

肝臓診療ガイドライン 2017 年版の外部評価

佐野圭二*, 伊佐山浩通**, 國分茂博***,
坂巻顕太郎†, 杉原健一††, 原留弘樹†††

はじめに

『科学的根拠に基づく肝臓診療ガイドライン 2005 年版』が初版として上梓され、3~4 年ごとの改訂という当初の予定通り、「2009 年版（第 2 版）」、「2013 年版（第 3 版）」に続いて「2017 年版（第 4 版）」が最新のエビデンスを加えて公表されることとなった。今までもガイドライン公表に伴って毎回外部評価が行われた。2005 年版では「ガイドライン評価委員会」がガイドライン完成直後に組織され、内容の妥当性、普及・利用可能性についての評価が同ガイドラインの巻末に掲載された。2009 年版では Web でのパブリックコメント募集と日本肝臓学会総会での討議を経て出版され、その後外部評価委員会が組織されその評価が日本肝臓学会誌に寄稿された¹⁾。2013 年版でも同じく Web 上でのパブリックコメント募集、日本肝臓学会総会での公聴会を経た後に 2013 年版が上梓され、その後外部評価委員会が組織されその評価が報告された。そのなかで、ガイドラインの公表に先立って外部評価がなされていないことを問題として指摘された。この評価を受けて、今回の 2017 年版ではガイドライン公表前に外部評価委員会が組織され、草稿の時点で評価が行われた。

目的

ガイドラインの定義として、平成 14 年度に公益財団法人日本医療機能評価機構が開始した EBM 普及推進事業〔Minds (マインズ): Medical Information Network Distribution Service〕のなかで 2014 年に定められた「診療上の重要度の高い医療行為について、エビデンスのシステマティックレビューとその総体評価、益と害のバランスなどを考慮し、最善の患者アウトカムを目指した推奨を提示することで、患者と医療者の意思決定を支援する文書」²⁾というものがある。この定義を念頭に置きながら、当ガイドラインの改訂過程の妥当性や臨床現場への適用可能性などを客観的に評価し、その結果を当該のガイドラインに反映させることが外部評価の目的である。2017 年版の外部評価はガイドライン公表直前に行われ、次回以降の改訂版の参考となるように外部評価をまとめることを目的とした。

評価方法

外部評価委員として、ガイドラインの改訂に携わっていない肝臓専門評価者（肝臓専門医）3 名（内科、外科、放射線科各 1 名）、専門外評価者 3 名（肝臓診療を専門としないが臨床ガイドライン

* Keiji Sano. 帝京大学医学部外科学講座 教授

** Hiroyuki Isayama. 順天堂大学医学部附属順天堂医院消化器内科 先任准教授

*** Shigehiro Kokubu. 新百合ヶ丘総合病院肝疾患低侵襲治療センター センター長

† Kentaro Sakamaki. 東京大学大学院医学系研究科生物統計情報学講座 特任講師

†† Kenichi Sugihara. 東京医科歯科大学 名誉教授

††† Hiroki Haradome. 北里大学医学部新世紀医療開発センター先端放射線画像解析学 教授

に精通する医師として内科、外科各1名と生物統計専門家1名)の計6名が選任された。その6名によって独立した評価を行い、その評価を集積して解析した。

今回用いた診療ガイドラインの評価ツールは、2013年版と同様、AGREE (Appraisal of Guidelines for Research and Evaluation)³⁾の改訂版であるAGREE II⁴⁾でのみ評価した(表1)。2005年版、2009年版の外部評価には、AGREE、Shaneyfelt法⁵⁾、COGS (Conference on Guideline Standardization)法⁶⁾の計3法が用いられたが、3法ともほぼ同様の質問項目から構成されていること、以前の評価で3法ともほぼ同じ結果であったこと、AGREE IIのみが領域別に質問が分類されており各質問項目が7段階評価であること(他の2法は全項目「Yes」「No」の選択)、などから今回もAGREE IIのみでの評価とした。

AGREE IIは、「対象と目的(3項目)」、「利害関係者の参加(3項目)」、「作成の厳格さ(8項目)」、「提示の明確さ(3項目)」、「適用可能性(4項目)」、「編集の独立性(2項目)」の6領域計23項目と、総合評価として最後に2つの質問が用意されている(表1)。各項目は、総合評価の最後の質問を除いて、「まったくあてはまらない(1点)」から「強くあてはまる(7点)」までの7段階で評価される。総合評価の最後の1つの質問は、「推奨する」「変更後推奨する」「推奨しない」の3段階評価である。また、各項目にはコメントを記載するための欄が設けてある。各項目の評価点から、各領域の満点を100%とした、獲得点数をパーセンテージに変換したスコア(領域別スコア, domain score)を計算する。そのスコアを用い、ガイドラインの優れている領域、劣っている領域を、付記されたコメントとともに、評価する。

