

診療ガイドライン策定の方法論

1. 一般的な留意事項

一般的疑問(Q: Question)と臨床疑問(CQ: Clinical Question)に分け、それに対する回答はステートメントという形で示した。CQに対するステートメント作成にあたっては、SR(Systematic Review)を行い、推奨グレードを提示した。われわれSRサポートチームは、文献検索を支援したり、エビデンスレベルの確認などを行った。また、作業手引をつくり、策定委員へ周知した。

一般的な留意事項を表1に示した。CQのステートメントの推奨度を判断するために必須である重要論文は、1つだけでなく、すべて引用した。引用すべき文献の検索過程については、その採用基準、および検索式などを策定委員は保存した。文献の採用基準は掲載したが、具体的な検索式は紙面の都合上割愛した。CQのステートメントで引用した文献はアブストラクトテーブルを作成したが、そうでない文献については作成しなかった。すなわち、CQの本文、Qのステートメント、本文の引用文献については、アブストラクトテーブルは不要とした。アブストラクトテーブルには論文コード・対象・方法・結果を含め、研究デザイン名称・エビデンスレベル、そしてバイアスに関する要約を含めた。CQステートメントの推奨グレードを決定するにあたっては、エビデンス総体の確実性・益害バランス・患者の価値観・費用を考慮した。

表1 一般的な留意事項

1. CQステートメントの推奨判断に必要な重要論文はすべて引用すること
2. CQステートメントに引用した論文の採用基準を記載すること
3. CQステートメントの引用文献についてのみアブストラクトテーブルを作成すること
4. アブストラクトテーブルは「デザイン名称」と「エビデンスレベル」を記載すること
5. 推奨グレードの決定にはエビデンス総体の確実性・益害バランス・患者の価値観・費用を考慮すること

2. 研究デザイン名称とエビデンスレベル

アブストラクトテーブルの論文コードの欄に、研究デザイン名称とエビデンスレベルを示した。研究デザイン名称は、①MA/SR(Meta-Analysis=メタ解析[メタ分析], Systematic Review=系統的レビュー), ②RCT(Randomized Controlled Trial=ランダム化比較試験), ③前向きコホート, ④事前設定RCTサブ解析, ⑤後ろ向きコホート, ⑥ケースコントロール, ⑦事後的RCTサブ解析, ⑧単群試験, ⑨横断研究, ⑩症例集積・報告の10種類とした(表2)。エビデンスレベル(証拠の水準)は、1+(最高位)/1/2/3(最低位)の4段階とした(表2)。MA/SRとRCTを除けば、エビデンスレベルはデザイン名称から一義的に決まる。

MA/SRについては、その質が高ければレベルは1+とし、低ければレベルは2とした。

表2 改訂エビデンスレベルの階層構造

デザイン名称	英語 / 説明	エビデンスレベル
MA/SR	Meta-Analysis (メタ解析 [メタ分析]) / Systematic Review (系統的レビュー) 質の高いMA/SR (下記をすべて満たす) ・バイアスリスクは低い ・臨床疑問に直接答えている ・研究結果はほぼ一致している ・誤差は小さく精確な結果である ・出版バイアスは疑われない	1 +
	質の低いMA/SR ・上記5項目のうち満たさない項目がある	2
RCT	Randomized Controlled Trial (ランダム化比較試験) 質の高いRCT (下記をすべて満たす) ・バイアスリスクは低い ・臨床疑問に直接答えている ・誤差は小さく精確な結果である	1
	質の低いRCT ・上記3項目のうち満たさない項目がある	2
前向きコホート	Prospective Cohort Study	2
事前設定 RCT サブ解析	Pre-Specified Sub-Analysis of RCT	2
後ろ向きコホート	Retrospective Cohort Study	3
ケースコントロール	Case-Control Study	3
事後的 RCT サブ解析	Post-Hoc Sub-Analysis of RCT	3
単群試験	Single-Arm (Self-Controlled) Trial	3
横断研究	Cross-Sectional Study	3
症例集積・報告	Case Series/Case Report	3

