

EBMと臨床への応用

EBMを超えて

診療ガイドラインの動向や新しい潮流 (GRADE) を
中心として

相原 守夫

- 診療ガイドラインのためのエビデンス検索は、EBM 情報源ピラミッドを活用し、「サマリーやガイドライン>事前評価済み研究>非事前評価研究の順」とすることが効率的である。
- 信頼できるガイドラインを作成するためには、GRADE-ADOLPMENT に準じて、既存の推奨の採用 (adoption)、適用 (adaptation)、新規開発 (de novo creation) のいずれかを利用することが迅速かつ効率的である。
- ROBINS-I と RoB 2.0 は論理的に共通の項目を使っているため、2 種類のエビデンス (RCT と NRS) に対処する際に、これらのバイアスのリスクツールを使って評価された場合、可能であれば両者を統合できる。
- 推奨を支持するエビデンス情報源はその最新性が重要であり、リビング推奨 (living recommendation) のためには、リビングシステマティックレビュー (living systematic review) の更新アプローチが重要である。

はじめに

診療ガイドラインと典型的な教科書の違いは、教科書が「どのように、なぜそれが機能するのか (how or why does it work)」という背景疑問を扱うのに対して、診療ガイドラインでは「何をすべきか (what to do)」という前景疑問に対して実行可能なステートメントとして回答を提供することにある。米国医学研究所による定義では、「診療ガイドラインは、エビデンスのシステマティックレビューと、複数の治療選択肢の利益と害の評価に基づいて患者ケアを最適化するための推奨を含む文書」である¹⁾。診療ガイドラインにおける推奨は、期限切れの古いエビデンスではなく最新のエビデンスに基づいた“生きている”ものでなければならない。すなわち、システマティックレビューを最新に保つこ

あいほら もりお：相原内科医院 (〒036-8062 青森県弘前市青山3-8-2) / 弘前大学医学部付属病院消化器血液内科

とは容易ではないが、そうしないと、レビューの正確性、実用性、信頼性が低下し、診療ガイドラインにおける推奨の決定が不適切なものになる可能性がある。

本稿では、診療ガイドラインやシステマティックレビューの効率的作成と最新性、信頼性を保つための新しいアプローチを述べ、このプロセスにおいて Grading of Recommendations Assessment, Development and Evaluation (GRADE) アプローチ (以下、GRADE) がどのように役立つのかを解説する。

GRADE アプローチ

GRADE は、エビデンスの要約 (例: システマティックレビューや医療技術評価) から効果推定値の確実性を評価するための構造を提供し、ガイドライン作成の過程では、エビデンスから推奨度を作成する構造を提供する²⁻⁴⁾。GRADE では、システマティックレビューにおけるエビデンスの確実性は、「ある効果推定値が正しいという確実性が、どの程度か」と定義される。診療ガイドラインにおけるエビデンスの確実性は、「効果推定値に対する確実性が、ある特定の推奨を支持するうえでどの程度十分か」と定義される²⁾。このプロセスでは、効果推定値に対する確実性は推奨の強さに重要な影響を及ぼすが、2つは全く別のものである。つまり、高い確実性の推定値と低い確実性の推定値のどちらも、強い推奨または弱い (条件付き) 推奨のいずれかと関連する可能性がある。

診療ガイドラインにおける推奨の強さ (強い、弱い/条件付き) と方向 (介入賛成/反対) を決定する2つの主な要因は、利益と害のバランス、および支持するエビデンスの確実性 (または質) であるが、他の要因 (例: 価値観や意向、コストや費用対効果、公平性、許容可能性、実行可能性) も検討する必要がある⁵⁻⁹⁾。

