

- Robertson GM, Lynch AC, Frizelle FA. The utility of laparoscopy in the diagnosis of acute appendicitis in women of reproductive age. *Surg Endosc*. 2003 Aug; 17(8):1311-3.
- 61) Ignacio RC, Burke R, Spencer D, et al. Laparoscopic versus open appendectomy: what is the real difference? results of a prospective randomized double-blinded trial. *Surg Endosc*. 2004; 18: 334-337.
- 62) Lintula H, Kokki H, Vanamo K, Valtonen H, Mattila M, Eskelinen M. The costs and effects of laparoscopic appendectomy in children. *Arch Pediatr Adolesc Med* 2004; 158: 34-37.
- 63) Katkhouda N, Mason RJ, Towfigh S, Gevorgyan A, Essani R. Laparoscopic versus open appendectomy: a prospective randomized double-blind study. *Annals of Surgery* 2005; 242: 439-450.
- 64) Moberg AC, Berndsen F, Palmquist I, Petersson U, Resch T, Montgomery A. Randomized clinical trial of laparoscopic versus open appendectomy for confirmed appendicitis. *British Journal of Surgery* 2005; 92: 298-304.
- 65) Olmi S, Magnone S, Bertolini A, Croce E. Laparoscopic vs open appendectomy in acute appendicitis: a randomized prospective study. *Surg Endosc* 2005; 19: 1993-1995.
- 66) Satoru Shikata, Takeo Nakayama, et al. Comparison of effects in randomized controlled trials with observational studies in digestive surgery. *Annals of Surgery*. 244(5): 668-676, 2006.

Laparoscopic vs. open appendectomy for appendicitis, odds ratio of wound infection risk. Cumulative M-A

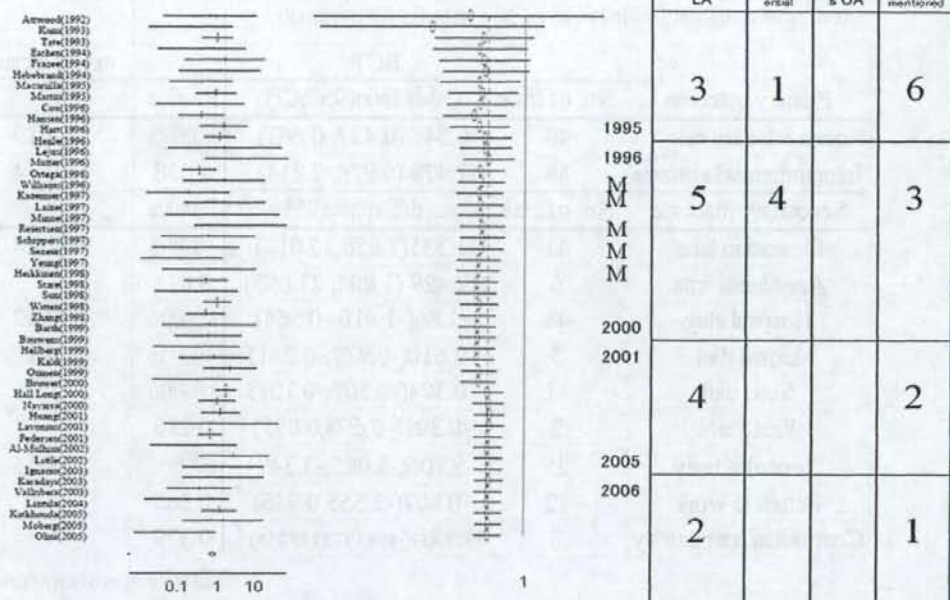


Figure 1.

Laparoscopic vs. open appendectomy for appendicitis, odds ratio of intraabdominal abscess. Cumulative M-A

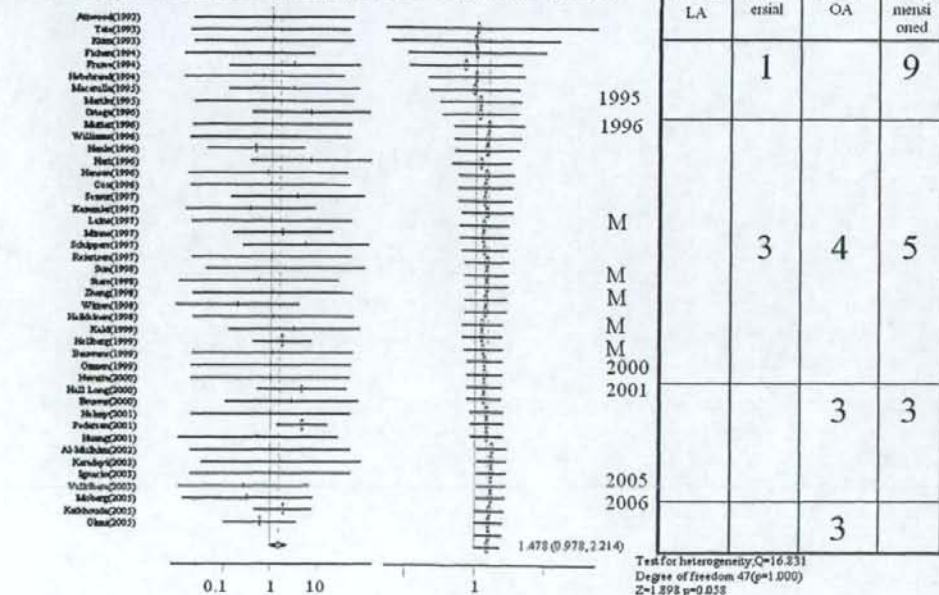


Figure 2.

Table 1.

Results of weighted pooled analysis and tests for homogeneity

Primary outcome	RCT			expert opinion
	No. of trials	Odds ratio(95%CI)	P value	
wound infection risk	48	0.546 (0.437, 0.682)	<0.005	14/19
Intraabdominal abscess	48	1.478 (0.978, 2.214)	0.058	10/14
Secondary outcome	No. of trials	Mean difference(95%CI)	P value	
Operation time	41	11.335(7.658,15.012)	<0.005	11/14
Anesthesia time	6	12.429 (1.804, 23.053)	0.022	
Hospital stay	44	-1.86(-1.410,-0.564)	<0.005	10/12
Liquid diet	5	-0.610(-0.979,-0.241)	<0.005	0/1
Solid diet	11	-0.304(-0.507,-0.101)	<0.005	
First stool	2	-0.301(-0.698,0.095)	0.136	
Normal activity	25	-5.706(-8.065,-3.347)	<0.005	14/16
Return to work	12	-0.807(-2.555,0.946)	0.366	
Cost within hospitalstay	8	371.439(-456.137,1189.015)	0.379	

