. Chest  
2006;129:182-7. ID:16424430



# What do Japanese women need to know to make decisions about treatment for benign gynaecological conditions?

Qualitative study comparing Canadian, UK and Japanese clinical guidelines

Rika Sakuma Sato (Ochanomizu Women's University)  
Takeo Nakayama (Kyoto University)

# Introduction

## Purpose of the study

- This focus group study aims to find out how guidelines can be utilized to assist more patient-centered decision-making for benign gynecological conditions, such as endometriosis and uterine fibroids, by asking gynae patients to read and evaluate existing guidelines on such conditions.

## Background 1

- Making decisions for benign long-term gynaecological conditions is difficult because of many choices.
- Clinical guidelines should enable such gynae patients and her health care providers to assess individually the risks and benefits of the treatment.
- Clinical guidelines on benign gynaecological conditions in Japan—GL for endometriosis published in 2004, none for uterine fibroids

## Background 2

- The Internet has enabled patients to find that there are clinical guidelines for endometriosis and fibroid tumors elsewhere in the world
- However, there are language barriers:  
English, medical terminology
- Designing an action research – Members of a self-help group asked a public health expert to conduct a research on the role of clinical guidelines in decision-making

## Research Questions

- How do patients respond to existent clinical guidelines? Do they find them too difficult?
- Can those clinical guidelines help patients make their treatment decisions?
- What is the most ideal format for guidelines to be utilized by patients?
- In what ways do the evaluation of guidelines by professionals and lay persons differ?

## Methods

8.9.2007BritsocMedCon

Rika S. Sato & Takeo Nakayama

7

## “deliberative” focus group

- 11 women (all Japanese) were asked to read and compare (at home) Japanese and UK guidelines for the management of endometriosis, & Canadian guideline for uterine fibroids.
- Women were later divided into two groups and were called in for a 2-3 hour discussion on the potential of these guidelines in assisting patients' clinical decision-making.

8.9.2007BritsocMedCon

Rika S. Sato & Takeo Nakayama

8

## Participants

- Discussants recruited from Tampopo, a self-help group for women with benign gynaecological problems.
- Group A: met on 8 Feb, 2006 (3hrs)  
4 fibroid patients and 2 endometriosis patient
- Group B: met on 19 Feb, 2006 (3hrs)  
3 fibroid patients, 1 endometriosis patient, 1 staff member who answers hotline calls

8.9.2007BritsocMedCon

Rika S. Sato & Takeo Nakayama

9

## Guidelines used in the study

- Society of Obstetricians and Gynaecologists of Canada, “Clinical Practice Guidelines: Management of Uterine Leiomyomas”. 2003.
- Royal College of Obstetricians and Gynaecologists, “Clinical Green Top Guidelines: The Investigation and Management of Endometriosis”. 2000.  
→ both professionally translated into Japanese
- Japan Society of Obstetrics and Gynecology, *General Rules for Clinical management of Endometriosis: Part II*. 2004: Chapter 4 “Infertility guideline” & “Pain control guideline”

8.9.2007BritsocMedCon

Rika S. Sato & Takeo Nakayama

10

## Discussions & Analysis

- Sato facilitated the discussions.
- Nakayama, an expert on GL development, joined the discussions to provide information on the medical significance of clinical guidelines.
- Discussions were audio-taped and transcribed.
- Sato coded the data using the software NVivo7 and identified emerging themes. The analysis was clarified and scrutinized by Nakayama.

8.9.2007BritsocMedCon

Rika S. Sato & Takeo Nakayama

11

## Results

8.9.2007BritsocMedCon

Rika S. Sato & Takeo Nakayama

12

## Impressions on the guidelines

- Canadian guideline for uterine fibroids was seen as most patient-friendly and useful, in terms of its language and format.
- Japanese guidelines on endometriosis-induced infertility & pain control were disliked by many, because of 1) the formal priority given to infertility treatment & 2) the inhuman impression of the title
- While Canadian guideline clearly identifies the affected women as a major decision-maker, Japanese guidelines do not include patients in the decision-making process.

8.9.2007BritsocMedCon

Rika S. Sato & Takeo Nakayama

13

## Are guidelines helpful in decision-making?

- Guidelines tend to address evidence dealing with short-term risks and benefits, but women want to know about long-term prognosis.
- Recommendations such as "Treatment X is the best option for women who wish to achieve Y as an outcome" are more useful than recommendations such as "Treatment X is effective for Y."
- When there is only insufficient evidence, women appreciate "Good Practice Points," i.e. the "recommended best practice based on the clinical experience of the guideline development group."

8.9.2007BritsocMedCon

Rika S. Sato & Takeo Nakayama

14

## Ideal format for patients' use

- The format of Canadian guideline, which has a list of recommendations at the beginning, is preferred.
- Some say that they do not need the lengthy references on evidences, while others want to know how the recommendations were derived.
- Flowcharts and glossaries may be useful.
- Internet access to guidelines is a must. A free brochure with a list of recommendation is desirable.

8.9.2007BritsocMedCon

Rika S. Sato & Takeo Nakayama

15

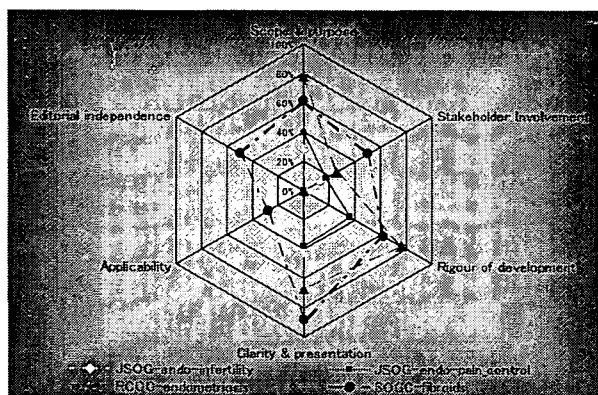
## Professional vs. lay perspectives

- Nakayama & I utilized the AGREE Appraisal Instrument ([www.agreecollaboration.org/](http://www.agreecollaboration.org/)) to evaluate the guidelines from a professional point of view.
- Canadian guideline is the most well-balanced of all, receiving highest scores on 4 out of 6 domains, while the UK guideline score high in "scope & purpose" and "rigour of development". Japanese guidelines fair poorly in all aspects.
- Although patients did not use analytical tools to assess these guidelines, the overall assessment was similar to the outcome of AGREE evaluation.

8.9.2007BritsocMedCon

Rika S. Sato & Takeo Nakayama

16



8.9.2007BritsocMedCon

Rika S. Sato & Takeo Nakayama

17

## Patients speaking out

- "When I read the SOGC guideline on uterine fibroids, I was impressed by the way in which the doctor-patient relationship was defined in a non-hierarchical manner. I began to wish for something like this in the clinical setting for the treatment of fibroids and endometriosis in Japan. I now think that patients should voice their opinions more explicitly in order to realize such guidelines in Japan."  
--fibroid patient in her late forties

8.9.2007BritsocMedCon

Rika S. Sato & Takeo Nakayama

18

## Conclusions

- Patients are capable of understanding clinical guidelines, if the guidelines are written in plain language and glossaries are provided. Their assessment of guidelines do not differ so much from that of professionals.
- In order to utilize clinical guidelines for more patient-centered healthcare, guideline developers should pay more attention to what patients consider relevant to their clinical decision-making.
- Internet access to clinical guidelines is likely to change patient-doctor relationship and promote “shared decision making” between them.