表1に、GRADE に従ったエビデンス総体の確実性評価を示す。GRADE の基本では、ランダム化比較試験 (randomized controlled trial: RCT) のエビデンス総体は、確実性の高いエビデンスとしてスタートするが、それらは5つのドメイン (バイアスのリスク、非直接性、不精確さ、非一貫性、出版バイアス) によって等級ダウンできる。一方、観察研究のエビデンス総体は、確実性の低いエビデンスとしてスタートするが、方法論的に妥当で、3つのドメイン (大きな効果の程度、用量反応勾配、もっともらしい交絡による提示効果の減少) によって、中または高の確実性のエビデンスに等級アップできる¹⁰⁾。エビデンスの確実性の評価はアウトカムレベル、研究全般にわたって行われることが重要である。また、エビデンスの確実性や推奨の強さを決定するためには、エビデンスプロファイルを作成し、推奨を決定するための GRADE アプリ GRADEpro GDT を使用するべきである¹¹⁾。

推奨のためのエビデンス検索

推奨を作成する際には、関連するすべての研究エビデンスの要約が不可欠で、理想的には質の高いシステマティックレビューに基づくべきである。意思決定のためのエビデンス

表1 エビデンスの確実性を等級付けするための GRADE アプローチ

確実性の初期レベル		確実性のグレードダウンとグレードアップ		確実性の最終レベル
研究デザイン	エビデンス総体の初期評価	グレードを下げる* (downgrade)	グレードを上げる* (upgrade)	エビデンス総体の最終評価 (GRADE)
ランダム化比較試験	高	<ul style="list-style-type: none"> ●限界 (risk of bias) <ul style="list-style-type: none"> -1 深刻 -2 非常に深刻 ●非一貫性 (inconsistency) <ul style="list-style-type: none"> -1 深刻 -2 非常に深刻 	<ul style="list-style-type: none"> ●効果の程度が大きい (large magnitude of effect) <ul style="list-style-type: none"> +1 大きな効果 RR>2 あるいは<0.5 +2 きわめて大きい効果 RR>5 あるいは<0.2 ●用量反応勾配 (dose-dependent gradient) <ul style="list-style-type: none"> +1 あり ●すべての公絡因子 <ul style="list-style-type: none"> +1 提示された効果を減弱させている +1 効果が観察されていないのに当該効果を増大させる方向に働いている 	<ul style="list-style-type: none"> ●高: ⊕⊕⊕⊕ ●中: ⊕⊕⊕○ ●低: ⊕⊕○○ ●非常に低: ⊕○○○
観察研究	低	<ul style="list-style-type: none"> ●非直接性 (indirectness) <ul style="list-style-type: none"> -1 深刻 -2 非常に深刻 ●不精確 (imprecision) <ul style="list-style-type: none"> -1 深刻 -2 非常に深刻 ●出版バイアス (publication bias) <ul style="list-style-type: none"> -1 ありそう -2 非常にありそう 		

* : 1=1段階グレードを下げる, あるいは上げる (例: 「高」から「中」へ). 2=2段階グレードを下げる, あるいは上げる (例: 「高」から「低」へ)

GRADEでは, システマティックレビューの場合は各アウトカムに関するエビデンス総体の確実性の等級付けを実施し, ガイドラインの場合には意思決定に重大なアウトカム全般にわたるエビデンス総体の確実性を1つだけ等級付けする。
(出版社の許可を得て文献10より引用)

を効率的に検索するためには, EBMの基本としての情報源ピラミッドを活用する必要がある (図1)¹²⁾.

情報源ピラミッドには, サマリーとガイドライン, 事前評価済み研究, 非事前評価研究の3つのカテゴリがある. サマリーは, 関連する疑問のトピックレベルでのエビデンス統合を目指して定期的に更新されるオンライン情報源である. この情報源は GRADE のような適切なシステムを用いて推奨の強さをグレーディングし, 基礎となる価値観や意向について明確な判断を提示している. サマリーのカテゴリには, UpToDate, DynaMed, Best Practice などがある. また, 診療ガイドラインの検索に役立つ情報源は, 米国 National Guideline Clearinghouse, Guidelines International Network (GIN), National Institute for Health and Care Excellence (NICE), 日本医療機能評価機構の Medical Information Network Distribution Service (Minds) などがある.