■ Heterogeneity was noted

研究協力者 四方 哲 蘇生会総合病院外科、京都大学大学院医学研究科健康情報学  
研究代表者 中山 健夫 京都大学大学院医学研究科健康情報学

はじめに

外科医の立場からすると、診療ガイドラインの記述内容と医師の裁量との葛藤や訴訟との関連が気になる場所である。自分の方針に自信を持つ外科医も、診療ガイドラインに沿うにしろ沿わないにしろ、なぜその治療を推奨するのかを患者や他の医師に説明できなければいけない。とはいえ、手術手技や術者の熟練度、人的・設備的・地理的な要素も大きく影響する外科治療すべてをエビデンスに基づき説明可能なわけではない。日々の外科診療の中では、「むしろ最適なエビデンスが存在しないことの方が多い」と思えるのではないだろうか。そのような中で、今後の外科診療を発展させる足掛かりとなる診療ガイドライン活用法を考えてみたい。

### I. EBM から診療ガイドラインへ

1991年に臨床疫学者 Guyatt によって提唱された EBM (Evidence-based Medicine) という言葉は急速に世界に広まった<sup>1)</sup>。EBM は臨床医の経験に加えて、患者集団を対象とした臨床研究を重視した個別医療の方法論である。EBM の実践は、データベース等から文献を探し、批判的吟味をしたうえで目の前の患者にこれを適用するかどうかを検討するプロセスである。この流れの延長で、多忙な臨床医に代わって批判的吟味をおこない高質なエビデンスをまとめた 2 次情報が利用可能になってきた。RCT やシステマティックレビューを中心に提供しているコクランライブラリ、ACP ジャーナルクラブ、クリニカルエビデンスなどがその代表であり、広範囲な領域を網羅しながら日々内容が改定されているオンライン教科書として UpToDate や DynaMed なども急速に普及している(表

1)。外科医といえどもこれらの 2 次情報を利用することは常識になりつつある。EBM の手法を用いた診療ガイドラインは EBM 実践を支援する 2 次情報の中核になる。

### II. 外科医と診療ガイドライン

診療ガイドラインとは、「特定の状況の下で臨床家と患者とが適切な判断や決断を下せるよう体系的に作成された文章」である<sup>2)</sup>。国内では 1999 年から厚労省が主要疾患の診療ガイドラインの作成を開始し、多くの学会や研究会がその作成・普及に取り組んできた。2007 年までに国内で書籍として刊行された診療ガイドラインは 188 件あるが、その約 1/3 (68 件) が EBM の手法に則って作成されたものである<sup>3)</sup>。海外のものまで考慮すれば、その種類と数は膨大で、平素から情報を整理しておく必要がある。表 2 に外科医が臨床をおこなううえで有用と思われる診療ガイドラインに関する情報の一部を示した。

一方で、「ガイドライン」というタイトルがついてさえいれば信用するという態度には注意が必要である。診療ガイドラインにおける推奨内容と作成者への企業献金との関連を指摘した報告もあり、世間の関心も高まりつつある<sup>4,5)</sup>。まず、ガイドライン作成母体や作成委員の公平性、資金源や利害衝突に関する記述の確認は必須である。その上で、記述内容は専門家の意見ではなく根拠(エビデンス)に基づいたものであること、作成プロセスが妥当であることを確認する。各外科医はガイドライン作成プロセスの原則を理解し、評価したうえで自分達が採用すべきガイドラインか否かを決めておく必要がある。

### III. 診療ガイドラインの拘束力

診療ガイドラインは、直接的に個々の臨床判断を拘束する力はない。そもそも「ガイドライン(または指針)」とされるものの意味が、国内ではほとんど議論されていないため、曖昧なままになっていて混乱がある<sup>6)</sup>。欧米では、拘束力が強い順に、Regulations(規制)、Directive(指令)、Recommendation(勧告・推奨)、Guideline(ガイドライン・指針)とされており、Recommendationでさえも直接的な拘束力はない。しかし、診療ガイドラインが存在することで、医師の判断の根拠を問われる場面がでてくる可能性があり、その結果、間接的に医師の臨床判断に「規制的な影響」を与える<sup>7)</sup>。近年、海外では“soft law”の考え方が提唱されている。これは明確な法・規則ではないが、基本的には遵守することが求められ、しない場合は、その理由の説明が必要とされるもので、“comply or explain” principle と呼ばれる<sup>8)</sup>。この流れは、社会の望む方向であることを認識する必要がある。しかし、当初懸念されていたような訴訟におけるガイドラインの誤った引用は、今のところ明らかではない。むしろ良識ある外科医の多くは、診療ガイドラインの登場により、大多数の医師が共有すべき診療ストラテジーがコンセンサスを得たことで、極端に独善的な診療が淘汰されるようになって良かった、と思っているのではないだろうか。

### IV. 活用①: 患者との情報共有ツール

外科医は、手術前に患者や家族に対して、手術の必要性や方法、合併症などの説明を行い、同意を得なければいけない。これは

省略することができないプロセスであり、かなり時間を要する日常業務である。筆者は、予定手術の説明を行う数日前に、診療ガイドラインやそのコピー、自作の資料などを患者に貸し出し、事前に予習していただくことにしている。これにより、いきなり何も無いところから説明を開始するよりもスムーズに本質的な説明に入ることができるようになった。現在、臨床医向けに作成した診療ガイドラインの一部を患者が理解できる形で作成し直した、いわゆる「患者向けガイドライン」も数多く公開されているのでこれらを積極的に活用しない手はない。財団法人日本医療機能評価機構 Minds では、医療者向けの診療ガイドラインを平易に書きなおし、用語解説(医学用語ヘルプ)を付した「一般向け MindsPLUS」をオリジナルコンテンツとして提供している。

表2にあげたガイドライン質評価ツールの項目にも、ガイドライン作成の手引書にも「患者の価値観や好みを考慮すること」の重要性が強調されている。近年では国内においても患者(団体)が何らかの形で参画して作成された診療ガイドラインが増えつつある。これらをコミュニケーションの基点として利用し、協働的に意思決定をおこなおうとする外科医の姿勢や態度は必ずや患者に通じ、信頼と安心感を与えるであろう。

### V. 活用②: 外科医の学習ツール

内科ならハリソン、外科ならサビストンが一流の教科書であることに変わりはない。それらを輪読したり、オリジナル文献を吟味したりする従来からの勉強会は大切である。「外科医は多忙なのだから、それらに加えていちいち各種ガイドラインなんか読んでいら