2017年版の外部評価と以前の外部評価との比較については、外部評価の目的から外れている点、評価ツールは同じであるが評価者が異なる点などから行わないこととしたが、特記すべきと考えられる変化に関してのみ記載することとした。

結果

領域ごと、項目ごとの得点と領域別スコアを表1に示す。肝臓専門評価者と専門外評価者を比較すると、「利害関係者の参加」以外の領域別スコアの差はどれも10%未満であり、ほぼ同等の評価がなされていると考えられた(表1)。評価者全体の領域別スコアを検討すると、「適用可能性(43%)」のみ低評価であり(その要因については後述する)、診療ガイドラインの目的である「最善の患者アウトカムを目指した推奨を提示する」ことを評価する領域の「提示の明確さ」については80%と高評価であった(図1)。

個々の項目についても、領域別スコアを出して肝臓専門評価者と専門外評価者を比較すると、その評価が20%以上乖離したものは、項目4、項目5、項目15、項目18(以上専門外評価者が低評価)、項目20(肝臓専門評価者が低評価)であった(図2:赤線)。項目4「ガイドライン作成グループには、関係するすべての専門家グループの代表者が加わっている(94%と67%)」では、特に専門外評価者から「メンバーの役割の記載がない」との指摘があがり、委員リストより詳細な記載が求められた。また緩和ケア・看護師・患者などから委員を選出すべきとのコメントが寄せられた。項目15「推奨が具体的であり、曖昧でない(89%と61%)」においては、推奨文から具体的なアクションにつながらない(CQ27「待機的肝切除において腹腔ドレーンは必ずしも必要としない」や、CQ47「分子標的治療では、各薬剤に固有の高頻度または重篤な副作用が報告されており、慎重に経

表1 AGREE II での各項目におけるスコア率

Domain (領域)	Item	評価項目	
1 Scope and Purpose (対象と目的)	1	The overall objective (s) of the guideline is (are) specifically described.	ガイドライン全体の目的が具体的に記載されている
	2	The health question (s) covered by the guideline is (are) specifically described.	ガイドラインで取り扱う健康上の問題が具体的に記載されている
	3	The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described.	ガイドラインの適用が想定される対象(患者、一般など)が具体的に記載されている
2 Stakeholder Involvement (利害関係者の参加)	4	The guideline development group includes individuals from all relevant professional groups.	ガイドライン作成グループには、関係するすべての専門家グループの代表者が加わっている
	5	The views and preferences of the target population (patients, public, etc.) have been sought.	対象集団(患者、一般など)の価値観や好みが十分に考慮されている
	6	The target users of the guideline are clearly defined.	ガイドラインの利用者が明確に定義されている
3 Rigour of Development (作成の厳密さ)	7	Systematic methods were used to search for evidence.	エビデンスを検索するために系統的な方法が用いられている
	8	The criteria for selecting the evidence are clearly described.	エビデンスの選択基準が明確に記載されている
	9	The strengths and limitations of the body of evidence are clearly described.	総体としてのエビデンスの強固さと限界が明確に記載されている
	10	The methods for formulating the recommendations are clearly described.	推奨を決定する方法が明確に記載されている
	11	The health benefits, side effects, and risks have been considered in formulating the recommendations.	推奨の決定にあたって、健康上の利益、副作用、リスクが考慮されている
	12	There is an explicit link between the recommendations and the supporting evidence.	推奨とそれを支持するエビデンスとの対応関係が明確である
	13	The guideline has been externally reviewed by experts prior to its publication.	ガイドラインの公表に先立って、外部審査がなされている
	14	A procedure for updating the guideline is provided.	ガイドラインの改訂手続きが示されている
4 Clarity of Presentation (提示の明確さ)	15	The recommendations are specific and unambiguous.	推奨が具体的であり、曖昧でない
	16	The different options for management of the condition or health issue are clearly presented.	患者の状態や健康上の問題に応じて、可能な他の選択肢が明確に示されている
	17	Key recommendations are easily identifiable.	どれが重要な推奨が容易に見分けられる
5 Applicability (適用可能性)	18	The guideline describes facilitators and barriers to its application.	ガイドラインの適用にあたっての促進要因と阻害要因を記述している
	19	The guideline provides advice and/or tools on how the recommendations can be put into practice.	どのように推奨を適用するかについての助言・ツールを提供している
	20	The potential resource implications of applying the recommendations have been considered.	推奨の適用に伴う付加的な資源が考慮されている
	21	The guideline presents monitoring and/or auditing criteria.	ガイドラインにモニタリング・監査のための基準が示されている
6 Editorial Independence (編集の独立性)	22	The views of the funding body have not influenced the content of the guideline.	資金源によりガイドラインの内容が影響されていない
	23	Competing interests of guideline development group members have been recorded and addressed.	ガイドライン作成グループメンバーの利益相反が記載され、適切な対応がなされている
Overall Guideline Assessment (全体評価)	1	Rate the overall quality of this guideline.	このガイドライン全体の質を評価してください
	2	I would recommend this guideline for use.	このガイドラインの使用を推奨する