情報源ピラミッドの第2階層に位置するのは事前評価済み研究で, 具体的にはシノプシスやシステマティックレビューである. つまり, サマリーやガイドラインが満足のいく回答を提供しない場合, まずシステマティックレビューから調べ, 必要に応じて, 1次研究の結果を読まなければならない. この情報源には, ACP Journal club, Cochrane

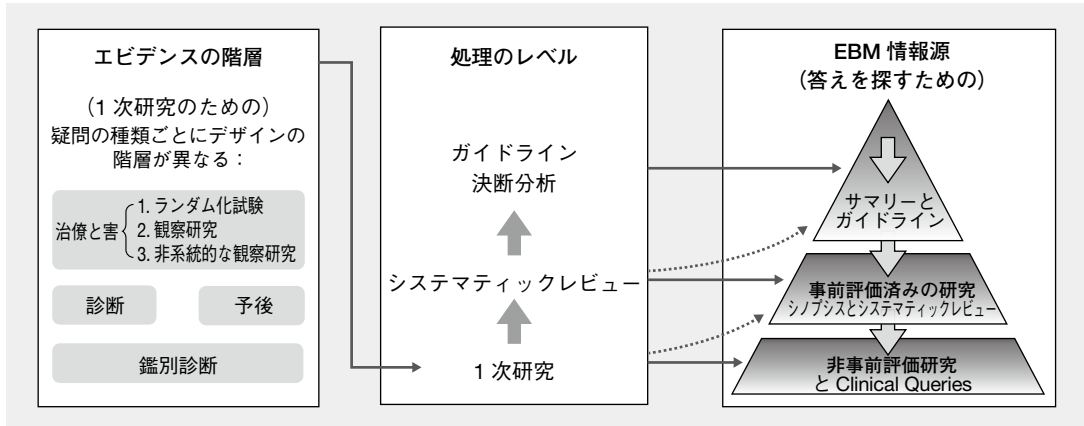


図1 EBM情報源ピラミッド

EBMにおいては、利用可能な情報源の検索方法を明確に確認するために、3つの分類システムがある。これには、①1次研究におけるエビデンスの階層 hierarchy of evidence (左欄)、②エビデンスの処理レベル (中央欄)、③EBM情報源のカテゴリ (右欄) が含まれる。合わせると、これらの3つの分類システムは、1次研究から既存のEBM情報源へのエビデンスの流れを表している。

(出版社の許可を得て文献12より引用)

review, McMaster Plus などがある。ピラミッドの最下層に位置するのが非事前評価研究であるが、MEDLINE や CINAHL のような大規模データベースで1次研究を検索する必要がある。これらのデータベースには数百万の論文が含まれているため、効率的に使用するには高度な検索スキルが必要となり、労力や作業時間も膨大となる。したがって、診療ガイドラインのためのエビデンス検索には、情報源ピラミッドのトップダウン（ボトムアップではなく）様式が効率的であり、このためのいくつかの連合検索エンジン（ACCESSSS, TRIP, SUMSearch, Epistimonikos など）が開発されている。推奨作成において利用するシステマティックレビューの質評価ツールには、PRISMA¹³⁾, AMSTAR¹⁴⁾, AMSTAR2¹⁵⁾ があり、診療ガイドラインの質を評価するためには、AGREE II¹⁶⁾ を利用できる。

GRADE-ADOLOPMENT

推奨作成を目的とするガイドライン作成グループには、通常、次の3つの選択肢がある。①既存の推奨事項を採用する (adoption)。②既存の推奨を各状況に適用させる (adaptation)。③利用可能な統合エビデンスに基づいて新規に推奨を開発する (*de novo* creation)。信頼できるガイドライン作成のためのこのアプローチを、GRADE working group は、GRADE-ADOLOPMENT と呼んでいる (図2)¹⁷⁾。GRADE-ADOLOPMENT を効率的、透明的に実施するためには、その中心要素である GRADE Evidence to Decision (EtD) フレームワークが必須で、実際にこのフレームワークを用いて22件の診療ガイドライン (226個の推奨事項) が1年未満の短期間で作成されている^{3,4)}。EtD フレームワークには、推奨の強さと方向を決定するための基本的な基準 (問題の優先性、望ましい効果、望