れないよ」という声が聞こえてきそうである。確かにその通りである。しかし一度、自分が担当する番に、ガイドラインの記述を徹底検証するスタイルのプレゼンを試してみたらどうであろうか。あるテーマについて、ガイドラインの引用文献を批判的吟味し直したり、海外のガイドライン内容と比較したりして推奨文の検証を試みるとガイドラインの記述を超えた背景知識が広がるはずである。そもそも普段使用するガイドラインに関するものであるから同僚の医師に興味をもって聞いてもらえるであろうし明日からの臨床に役立つはずである。

現在、専門医制度の在り方が議論されているが、各学会のポリシーが反映された診療ガイドラインから専門医試験問題が出題されるようにするなど、生涯学習ツールとしての有効利用も検討されている<sup>9)</sup>。

#### VI. 活用③:臨床研究テーマの宝庫

先に挙げたようなガイドラインを足掛かりとした勉強をすると、臨床的に重要であるがエビデンスに乏しく未解決な領域が何であるのが明確になる。「次の改訂でこの部分の記述を変えてもらえるような研究をしよう」と思い立つチャンスに繋がる。外科医でなくとも、自分の研究が教科書やガイドラインの記述を変えたり引用されたりすれば光栄なことに違いない。

診療ガイドライン作成者からは、「外科には RCT という高レベルのエビデンスが乏しい」という意見がくりかえし述べられ、第一線の外科医からは「自分は RCT をできる環境にはない」とか「同一症例が少なく n(サンプルサイズ)が小さいので研究になりにくい」といった元気がない声がよく聞かれる。これら

は、外科でも RCT が最高のエビデンスである、大規模ほどよい試験だ、という前提に基づく誤解かもしれない。筆者らの研究によれば、外科領域では、RCT の結果と観察研究(コホート研究)の結果に大差はないので、よくデザインされた観察研究であれば最高のエビデンスになりうる<sup>10)</sup>。また、大規模にしなければパワー不足になり有意差がでないような治療は、そもそも効果や患者にとってのメリットは少ないのかも知れない。ならば、数値化できる客観的な指標のみにとらわれずに患者の視点に立った主観的な指標を調べる研究(患者立脚型アウトカム研究)なども外科領域では重要と考えられる。また、多様化する価値観や術前・術後患者の特性、思考・行動様式を質的研究で明らかにすることも重要な研究として注目されており、今後は評価されるようになるであろう<sup>11,12)</sup>。

各学会はガイドライン公開前の公聴会だけではなく、改訂に向けたセッション等を継続して開催し、一人ひとりの外科医がガイドラインを起点とした研究成果を存分に発表できる場を与え、それらを反映した改訂を重ねてゆく必要がある。

#### おわりに

広い意味での EBM には「エビデンスを作る」、「伝える」、そして「使う」の三つの局面がある<sup>13)</sup>。本稿ではガイドラインを利用した各局面の一例を解説した。新たなエビデンスを取り入れて改訂される診療ガイドラインは、個々の臨床場面で役立つだけでなく医師の生涯学習、臨床研究の方向性を検討することにも連繋する。また、ガイドラインをコミュニケーションの基点として患者らと情

報を共有することで協働的な意思決定を行うスタイルは医療における信頼の基盤となる。

#### 引用文献

- 1) Guyatt G: Evidence-based Medicine. ACP Journal Club 114:16, 1991
- 2) Field MJ, Lohr MJ: Clinical Practice Guidelines: Directions for a new program, ed by Field MJ, Lohr MJ, National Academy Press, Washington DC, p1-38, 1990
- 3) 奥村晃子, 中山健夫: 日本の診療ガイドラインの評価—利害衝突の開示状況を中心に—平成19年度厚生労働科学研究費補助金 医療安全・医療技術評価総合研究事業報告書「診療ガイドラインの新たな可能性と課題:患者・一般国民との情報共有と医療者の生涯学習」p211-221, 2008
- 4) Choudhry NK, Sreelfox HT, Detsky AS: Relationships between authors of clinical practice guidelines and the pharmaceutical industry. JAMA 287: 612-617, 2002
- 5) Papanikolaou GN, Baltogianni MS, Contopoulos-Ioannidis DG et al: Reporting of conflicts of interest in guidelines of in guidelines of preventive and therapeutic interventions. BMC Med Res Methodol 1(3), 2001
- 6) 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 15: 359-60, 2003
- 7) 中山健夫:4)拘束力について、EBM を用いた診療ガイドライン作成・活用ガイド、第1版、中山健夫編集、金原出版、東京、15-17、2004
- 8) Campbell A, Glass KC. The legal status of clinical and ethics policies, codes, and guidelines in medical practice and research. McGill Law J 46:473-89, 2001
- 9) 中山健夫. 平成19年度厚生労働科学研究費補助金 医療安全・医療技術評価総合研究事業報告書「診療ガイドラインの新たな可能性と課題:患者・一般国民との情報共有と医療者の生涯学習」1-6, 2008
- 10) Shikata S, Nakayama T, Noguchi Y et al: Comparison of effects in randomized controlled trials with observational studies in digestive surgery. Annals of Surgery 244:668-676, 2006
- 11) Qualitative Research in Health Care Third edition Ed by Catherine Pope and Nicholas Mays, Blackwell Publishing Ltd, Oxford, 2006
- 12) 質的研究実践ガイド 保健医療サービス向上のために 第2版 大滝純司(監訳) 医学書院、東京、2008
- 13) EBM のための情報戦略:エビデンスをつくる、つたえる、つかう.中嶋宏監修、津谷喜一郎、山崎茂明、坂巻弘之編集.中外医学社、東京、2000.
- 14) Minds 版診療ガイドラインの手引き 福井次矢、吉田雅博、山口直人 著. 医学書院、東京、2007.
- 15) Guirguis-Blake J, Calonge N, Miller T et al: Current processes of the U.S. Preventive Services Task Force: refining

evidence-based recommendation development. *Ann Intern Med* 147:117-122, 2007

- 16) Shaneyfelt TM, Mayo-Smith MF, Rothwan J: Are guideline following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 281: 1900-1905, 1999
- 17) Cluzeau FA, Littlejohns P, Grimshaw JM et al : Development and application of a generic methodology to assess the quality of clinical guidelines. *Int. J. Qual. Health Care* 11: 21-28, 1999

#### 【論文発表】

- 1) 四方 哲、中山健夫： Multiple problems:エビデンスが当てはまらないケースはどうするか？ *Medicina* 45(9): 1585-1588, 2008.
- 2) 四方 哲、中山健夫： 外科医のための臨床研究講座 第9回 なぜ診療ガイドラインが必要なのか？—外科医による診療ガイドラインの活用 *外科* 71(1):71-75, 2009.
- 3) Shikata S, Nakayama T, Yamagishi H. Quality of surgical randomized controlled trials for acute cholecystitis: Assessment based on CONSORT and additional check items. *J Hepatobiliary Pancreatic Surg.* 2008;15(3):297-303.
- 4) Yoshitomi H, Shikata S, Ito H, Nakayama T, Nakamura T. Manufacturers Affect Clinical Results of THA with Zirconia Heads: A Systematic Review. *Clin Orthop Relat Res.* 2009 Jan 31.