	肝臓専門			専門外			全体		
	得点	率	領域別スコア	得点	率	領域別スコア	得点	率	領域別スコア
	15	67%	72%	17	78%	70%	32	72%	71%
	21	100%		19	89%		40	94%	
	12	50%		11	44%		23	47%	
	20	94%	83%	15	67%	61%	35	81%	72%
	14	61%		9	33%		23	47%	
	20	94%		18	83%		38	89%	
	14	61%	78%	12	50%	76%	26	56%	77%
	14	61%		14	61%		28	61%	
	18	83%		16	72%		34	78%	
	19	89%		19	89%		38	89%	
	16	72%		15	67%		31	69%	
	15	67%		16	72%		31	69%	
	21	100%		21	100%		42	100%	
	19	89%		21	100%		40	94%	
	19	89%	83%	14	61%	76%	33	75%	80%
	15	67%		16	72%		31	69%	
	20	94%		20	94%		40	94%	
	15	67%	42%	11	44%	44%	26	56%	43%
	18	83%		16	72%		34	78%	
	6	17%		11	44%		17	31%	
	3	0%		6	17%		9	8%	
	19	89%	78%	19	89%	75%	38	89%	76%
	15	67%		14	61%		29	64%	
	17	78%		16	72%		33	75%	
	Yes	2		Yes	1		Yes	3	
	Yes with modification	1		Yes with modification	2		Yes with modification	3	
	No	0		No	0		No	0	

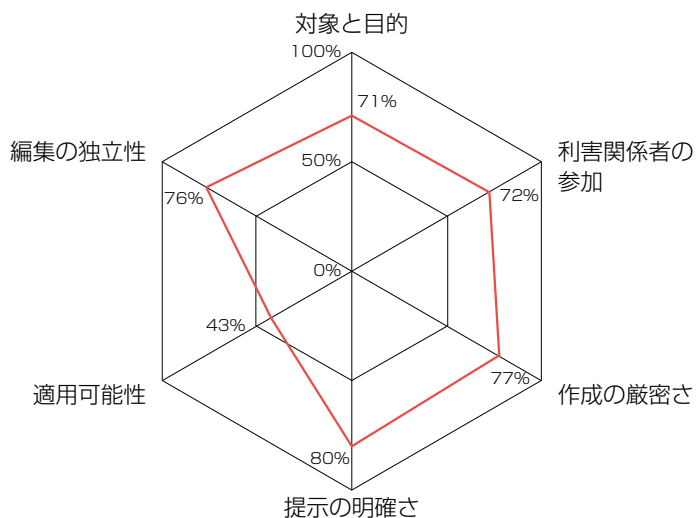


図1 領域別スコア

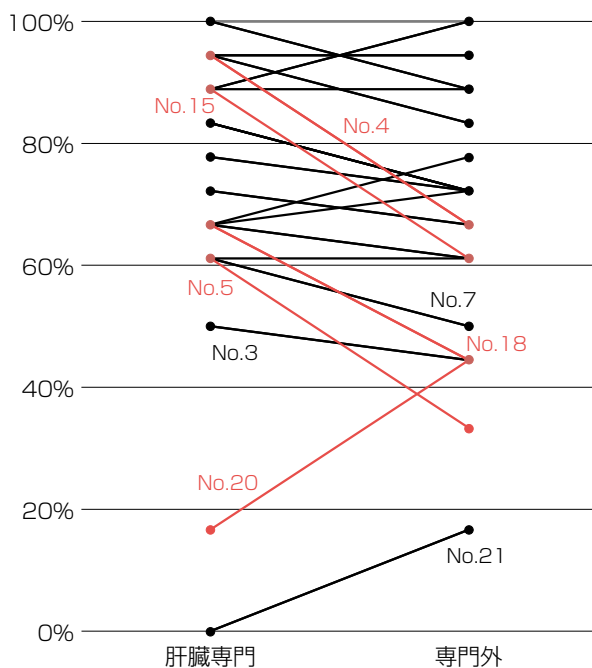


図2 肝臓専門評価者と非専門評価者による各項目別スコアの相違
赤線：肝臓専門評価者と専門外評価者の評価が20%以上乖離したもの

過観察し、減量・休薬を含め適切に対処する必要がある」など)ものが専門外評価者からの低評価につながった。専門、専門外合わせて低評価であった項目5, 18, 20は次段に記載する。

評価者全員で特に評価が低かったもの(スコア率60%以下)は6項目に及んだ。「対象と目的」の領域では、項目3「ガイドラインの適用が想定される対象(患者、一般など)が具体的に記載されている(47%)」のみが専門・専門外ともに低評価であった。理由は当ガイドラインの表題含め「肝