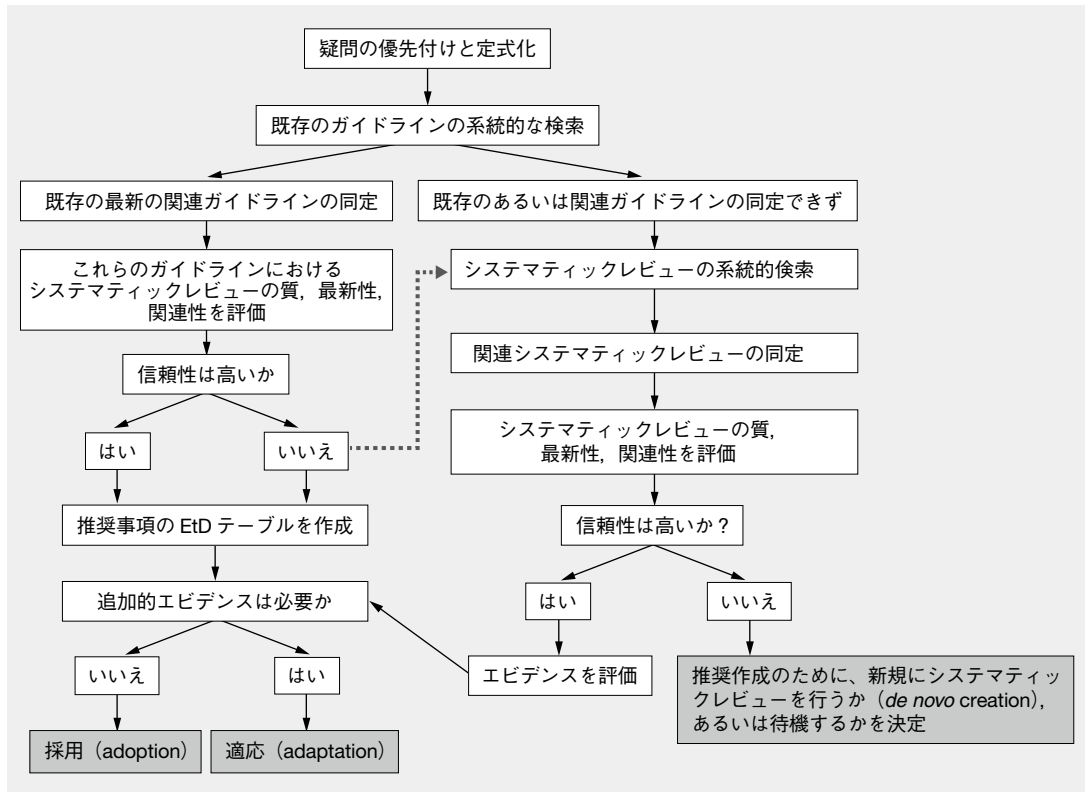


図2 GRADE-ADOLOPMENT

ガイドライン作成者は、他所から既存の推奨を採用する (adoption) か、既存の推奨またはシステマティックレビューを自身の状況に合わせて適応させる (adaptation) か、新規作成する (de novo creation) ことができる。
(出版社の許可を得て文献10より引用)

ましくない効果、エビデンスの確実性、価値観や意向、効果のバランス、資源の利用、公平性、許容可能性、実行可能性などが含まれている。また、GRADE-ADOLOPMENTを使った推奨 (adolop 推奨) は、介入効果について同じエビデンスを使用した他のガイドラインの推奨と異なる場合がある。これは、各臨床疑問において、採用、適応、de novo 開発の組み合わせの重要性を強調している (例：ベースラインリスクの違い、介入効果に関する新しいエビデンス、価値観や意向、リソースの考慮による異なる推奨)。