#### 【学会発表】

- 1) 2008(平成 20)年 7 月 16 日 里輝幸、中山健夫、四方哲 累積メタアナリシスを用いた臨床研究の動向評価:腹腔鏡下虫垂切除術のRCTを例に. 第63回日本消化器外科学会定期学術総会(札幌)
- 2) 2007(平成 19)年 4 月 11 日 四方哲、中山健夫、野口善令 —外科領域における臨床試験の特異性と将来—「消化器外科では観察研究がRCTの代わりになる」第 107 回 日本外科学会定期学術集会(大阪、シンポジウム)
- 3) 2008(平成 20)年 7 月 18 日 四方哲、中山健夫、里輝幸 —消化器外科領域の動向と日本の国際的役割: 高レベルのエビデンスを提案できる国を目指して—「消化器外科領域における新エビデンスレベルの提案に向けて」第 63 回日本消化器外科学会定期学術総会(札幌、特別企画)



表 1. EBMに必要な2次情報

Cochrane Library(コクランライブラリ)	<a href="http://www.cochrane.org/">http://www.cochrane.org/</a>
ACP Journal Club	<a href="http://www.acpjg.org/">http://www.acpjg.org/</a>
UpToDate	<a href="http://www.utdol.com/home/index.html">http://www.utdol.com/home/index.html</a>
DynaMed	<a href="http://www.ebsco.co.jp/medical/dynamed/">http://www.ebsco.co.jp/medical/dynamed/</a>
Clinical Evidence(クリニカルエビデンス)	<a href="http://clinicalevidence.bmj.com/ceweb/index.jsp">http://clinicalevidence.bmj.com/ceweb/index.jsp</a>

表 2. 外科医に必要なガイドライン情報

<b>国内のガイドライン利用に有用な情報</b>	
Minds 医療情報サービス	日本医療機能評価機構による医療情報サービス。 国内の診療ガイドライン約 70 種類の全文が公開されている。 「医中誌 Web」に 1999 年以降に収録された国内の診療ガイドラインを疾患別に分類し、書誌事項を掲載している。
http://minds.jcqhcc.or.jp/index.aspx	
医学中央雑誌「診療ガイドライン」	日本癌治療学会の後援で作成した 10 種類のがん診療ガイドラインが公開されている。今後 23 種類のがん診療ガイドラインが作成、公開される予定である。
http://www.jamas.or.jp/	
がん診療ガイドライン	
http://jsco-cpg.jp/	
<b>海外のガイドラインデータベース</b>	
NGC (National Guideline Clearinghouse)	米国 AHRQ によるガイドラインのデータベース。系統的な方法で作成された英語記述のガイドライン 5 年分を検索できる。 ガイドライン間の比較機能などもある。
http://www.guideline.gov	
NICE (National Institute for Health and Clinical Excellence)	英国立審査機関が実施する情報サービス。公開されたガイドライン
http://www.nice.org.uk/Guidance/Type	約 80 種類と現在作成中である 50 種類が公開されている。
CMA Infobase (Canadian Medical Association)	カナダ医師会によるガイドラインのデータベース。カナダの保健機関の後援で作成された 1200 種類のガイドラインすべてが公開されている。
http://www.cma.ca/index.cfm/ci_id/54316/la_id/1.htm	
<b>ガイドラインの理解・作成に有用な情報</b>	
EBM を用いた診療ガイドライン作成・活用ガイド	EBM を背景とした診療ガイドライン作成について解説している。
文献 7)	
Minds 版診療ガイドラインの手引き	Minds におけるガイドライン作成のためのガイドライン (GLGI) を解説している。
文献 14)	
ACCP ガイドライン作成マニュアル	米国胸部専門医学会 (ACCP) がエビデンスに基づく診療ガイドラインの作成プロセスの全容を公開している。
http://www.chestnet.org/education/guidelines/development/	
USPSTF エビデンスに基づくガイドラインの作成	U.S Preventive Services Task Force によるガイドライン作成の手順。
文献 15)	
GRADE エビデンスの質とお勧め度	国際的任意団体による診療ガイドライン作成システム。多くの団体がこのシステムを採用しつつある。
http://www.gradeworkinggroup.org/index.htm	

ガイドラインの批判的吟味に有用な情報  
AGREE(Appraisal of Guidelines Research and  
Evaluation)

<http://www.agreetrust.org/>

国際プロジェクトが開発した診療ガイドラインの質を評価するツール。

6つの領域における23項目について4段階の評点で評価する。

3つのカテゴリーにおける25項目についてYes、Noで評価する。

3つのカテゴリーにおける37項目についてYes、Noで評価する。

文献 16)

文献 17)

Shaneyfeltらのガイドライン評価

Cluzeauらのガイドライン評価

## 診療の質指標(Quality Indicator)とガイドライン

研究協力者 東 尚弘 国立がんセンターがん予防・検診研究センター検診研究部

### 1. 医療の質とは何か

医療の質という言葉は氾濫している。医療機関のホームページなどでは、院長の挨拶や病院の理念といった場に必ずといってよい程、質の高い医療が触れられているし、医療政策に関する新聞記事にも医療の質を確保するためには何かかくかくしかじか、といったような言葉が頻繁に出てくる。もはや、医療の話題で触れるべき言葉のお約束でさえあるかもしれない。

しかし、いったいこの医療の質とは何だろうか？そして、どうやって評価するのか、ということは、これらの言葉を使っている人々には何か見通しがあるのだろうか？ともすると「愛」とか「友情」などと同じように、美しい言葉でありながらよくわからないものではないかという気さえする。またそれが故に逆に便利によく使われる、それが医療の質という言葉なのではないだろうか。

少し掘り下げるとどういうことになるのか。医療の質研究の基本的な枠組みを広めた Donabedian は対人関係的な質と医療技術的な質に大きく分けられると述べた<sup>1</sup>。前者は患者に対して人間として親身になり、人として尊重し、きちんとしたコミュニケーションをとるということである。それが、疾患を治癒する、生命を延長する、生活の質を向上させるという医療本来の目的にどれほど貢献するのかは定かではない面があるが、患者の満足度などで評価測定が可能である。