癌」という記載が「原発性肝癌」すべてを含むのか「肝細胞癌」のみなのか曖昧であったためである（筆者注：しかし外部評価施行後に草案も変更されていたので公表時は問題ないと思われる）。「利害関係者の参加」の領域では、項目5「対象集団（患者、一般など）の価値観や好みが十分に考慮されている（47%）」が特に専門外評価者から低く評価された。これはガイドライン改訂委員にも外部評価委員にも患者代表が選ばれておらず、ガイドライン作成、日本肝癌研究会での公聴会、日本肝臓学会ホームページ上でのパブリックコメント募集、外部評価、いずれの段階でも患者の声が極めて反映されにくい状況であったと推測されたことが理由である。「作成の厳密さ」の領域では、項目7「エビデンスを検索するために系統的な方法が用いられている（56%）」の評価のみが低かった。これは文献検索式についての記載が、前版までは巻末に記載されていたものが今回前版同様であることが多いためかまとまった記載がなくなり、前版を参考にしないと把握できないからである。「提示の明確さ」の領域ではすべて高評価であったが、「適用可能性」の領域では4項目のうち3項目で低評価であった。まず項目18「ガイドラインの適用にあたっての促進要因と阻害要因を記述している（56%）」では、移植医療、分子標的治療薬、粒子線治療など高額医療に関してそのコストベネフィットや保険適用に関する記載が不十分との指摘がなされた。それらの治療法に関しては項目20「推奨の適用に伴う付加的な資源が考慮されている（31%）」でも、前版巻末に掲載されていたソラフェニブ治療の費用対効果についての医療経済委員のコメントのような記載がないとの指摘がなされ低評価につながった。項目21「ガイドラインにモニタリング・監査のための基準が示されている（8%）」に関しては、全項目中最低評価であり、全委員から「記載がされていない」との指摘があったが、これは2005年版から毎回指摘され続けている。「編集の独立性」の領域では低評価項目はなかった。

その他スコア率は60%以上であった項目に対する特筆すべきコメントを列挙する。項目9「総体としてのエビデンスの強固さと限界が明確に記載されている（78%）」と項目10「推奨を決定する方法が明確に記載されている（89%）」において、今回採用されたGRADEシステムについての説明が乏しいこと、合議や挙手による推奨度の決定はバイアスが生じるので好ましくないこと、とのコメントがあった。項目23「ガイドライン作成グループメンバーの利益相反が記載され、適切な対応がなされている（64%）」では、今回利益相反（COI）に関して初めて記載されたが、より詳細な記載（各委員の各CQに関するCOI、そしてfundに関するもののみならずacademic COIも含め）をすべき、との意見が全外部評価委員から寄せられた。

全体評価としては、「ガイドライン全体の質」のスコア率が75%と高値であり、「推奨するかどうか」の質問には、「Yes」が3人、「Yes, with modifications」が3人で、「No」の回答はなかった。

まとめ

今回、肝臓診療ガイドライン 2017年版の外部評価を AGREE II に沿って行った。次のガイドライン改訂の際に参考にしてもらいたいと考える評価を列挙する。

- ・ガイドライン改訂委員については、以下2点を留意されたい。①緩和ケア・看護師・患者などからも選出すること、②各委員の役割のみならず、経済的・業績的COIを明記すること
- ・ガイドライン改訂の具体的方法（CQの設定、文献検索、推奨の決定など）について、より詳細に網羅的に記載すること

- ・各 CQ の推奨文をできるだけ曖昧でなく具体的なものにすること
- ・ガイドラインの適用にあたっての促進要因と阻害要因（コストなど）、施設資源・医療経済の観点（特に移植医療・分子標的治療薬・粒子線治療などにおいて）の検討を加えること
- ・当ガイドラインの汎用性や有用性の評価法について記載すること

引用文献

- 1) 山口拓洋, 角谷眞澄, 泉並木, 山口俊晴, 久保正二, 菅野健太郎. 肝癌診療ガイドライン改訂版 (2009 年度版) 外部評価の結果. 肝臓, 2010; 51: 591-8.
- 2) 森實敏夫, 吉田雅博, 小島原典子 編. Minds 診療ガイドライン作成の手引き 2014, 東京, 医学書院, 2014.
- 3) 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. PMID: 12571340
- 4) Brouwers M, Kho ME, Browman GP, et al. AGREE II: Advancing guideline development, reporting and evaluation in health care. J Clin Epidemiol. 2010; 63: 1308-11. PMID: 20656455
- 5) Shaneyfelt TM, Mayo-Smith MF, and Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. JAMA 1999; 281: 1900-5. PMID: 10349893
- 6) 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. PMID: 13679327