ランダム化試験と非ランダム化研究の統合

2016年に非ランダム化研究 (non-randomised study : NRS) のバイアスのリスク評価法として、risk of bias in nonrandomized studies of interventions (ROBINS-I) ツールが開発され¹⁸⁾、ROBINS-Iの7項目を使って個々の研究におけるランダム化の欠如がバイアスのリスクにどのように影響しているかを評価できるようになった。NRSのバイアスのリスクを評価するためにROBINS-Iを使用する場合、選択バイアスと交絡の

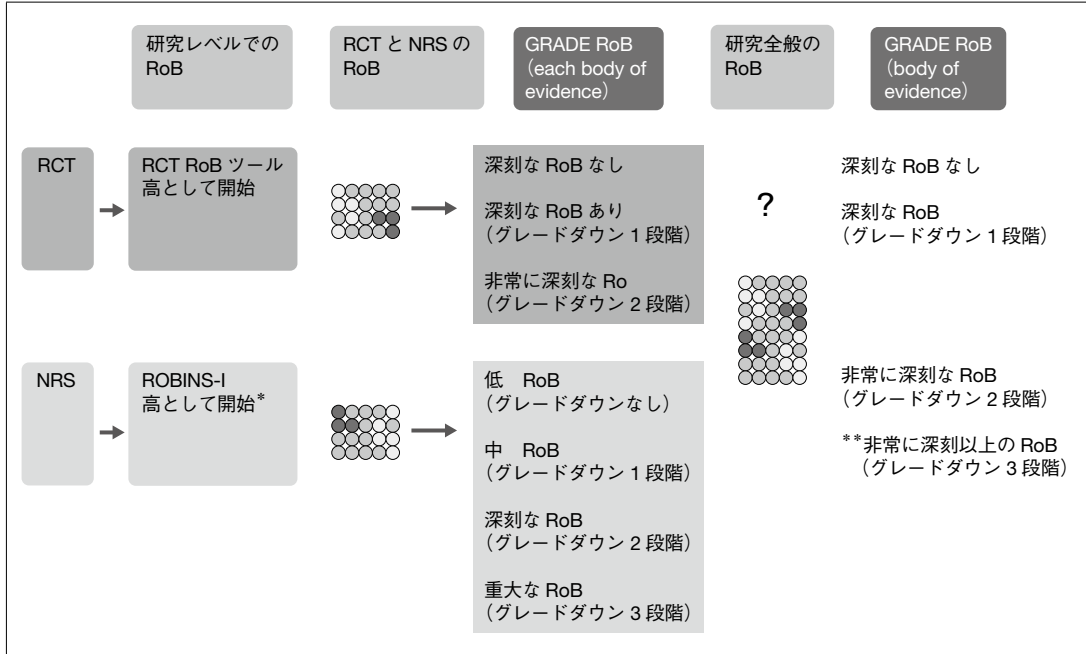


図3 ランダム化試験と非ランダム化研究を GRADE で評価する (ランダム化および非ランダム化研究のバイアスのリスク評価)

2つの種類のエビデンス (ランダム化と非ランダム化) に対処するために、さまざまなツールが使用されている。バイアスのリスク (RoB) は、複数の研究を通して評価され、可能であれば統合される。ROBINS-I を使用する場合、そのツールで重大 (critical) と評価されたエビデンス総体には2段階以上の等級ダウンが必要である。

* : 実際には、NRS についての確実性は少なくとも低または非常に低い2段階等級ダウンにつながるものが一般的である。しかし、大きな効果や用量反応がある結果、またはもっともらしいバイアスにより推論が強化される結果については等級ダウンの程度を下げる可能性がある。ROBINS-I 評価だけで等級ダウンなし、または1段階だけの等級ダウンとなるエビデンス総体はまだ同定されていない。RCT と NRS の統合方法については、今後のGRADE ガイド論文でさらに論じる予定である。