一方、医療技術的な質は、その時々の医療水準で患者がよくなるために方策を提供している度合いであるといえる。これは医療本来の目的、つまり患者の健康状態の改善を、ど

れだけ達成するか、また達成するような方策をとっているか、とも言い換えることができる。しかし、これは評価が非常に難しい。個別の例に限って考えれば、例えば最善のことをやっても患者がなくなることもあり、また放っておいても患者はよくなる、ということも医療者ならば往々にして経験することである。そのため集団として初めて可視化される確率論でしか評価のできない面がある。また、個人々人にとってはよい医療と考えられても集団にとっては害をなすこともある。抗生物質の処方などはその一例であり、個人々人への抗生物質の使用は耐性菌の出現をもたらすため、最小限度としなければならないとされている。これら諸点をふまえて、米国の医学研究所 (Institute of Medicine) の、「診療の質とは、個人および集団に対する診療行為が望まれた健康状態をもたらす確率を上げ、かつ、最新の専門知識と合致する度合いをいう」と定義した<sup>2</sup>。また、付記するならば、この定義単純に、健康状態の改善とせずに、「望まれた健康状態をもたらす」として、患者中心の目標設定を盛り込んでいるところにも深い思考過程が伺える。

### 2. 医療の質の評価・測定

#### a. 過程・構造・結果の視点

先述の Donabedian が提唱したモデルにおいて医療評価には構造・過程・結果の3つの視点があるといわれている<sup>1</sup> (図1)。構造とは、診療を取り巻く環境・資源要因、例えばスタッフの数、診療機器の有無などを評価対象とするものである。また経営状況は健全かといったことも構造評価の一部であり、分類

によっては特定の手技・手術・治療の数などいわゆる Volume を構造と見なすこともある。過程は、診療内容、つまり医療者の提供する医療行為を直接評価対象とする。スタンダードに沿った治療が行われているか、また、沿わない場合は理由（患者の拒否）が明記されているか、さらには薬物療法のタイミングなども評価対象となる。結果の視点からは診療の結果として患者に起こる事柄を評価する。つまり、望ましい事柄（治癒、健康状態の改善）がどれだけ起きたのか、望ましくない事柄（死亡、再入院、再発など）がどれだけ少なく抑えられたのかを評価する。もちろん過程とは医療そのものであるためそれが直接評価できるに越したことはないわけであるが、それが困難を伴うからこそ、間接的な指標である構造、結果という視点が存在することになる。以下にその長所、短所を検討する。

#### b. 各視点の長所、短所

まず、構造であるが、これは医療が起こる前の状況を評価するという意味では未来志向であり、今後受診する患者の視点からは最適の視点ともいえるかもしれない。また、評価そのものの信頼性、という点からは、構造は評価しやすく、ある意味「数えることができる」といったやすさがあり、専門知識もさほど必要ない。しかし、構造はあくまで間接的な視点であり、適切な医療がそこから行われてよい結果をもたらすのかは検証の余地がある。例えば、最新の検査機器があったからといってきちんと使いこなせるかはスタッフの力量によるだろうし、正しい診断の後に適切な治療が行われなければならない。また、よさそうに見える治療機器も過剰な診療を生むだけで、実はアウトカムを改善しないなどの問題点が指摘されている（CTで十分な疾患に対して、MRIを撮影するなど）。

次に順番を入れ替えて結果評価について検討する。結果の改善が医療の目的であるため、目的の達成により評価するのは非常に自然である。生存率、再入院率、合併症発生率などはしかも、患者にもわかりやすい。歴史的にも、欧米では、ナイチンゲールが各施設の死亡率のデータをとり比較した<sup>3</sup>、とか、米国ボストンの外科医 Codman が、外科医の手術成績を追跡した<sup>4</sup>、といった事例が頻繁に紹介される。わが国でも、江戸幕府の設立した小石川養生所において、医師ごとの成績が評価されたという記録も残っている<sup>5</sup>。

しかし、これらの結果評価の数字が他の施設や、他の時期などと比較ができるのか、ということに大きな問題がある。というのは、評価として出てきた「結果」は、ただ、「医療を受けた後の患者の状態」というだけで、必ずしも医療の結果とは限らないということである。何よりも、結果は患者の医療前の基礎状態に大きく左右される。もともと重症な患者は、どれだけ医療が適切であっても悪い結果となりがちであるし、軽症な患者は最悪の医療を受けていても改善して健康になることがある。そのため、結果を比較するためには基礎状態が同じ患者を比較することが必要となるわけであるが、各施設や評価対象となる患者集団が均一であることは珍しく、そのため、実際には多様な患者を統計的な調節処理をして均一な「仮想患者」について比較するということが行われる。そこで均一な仮想患者がどれだけうまく作れるかということが非常に大きな問題であり、順位などを付けようとする、統計処理方法によってそれが大きく異なるということが明らかになっており、発展途上であるという。1980年代末に米国で公的保険メディケアの契約病院の死亡率を公表したことがあったが、実際の死亡率が、患者の重症度を「統計的に調

節」した後の「あるべき死亡率」の4倍である施設（つまり、患者の重症度から算定される死亡率よりも、実際死亡している割合が非常に高い劣悪な施設と考えられる）が、実はホスピスであり、ホスピスに入所する患者の重症度をうまく統計的調節法が捉え切れていないという批判を浴びたこともあり<sup>6,7</sup>、統計的調節は「行う」「行わない」の問題ではなく「どのように行うのか」というのが問題であることを常に意識する必要がある。もちろん、偶然や運、医師の指示が正しくても患者のコンプライアンスなど、実に様々な因子が結果に影響してしまい、どこまでが診療の純粋な結果なのかは難しい問題である。（図2）

では、過程評価はどうか。過程評価とは、行われた医療の適切性を、医学知識、標準に照らして評価するものである。主に大きく2通りの方法がある。最初に基準を明示的に決めて、その基準をどれだけ満たすかによって実際の診療内容を評価するという、明示的（Explicit Review）、逆に基準は明確に決めずに実際の症例の診療記録を見ながら専門家がその診療内容がどの程度の質であるかを評価する内示的レビュー（Implicit Review）というものである<sup>8</sup>。Explicit reviewは明示的な基準を作成するのが困難である反面、いったん基準が決まってしまうえばその適応には高度な専門家を要することなく、データを収集する人員のみで評価でき、また基準を統一することで、広く多施設で比較が可能である。しかし逆に基準が作られている範囲しか評価ができないことから、「医療の質」といえる程に広い範囲の事柄について医療全体を評価しようとする多くの基準が必要で、かつ、それらを適用するために集めなければならないデータの量が非常に多くなるという欠点がある。一方 Implicit reviewの方は吟味