*: デザイン名称は色字で示した 10 種類である。

レベル1ではなく2とした理由は、MA/SRの元となる研究のレベルが2であれば、それらを統合してもレベルが上がることはないだろうと判断したためである。質の高いMA/SRとは、バイアスリスクが低く、臨床疑問に直接答えており、研究結果はほぼ一致しており、誤差は小さく精確な結果であり、出版バイアスが疑われないものと定義した(表3)。

RCTについては、その質が高いものはレベルは1とし、低ければレベルは2とした。質が高いRCTとは、バイアスリスクが低く、臨床疑問に直接答えており、誤差は小さく精確な結果なものと定義した。アウトカム評価の不完全性、非盲検性、割付時の不備、クロスオーバー法での持ち越し効果、多くの中止脱落例があれば、バイアスリスクは高いと判断した。CQのステートメントがPICO (Participants ; Intervention ; Comparator ; Outcome) に合致しているか否かで、その直接性について判断した。例数が小さ過ぎるのは不精確と判断した。

前向きコホートや事前に設定したRCTによるサブ解析は、レベル2とした。前向きコホートによるMA/SRは、レベルを変えず2とした。質の高いRCTには及ばないだろうという判断であると同時に、MA/SRはレベルを1+か2のみとしたためである。

後ろ向きコホート、ケースコントロール、事後的RCTサブ解析、単群試験、横断研究、症例集積・報告は、すべてレベル3とした。それらに基づくMA/SRも、そのレベルは2とした。

表3 MA/SRおよびRCTの質を評価する項目

MA/SR/RCTのみ、アブストラクトテーブルに質の評価結果を載せた。MA/SRでは5つの要素について、RCTについては3つの要素について評価した。これらの評価に基づいてエビデンスレベルを決定した。

- ①バイアスリスク (Risk of Bias)
アウトカム評価の不完全性、盲検化の欠落または不備、選択的報告、割付法の不備 (RCTの場合)、持ち越し効果が存在 (クロスオーバー法の場合) などのバイアスリスクを評価する。
- ②直接性 (Directness)
対象者、介入内容、比較対照、アウトカム指標が、CQステートメントに適合しているかを評価する。
- ③試験間の結果の一致性 (Consistency, MA/SRのみに該当)
MA/SRでは複数の研究を統合するが、研究結果がほぼ一致 (consistency) していることが前提になる。そうしたことを、フォレスト・プロットや検定結果などから判断する。
- ④精確さ (Precision)
誤差が小さく正確な結果であるかどうかは、例数の目安は設けられないが、95% CIが適切かどうかなどで判断する。統計学的に有意ということだけで判断せず、臨床的有意性も考慮する。
- ⑤出版バイアスがないこと (Publication bias, MA/SRのみに該当)
MA/SRに含めた研究に漏れがないかどうかを判断する。

3. 推奨グレード

策定委員はCQのステートメントに対して、推奨グレードを付した。推奨グレード決定に際して考慮したのが、表4に示した4項目である。引用した文献のエビデンスレベルの高さ、益が害を上回るか、患者様に価値があるか、費用は益に見合うか、という4項目である。これら4項目について、結果と判定根拠を提示した。そして、表5に示したように、CQのステートメントへグレードA(強い推奨)、またはグレードB(弱い推奨)を付した。グレード判定の根拠説明も掲載した。

表4 推奨グレード決定のための4項目

項目	内容	判定
エビデンス総体の確実性	エビデンスレベルの1+/1が含まれるか	はい・いいえ
益害バランス	益が害を上回るか	はい・いいえ
患者の価値観	価値観は患者間で一様であるか	はい・いいえ
費用	費用は正味の利益に見合うものか	はい・いいえ

表5 推奨グレードの決定の仕方

推奨の強弱	グレード	説明
強い推奨	グレードA	推奨グレード決定のための4項目のうち、「はい」の判定が多く、ほぼすべての患者にとって推奨の対象となる行為を行うべきと考えられる。
弱い推奨	グレードB	4項目のうち、「いいえ」の判定が多く、患者により実施すべき行為が異なると考えられる。

策定委員が付した推奨グレードについては、21名の策定委員によるアナライザーを使った投票により、70%以上の合意率で最終決定した。70%に満たない場合は文言などを見直し、再度投票した。3回投票を行っても70%以上の合意が得られなければ、「推奨なし」とした。