** : GRADE の場合、対応する用語は、深刻ではない (not serious)、深刻 (serious)、非常に深刻 (very serious)、および第4の RoB のレベルである。GRADE は現在、第4レベルの適切な用語を探求している。

ROBINS-I : risk of bias in nonrandomized studies of interventions ; NRS, nonrandomized study ; RCT, randomized controlled trial. RoB : risk of bias (文献 19 より翻訳転載)

※著者注 : 第4のレベルについて、GRADE working group は最近、“きわめて深刻 (extremely serious)” が好ましいと発表した²⁰⁾。

評価が ROBINS-I の不可欠要素であることから、NRS のエビデンス総体における GRADE の最初の確実性は高として開始する (図3)^{19, 20)}。これは、GRADE がランダム化が交絡バイアス防御の唯一の方法であるという従来の見解を変えたことを意味するものではない。したがって、NRS からのエビデンス総体を低い確実性から始めて等級アップの理由を探るか、あるいは高い確実性から開始して等級ダウンの理由を探るかどうにかかわらず、最終的な確実性の等級は同じでなければならない。

一方、同年に発表されたランダム化試験のための改訂 Cochrane RoB 2.0¹⁹⁾ は、結果にバイアスがもたらされる可能性がある5項目で構造化され、RCT の結果に影響を与える可能性があるすべての種類のバイアスをカバーしており、さらなる項目を追加する必要

はない。ROBINS-I と RoB 2.0 は論理的に共通の項目を使っているため、2 種類のエビデンス (RCT と NRS) に対処する際に、これらのバイアスのリスクツールを使って評価された場合、可能であれば両者を統合できると GRADE working group は発表している (図 3)²¹⁾。

リビングシステマティックレビュー (living systematic review)

推奨を支持するエビデンス情報源は、その最新性が重要であり、エビデンスが最新かどうかを判断する簡単な方法は、引用された最新文献の日付を探すことである。もし数年以上古いものであれば、その後に発表された研究や将来の研究によって異なる結論がもたらされる可能性がある。つまり、システマティックレビューは“生きているもの”でなければならない。このような観点から、システマティックレビューの更新アプローチとして、living systematic review (LSR) が 2014 年に提案された^{22~25)}。LSR は、「関連する新たなエビデンスが利用可能になった時点でレビューに組み込むことによって継続的に更新するシステマティックレビュー」と定義されている。LSR においては、効率的なレビュー更新のため、人的検索と機械検索を融合したシステム (例: Cochrane Crowd²⁶⁾) が開発されつつあり、さらには、エビデンスのバイアスのリスク評価 (例: RobotReviewer ツール²⁷⁾) やシステマティックレビューの一部セクションの作成を自動化する試み (例: RevMan HAL: schizophrenia.cochrane.org/revman-hal-v4²³⁾) も登場している。2019 年 7 月時点で、5 件の LSR が Cochrane web サイトに紹介されており (LSR Network: <https://www.cochrane.org/lsr>)、さらには、ネットワークメタアナリシスの living NMA SR への取り組みもある (<http://livenetworkmetaanalysis.com/>)^{28, 29)}。

一方、レビューを最新に保つために、メタアナリシスも更新する必要があるが、標準的なメタアナリシスの方法を使用する場合、より多くの更新が行われるにつれて、それまでの効果なしのメタアナリシスが統計的に有意であると誤って結論付ける可能性 (タイプ I エラー) が急速に高まる。LSR では、このタイプ I エラーやタイプ II エラー (誤って真の効果を検出できない) を調整した手法 (例: 試験逐次分析や逐次メタアナリシスなど) が必要となる²⁴⁾。

リビングガイドライン (living guideline)