を担当する専門家の知識が及ぶ範囲で評価が可能であり網羅的であるという利点があるが、一般化が難しいという問題もある。また、比較可能にするためには、専門家が同じ視点で同じ程度の厳しさを評価を行うことが必須であるが、それを確保するのが難しいという面がある。そこで、最低限視点のみをそろえて評価を行う、「構造化内示評価」（Structured Implicit Review）というものもある<sup>9</sup>。また、Implicit Reviewは診療記録をレビューすることで逆に結果悪い転帰をたどった症例に関しては、診療への評価も厳しくなるという研究結果もある<sup>10</sup>。これらの点から、広く医療の質評価を系統的に行うという場合には、明示的評価の手法がとられることが主流である<sup>11,12</sup>。過程評価全体の特徴としては、結果評価で必要だった、患者の基礎状態に対する調節が不要である点が大きな利点であるという。これは過程評価の基準が「このような患者にはこのようなことをしているか」というものであることから、患者の基礎状態が「このような患者」という部分に暗黙のうちに考慮されているからである。逆に、診療の質を総合的に評価するのが過程評価においては難しいという欠点がある。まず、指標の設定されている範囲や、判断者の知識や観察が及ぶ範囲でしか評価ができない。これは必然的に一部であり、氷山の一角であるともいえる（図3）。特に Explicit Reviewにおいては、あらかじめ質の基準を設けておくため、総合的に評価しようとする、非常に数多くの指標を用意しなければならず、それを作成、使用する、両方の局面で非常に手間がかかってしまい、資源の面からの必然的な限度が決まってしまう。

さらに、過程評価で評価できるのは基準が必要であることから、基礎的な標準診療が中心になる。画像で他人には見えないような腫

瘍が見つけれられる、とか、人にはできない手術ができる、というのは過程の基準を作るのが難しく、対象とはできない。このような事柄は、基本的に結果評価で評価するしか方法がないのが実情である(図4)。

### c. 評価視点の相補関係

明示的評価の欠点である網羅性については、結果評価を組み合わせることで相補うことが考えられる。結果評価は網羅的である反面医療以外の因子、いわば雑音が混入する問題があるが、結果評価と過程評価は本来雑音がない場では相関するはずなので、それぞれの関連を証明することが評価の妥当性を検証することになる。これらを Donabedian は原因妥当性 (Attributional Validity)、寄与妥当性 (Contributional Validity) と呼んだ<sup>1</sup>。つまり結果基準の妥当性(診療の質を捉えているかどうか)を「医療の質が原因といえるか (Attributable)」で検討する原因妥当性、過程基準に示された診療過程が、アウトカムの改善へ寄与しているかどうかを検証する寄与妥当性、という意味である。

ここで、関連が証明できなかつたときには、どちらが質の基準として強いかという過程がどれほど医師の臨床知識水準に関して規範となっているか、による。つまり、規範であることについては結果との関連がない場合には結果の測定の不備が推定されることになる、規範として少し怪しい(エビデンスが弱い、など)場合には、結果評価の方が重視されることになる。これらは、どちらかが絶対という正解はなく、そのため個別に基準を評価することが必要である。

極論すれば、過程評価は規範性が強いものについては絶対評価が可能である。つまり、こういった患者についてはこうすべき、ということがいえるならば 100%をめざしてそれ

に近い方が質が高いということが絶対的にいえ、また 100%となった時点でその点に関してはそれ以上改善の余地はない、ということになる。(もちろん、おおざっぱな評価のためどの程度がよいのかわからないという過程評価もありえる。これはあくまで基準を精緻に作成した場合ということである。)一方で結果評価については最善を尽くしていても自然のばらつきから結局のところ死亡する症例がある、といった理由から死亡がありえる疾患については、どこまで達成するのが最善ということがどこまでも不明であり、結局平均などの外的な参照値がないと、達成度が不明ということになる。

医療の質として昨今世界的な問題となっているのは、最新の臨床知識が現場で生かされていない、現場で普及していない (Evidence-practice gap) ということである<sup>13</sup>。つまり、知識的には標準と考えられるものが、現状の標準となっていない、ということである。この問題意識を基本として医療の質評価を考えたときには、過程評価で絶対基準を作るところが必要である。

しかし、医療に絶対があるのか、という問題がある。使いたい薬が基礎疾患(肝障害、腎障害)などのために使えないといったことは往々にしてあるので、100%満たすべき基準なんて存在しないという意見もある。しかし、例外については例外とした記載を加味することで 100%をめざした、ということもいえるのではないか。この際の記事はもちろん評価をパスするための記載ではなく、患者の病状の記録を医療チームで共有するという意味で重要なはずであり、主治医のみが例外であることを知っていればよいという性質ではないと考えられる。例外の記事を強調すると、評価しているものが医療の質なのか、記録の質なのかという綱引き議論となって本

質からずれてしまうことがあるが、それは、おそらく結果評価を同時に組み合わせて寄与妥当性を検証することで避けるべき性質のものであろう。

#### d. 診療内容のばらつきは質の問題か

診療の質評価に関する初期の研究では主に「ばらつき」に焦点が当てられて評価されてきた経緯がある。Wennberg<sup>14</sup>は1973年に米ニューハンプシャー州バーモント市における受療・医療費・診療内容のばらつきを記述して科学誌 Science に報告した。ここでは、狭い地域の中で非常に大きなばらつきが見られており、このばらつきは、患者特性のばらつきでは説明できる範囲よりも大きいのではないかと述べている。ここで重要なのは「患者特性で説明できるか」ということなのだが、昨今のばらつき論はそれが置き去りにされているリスクが懸念される。各種ガイドラインでも主張されている通り、診療とは個々の患者に標準を応用して当てはめるものである。個々の患者はそれぞれの併存症、重症度、臨床状況があり、標準をその状況に当てはめることが本来医師・医療者の責務である。そのため、ばらつきが生じるのは当然である。一言でいうならば「よいばらつき」「悪いばらつき」が存在するということで、「よいばらつき」とは患者の状況にあわせて診療内容が適合されているというものであり、「悪いばらつき」とは患者の状況以外の要因、例えば医療者の都合や技術、保険制度の違いなどのために起きるものである。ばらつきを即医療の質の問題として見なされることにはここであらためて警鐘を鳴らしたい。

また、この観点からもばらつきの指摘は問題の出発点でしかなく、問題の存在の結論とはなりえないことも付記しておく。

### 3. 明示的基準としての評価の指標 (Quality Indicator)

先述の明示的基準を「過程を対象とした Quality Indicator (QI)」と呼ぶことがある。Quality Indicator は質の指標ということなので、一般に Quality Indicator 自体は何も過程評価に限ったことではなく死亡率を Quality Indicator としてもよいわけだが、過程評価の指標を表すよい言葉がないので、本稿においては特に指定がない限り過程を対象とする Quality Indicator を指すことにする。