推奨を更新する現行モデルは、その公表後の特定の時間間隔 (例: 3 年) でガイドライン全体を更新するという手段を用いている³⁰⁾。この更新アプローチには 2 つの大きな限界がある。第一は、推奨事項の期限切れであり (支持しているシステマティックレビューが期限切れになる)、推奨の妥当性に悪影響を及ぼす。第二の限界は、いくつかの推奨事項については、根底にあるエビデンスが変わらないため、それらの更新に投資された努力が無駄になることである。これは更新プロセスの効率に悪影響を及ぼす。このため、診療ガイドラインにおける推奨とエビデンスの確実性を最新に保つための効率的な作業が必要

となる。

living guideline は、関連する新しいエビデンスが利用可能になった時点ですぐに個々の推奨事項の更新を可能にするためのガイドラインである²⁵⁾。これは、新しいアプローチを表すものではなく、むしろ関連する新しいエビデンスが利用可能になるとすぐに個々の推奨の更新を可能にするための標準ガイドライン作成方法の最適化（例：Guidelines Development Checklist⁶⁾に従う）を表す。ガイドライン更新の効率を最大にするには、living guideline チームと LSR チームの円滑な協調作業プロセスが必須となる。また、このモデルでは、ガイドライン全体としては更新されなくなる。その代わりに、根拠となるエビデンス基盤に大きな変化があるときはいつでも、個々の推奨が更新され、これをリビング推奨（living recommendation）と呼ぶ。この作業には、GRADE を利用した Living Evidence Profile や living EtD テーブルなどの living summary table が必要である²⁵⁾。また、個々の更新推奨を公開する、MAGIC Authoring and Publication Platform (MAGICapp)³¹⁾ や BMJ Rapid Recommendations³²⁾ のようなプラットフォームが必要である。

おわりに

GRADE の登場と普及により、診療ガイドラインや公衆衛生ガイドライン、医療システムに関する推奨や決断を作成できるようになった現在、LSR を含めたガイドライン作成の効率性と最新性がテーマとなりつつある。living guideline の基本となる LSR は新しいアプローチであるが、これにより従来のレビュー更新に伴う作業や資金投資の無駄を大幅に減らすことができるだろう。living guideline の概念の成功と持続可能性は、その主要基盤となる LSR にかかっているが、ガイドラインにおける個々の推奨は“生きている”べきで期限切れのものであってはならない。

● 文献

- 1) Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice Guidelines : Clinical Practice Guidelines We Can Trust. 2011. <https://www.nap.edu/catalog/13058/clinical-practice-guidelines-we-can-trust>. Accessed August 15, 2019.
- 2) Guyatt GH, et al : GRADE : an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336 : 924–926, 2008
- 3) Alonso-Coello P, et al : GRADE Evidence to Decision (EtD) frameworks : a systematic and transparent approach to making well informed healthcare choices. 2. Clinical practice guidelines. *BMJ* 353 : i2089, 2016
- 4) Li SA, et al : Evidence to Decision (EtD) framework provides a structured ‘roadmap’ for making GRADE guidelines recommendations. *J Clin Epidemiol* 104 : 103–112, 2018
- 5) Moberg J, et al : The GRADE Evidence to Decision (EtD) framework for health system and public health decisions. *Health Res Policy Syst* 16 : 45, 2018 (doi : 10.1186/s12961-018-0320-2)
- 6) Schünemann HJ, et al : Guidelines 2.0 : systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ* 186 : E123–E142, 2014
- 7) MacLean S, et al : Patient values and preferences in decision making for antithrombotic therapy : a systematic review : Antithrombotic Therapy and Prevention of Thrombosis, 9th ed. American College of chest Physicians evidence-based clinical practice guidelines. *Chest* 141 : e1S–e23S, 2012
- 8) Hultcrantz M, et al : The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol* 87 : 4–13, 2017
- 9) Zhang Y, et al : Using patient values and preferences to inform the importance of health outcomes in practice

- guideline development following the GRADE approach. *Health Qual Life Outcomes* 15 : 52, 2017 (doi : 10.1186/s12955-017-0621-0)
- 10) 相原守夫 : Part 1 診療ガイドラインと GRADE システム. 診療ガイドラインのための GRADE システム第3版. pp1-122. 中外医学社, 2019
 - 11) GRADEpro GDT : GRADEpro Guideline Development Tool [Software]. McMaster University, 2015 (developed by Evidence Prime, Inc.). (Available from grade.pro.org)
 - 12) 相原守夫 : 最新のエビデンスを探す. 医学文献ユーザーズガイド—根拠に基づく診療マニュアル第3版 (訳). pp37-63. 中外医学社, 2019
 - 13) Moher D, et al : Preferred reporting items for systematic reviews and meta-analyses : the PRISMA statement. *PLoS Med* 6 : e1000097, 2009 (doi : 10.1371/journal.pmed1000097)
 - 14) Shea BJ, et al : AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 62 : 1013-1020, 2009
 - 15) Shea BJ, et al : AMSTAR 2 : a critical appraisal tool for systematic reviews that include randomized or non-randomised studies of healthcare interventions, or both. *BMJ* 358 : j4008, 2017 (doi : 10.1136/bmj.j4008)
 - 16) Brouwers MC, et al : AGREE II : Advancing guideline development, reporting and evaluation in health care. *CMAJ* 182 : E839-E842, 2010
 - 17) Shunemann HJ, et al : GRADE Evidence to Decision (EtD) frameworks for adoption, adaptation, and de novo development of trustworthy recommendations : GRADE-ADOLPMENT. *J Clin Epidemiol* 81 : 101-110, 2017
 - 18) Sterne JAC, et al : ROBINS-I : a tool for assessing risk of bias in non-randomized studies of interventions. *BMJ* 355 : i4919, 2016 (doi : 10.1136/bmj.i4919)
 - 19) Sterne JAC, et al : RoB 2 : a revised tool for assessing risk of bias in randomised trials. *BMJ* 366 : 14898, 2019
 - 20) Piggott T, et al : GRADE Working Group, GRADE notes : Extremely Serious, GRADE's Terminology for Rating Down by 3-Levels, *Journal of Clinical Epidemiology* (2020), doi : <https://doi.org/10.1016/j.jclinepi.2019.11.019>
 - 21) Schünemann HJ, et al : GRADE guidelines : 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. *J Clin Epidemiol* 111 : 105-114, 2019
 - 22) Elliott JH, et al : Living systematic review : 1. Introduction—the why, what, when, and how. *J Clin Epidemiol* 91 : 23-30, 2017
 - 23) Thomas J, et al : Living systematic reviews : 2. Combining human and machine effort. *J Clin Epidemiol* 91 : 31-37, 2017
 - 24) Simmonds M, et al : Living systematic reviews : 3. Statistical methods for updating meta-analyses. *J Clin Epidemiol* 91 : 38-46, 2017
 - 25) Akl EA, et al : Living systematic reviews : 4. Living guideline recommendations. *J Clin Epidemiol* 91 : 47-53, 2017
 - 26) Cochrane Crowd. <http://crowd.cochrane.org/index.html> (Accessed on 15 August 2019)
 - 27) Marshall I, et al : RobotReviewer : evaluation of a system for automatically assessing bias in clinical trials. *J Am Med Inform Assoc* 23 : 193-201, 2016
 - 28) Créquit P, et al : Living network meta-analysis was feasible when considering the pace of evidence generation. *J Clin Epidemiol* 108 : 10-16, 2019
 - 29) Lerner I, et al : Automatic screening using word embeddings achieved high sensitivity and workload reduction for updating living network meta-analyses. *J Clin Epidemiol* 108 : 86-94, 2019
 - 30) Alderson LJ, et al : Median life span of a cohort of National Institute for Health and Care Excellence clinical guidelines was about 60 months. *J Clin Epidemiol* 67 : 52-55, 2014
 - 31) MAGIC Evidence Ecosystem Foundation <http://magicproject.org/about/> (Accessed on 15 August 2019)
 - 32) BMJ Rapid Recommendations <https://www.bmj.com/rapid-recommendations> (Accessed on 15 August 2019)