#### a. Quality Indicator, (QI) の構造

先述の通り、基本的に QI は標準医療を行っている割合を評価するものである。そのため、分母に対象とする患者や臨床状況を記述し、分子に、そのような患者に行われるべき診療内容を示している。これら分母の患者数のうち、分子を満たす患者数でその QI の準拠率が計算される。例えば、大腸癌の標準診療として、治療前に血清癌胎児性抗原 (CarcinoEmbryogenic Antigen, CEA) を測定するのが標準であれば、それを QI としてスコアは、治療を受けた大腸癌患者が分母であり、治療前に血清 CEA 値が測定され結果の診療録記載がある患者が分子になる。こうしてこの QI の準拠率が計算され、QI それぞれについて準拠率が計算され、それを QI ごとに他施設と比較したり、同一施設の中で他の項目を規定する QI 間で比較したりすることで、自施設における改善点が明らかになる。

#### b. 用語について

本稿では一貫して Quality indicator、質指標という言葉を使用するが、類義語として、臨床指標 (クリニカルインディケーター、Clinical Indicator) やパフォーマンス指標 (Performance indicator) といったものがある。



あまりそこに本質的な違いはなく方言のようなものである。ただ、語感には若干の違いがあり、臨床指標といった場合には、医療費や在院日数などのような経営指標と対比して、指標が「臨床」的事項であるということが意識されており、パフォーマンスと呼んだ場合には、「質」という言葉を避ける意識が働いている印象がある。さらに、測定を強調して質やパフォーマンスの尺度 (Quality measure, performance measure) という場合もある。(図5)

#### 4. 過程を対象とした医療の質指標とガイドライン

##### a. 類似点

標準医療がどれだけ行われているか、ということを考えるにあたって標準が何かということ定義する必要がある。そのために手取り早い情報源が診療ガイドラインといえる。診療ガイドラインは政府の推進もあってかなりの分野で整備されてきている。そこで次の段階としては、本当にガイドラインで示されたような診療が行われているのかを検証することにある。診療ガイドラインの評価ツールとして有名な AGREE 評価<sup>15</sup>の中に「診療の監査やモニターのためのツールが用意されているか」という項目があり、この「ツール」が QI に相当すると考えられるが、わが国の診療ガイドラインではそのようなツールが用意されていることは筆者の知る限り、まだこれから、という状況である。

##### b. 相違点

ガイドラインと QI は同じ標準的診療を使っているという点で類似している。しかし、ガイドラインの目的は「最新の知見に関する情報の普及による診療の支援」であり、QI の目的は「評価」であることから、いくつか

意識すべき相違点がある<sup>16</sup> (図6)。まず第1に対象とする分野である。ガイドラインは情報の普及であることからコンセンサスがない分野についても「コンセンサスがない」ということや、「こうすることが有効な可能性がある」といった情報でも提供する意味がある。一方で QI では、コンセンサスがない事柄は QI にならないのでそれは QI ではカバーしない、または、そのような診療を行う場合には「標準でない旨を患者に説明する」ことが QI となる。このような QI が、ある治療法の適応拡大などが積極的に行われている分野においてはそのような先進的試みを妨げないように考えて工夫しなければならない。

次にガイドライン推奨は情報の普及が主目的のため、推奨自体は簡潔に書かれて細部は使用する医師の判断にゆだねていることも多い。QI では評価が目的のため、あまり簡潔であることは意識されずに解釈に幅がないことを旨としている。例えば、手術不能の場合には〇〇を行うのがよいといった推奨は、「手術不能」の判断は各医師が患者ごとに判断することになるが、QI ではできるだけ判断が入らずに評価ができるように、手術が診断から〇〇の期間行われない場合を対象とする、といったように、明確な事実を元に対象を定義する構造としている。また「〇〇のような患者に対しては□□が有用である」、といった推奨に関しては、QI に変換する際には、それが果たして標準であり、原則として□□をやるのが求められるのかといったことを検討して、そうでない場合には QI として当該推奨は使えない、といった判断をすることになる。

また、ガイドラインで情報を普及する必要もない事柄であり、行った方がよいが、実際には行われていないようなこと、特に診療を

円滑にし、それを通じ患者アウトカムを改善するような事柄については、QI となる。例えば手術所見・病理所見の適切な記載などはガイドラインには含まれることは少ないものの、情報が見あたらずに診療が阻害されることもありえるため、QI の内容には含まれることがある。これらは、例えば、がんの分野では取り扱い規約に沿った診療録記載など、情報共有に役立つような統一された記載方法であるといったことが挙げられる。

また、QI については「質の評価」をうたっている以上は、個々の患者の状態に応じた標準から離れた調整といった事柄も加味する必要がある。患者の状態に応じた調節を標準医療から外れているからといって「質が低い」と評価してはならない。そのため評価する側の責任も認識しつつ運用する必要がある。

このように細かい点が異なるため、ガイドライン推奨と QI は表裏一体の関係にあり、同一視するべきではなく、解釈には十分な注意が必要である。

## 5. どのようにして Quality Indicator を作るのか

評価の基準を作成するということは慎重に行わなければならない。そのため、基準を作成する手順は特定のグループに偏ることなく、広く納得されるものでなければならない。そのようなことに気を遣った方法として、米国カリフォルニア大学ロサンゼルス校およびランド研究所で開発された、RAND/UCLA 適切性評価法の手順が(デルファイ変法と呼ばれることもある)世界中に広がって国際標準となっている<sup>17)</sup>。全体の流れを図7に示す。具体的には、まず、専門家数名が QI の候補を選び、エビデンスに基づいてそれぞれの QI 候補がなぜ QI として適切か、という根拠のまとめを作成する。同時に各臓

器のがんの診療にあたる専門家をできるだけ偏りなく 10 名程度組織して専門家パネルとする。これは内科、外科などの診療科だけではなく地域的な多様性も考慮する。次に、この専門家パネル委員それぞれに、各 QI 候補および根拠のまとめを郵送し、各委員は個別に QI 候補を吟味、その QI としての適切性について評価する。その評価結果をまとめ、集計を見ながら、1 日程度、検討会議を開催する。その際に QI の修正なども含めて再検討し、会議の内容を元に各委員が QI に対して第 2 回評価を行う。その集計で、委員の多くが適切だと評価し、かつ反対者がごく少数であったものだけを最終的な QI として採用する。このような手順により、それぞれの QI 候補について詳細かつ多角的な吟味が加えられて完成することになる。以下にそれぞれの段階の詳細を述べる。

### a. QI 候補作

QI の候補となる情報源は主に 3 つある(図 8)。1 つ目は海外で既に QI として使用されているものである。もちろん、技術や医療制度の異なる海外の標準がそのまま日本で使えるわけではないが、明文化されており、吟味もしやすく、検討の俎上にのせる候補として扱うには便利である。米国では、医療研究品質局 (Agency for Healthcare Research and Quality) が、National Quality Measures Clearinghouse という、Quality Indicator (ここでは、Quality Measure と呼ばれている) のデータベースを作成している。先行文献やこのデータベースなどから指標の候補を作成することができる。

2 つ目は、各疾患について専門学会が発行する診療・治療ガイドラインである。各ガイドライン推奨で、推奨度の低いものは QI として使いづらいが、推奨度の高いものを参考

にして作ることができる。この中から例外がかなり多いと考えられるものや、判断の幅が大きくて基準としては使いづらいものを除外・改変して QI 候補を作成する。ここでの留意点がガイドラインは情報の伝達を目的としているため、原則を述べており例外についての考察を紙面の都合などで省略していることもある。そのような例外に対する考察を行い、基準として明確化していくこともガイドライン推奨から QI へ変換する上での重要なポイントである。

3つ目の情報源は専門家の考える新しいアイデアである。手術記載でこのような事柄を記載するのがよいとか、術前検査でこの部分の画像が注意されているか、また日常診療の中でこのようなことが抜け落ちていることが多くて困る、などのアイデアを元に、QI 候補としたものも数多くある。

候補が決まったら、次に、その根拠のまとめを作成する。全ての QI 候補に根拠となるランダム化比較試験が存在するわけではないが、疫学データなどを総合して、ベスト・エビデンスに基づいて根拠づけをする。ガイドラインや先行研究を元に作成したものは、既にその作成時点までの文献検索は行われているので、その作成時以降の文献を中心に検索を行い最新の知見を盛り込んだ。この作業は各臓器 2~5 名で分担して行うのが作業量的にも望ましい。

#### b. 専門家パネル委員の選定

各臓器において 10 名程度の専門家パネルを選定する。この数に決まった定数は存在しないが、伝統的には 9 名で行われてきた。しかし、最近の同様の手法については、診療が多様化して様々な職種のチーム医療が強調されるに至り、15 名程度の大きなパネルが構成されることも多い<sup>12,18</sup>。あまり人数が多

いと、多様な意見を拾い上げて集約するという意味では人数を増やすことがよい方向に作用するが、逆に人数が多いと一人一人の意見を表明する機会が減るといった問題もあるので、バランスを考えることは非常に重要である。

パネル委員の構成としては、最低限各疾患の診療にかかわる内科・外科医、癌であれば化学療法の専門家、内視鏡治療の専門家など関連する各科から最低 1 人委員を選出することが必要である。また、大学病院や専門施設だけではなく、診療の場としても一般病院の臨床医の委員も意識して選出することが望ましい。また、診療パターンや医療状況に地域特性も考えられるため地域的な多様性も考慮したパネル選出を行う。可能であれば世代的な多様性も確保して、議論のできる若手なども含まれると、検討の幅が広がるといわれる。

蛇足ながら可能であれば同様の立場の人は 2 人以上含められた方がよい。これは、1 人だけで、特定のグループ（や診療科）の意見を代表するのは、委員自身にとっても心理的に重荷であることもあるし、個人的な意見なのか、その集団の代表的意見なのかという判断が難しくなるためである。

#### c. パネル専門家による事前の QI 候補評価

選定された QI 候補に図 9 のような 1-9 の評価スケールを付け、パネル検討会の 3 週間程度前に、根拠のまとめとともに各パネル委員に郵送する。この時期についてはリストの長さにもよるため、定まったものはないが、大体 1 ヶ月半前から 3 週間前とするのが適切であろう。忙しいパネル委員が時間を見つけるに当たってあまり時間がないのもよくないが、回答からパネル検討会まで時間が空いてしまっても内容が不確実になる側面があ

る。

各パネル委員は根拠と自らの意見を元に、各 QI 候補が QI として適切かを 1=きわめて不適切、9=きわめて適切、のスケールで評価する。その際 QI 候補の表現や内容に関して修正の提案があれば、コメントとして記入する (図 8)。

QI としての適切性は、「原則としてそのような診療を行うべきといえるかどうか」、また、「行っていないことは、概して質が低いといえるのか」を出発点とするのが手始めにはよいと考えられ、そこから派生して質指標として「準拠している割合が多い程質が高い診療を提供している」といえるかどうか、が基準となる。このように「行っていないことが概して質が低いといえる」ことを基準とするのは非常に低いレベルのように考えられるが、これから質の問題を見つけて改善しようという際には、このレベルとすることが妥当と考えられる。もちろん、後者の測定の部分が重要であり、「行われない場合に質が低い」とはいえないまでも「行っている方が、行っていない場合よりもほぼ確実に質が高いといえる」場合には、多少ハードルが高いことであっても質の指標となりうる。

パネル委員は評価後に事務局に評価結果を郵送し、それを元に事務局で集計を用意、検討会に備える。

#### d. パネル検討会と第 2 回評価

集計を元に各臓器、全員参加でパネル検討会を開く。ここでは、誰がどの評価を付けたかについては伏せたまま評点のばらつきなどを見ながら検討を行う。QI の候補を会議室前方のスクリーンに投影し、適宜表現の修正を加えたものが共有できるようにする。その修正後の QI について、ひとつひとつの QI について議論が終わり次第、第 2 回評価を各

委員が個別に行う。その中で QI の使われ方を再度確認し、また適宜根拠についても吟味する。また、議論の過程において新しい QI 候補が提案された場合には、その QI について多角的に評価して第 2 回評価に加えることになる。

#### e. 最終集計

第 2 回評価を元に、基本として中央値が 7 以上で、1~3 とした評価委員数が 2 名以下のものを最終的な QI として採用し、最終的な QI セットとするのが伝統的に標準である<sup>17)</sup>。パネルの人数が多くなる場合には、この人数を調節する必要がある。

同様の QI で同時に成り立たないもの、例えばフォローの間隔のみが異なり「手術後の画像検査によるフォローは 3 ヶ月おき」、というものと「6 ヶ月おき」というものが両方も採用となった場合には、反対の少ないもの、中央値の高いものを採用、まったく同程度の分布の場合には、基準として緩やかな方を採用とする。

#### f. 本方法の特徴

以上のような手順が、国際標準とされる RAND/UCLA 適切性評価法である。この方法には、

- ① 検討会前に各委員が個別に検討する機会を提供する。
- ② 数値で適切性を評価することで、各委員の意見が可視化される。
- ③ 関連する診療科や地域特性などパネル委員構成を工夫することで多角的検討が可能になる
- ④ 検討会で議論することで、評価方法が確認できるとともに、QI の解釈などに様々な見方が反映され、完成度の高い QI の作成が可能になる