

診療ガイドライン策定の方法論

1. 全般的な留意事項

本診療ガイドラインの策定にあたっては、臨床で生じる疑問のうち、実臨床における問題に対する回答を導き出したいものを厳選して、臨床疑問 (Clinical Question : CQ) を設定し、CQ には該当せず、診療ガイドラインで取り上げる必要性が高い疑問は一般的疑問 (Question : Q) として設定した。Q に対する回答はポイント、CQ 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、系統的レビュー (Systematic Review : SR) を行い、推奨グレードを提示した。“科学的根拠 (エビデンス)” となる文献については、システマティックレビューチーム (SR チーム : SR 担当者あるいは策定委員) が CQ ごとに検索を行った。文献検索においては、まず先行する SR/メタ解析 (Meta-Analysis : MA) を検索し、最新で質の高いものがみつかった場合は、個別研究の検索を省略可能とした。RCT が複数抽出されても、特に害の評価においては RCT のみでは限界があるため、可能な限り、害についてはあらゆるデザインの研究をも評価の対象とした。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。

全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、書籍掲載するものは代表的なものを 5 つまで、それ以外については web [日本糖尿病学会ホームページ内 書籍ページ URL (http://www.jds.or.jp/modules/publication/index.php?content_id=4)] に掲載した。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 全般的な留意事項

1. CQ ステートメントの推奨判断に必須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
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

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)**
誤差が小さく精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。

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。例数が少ない場合や信頼区間が広い場合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。

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 のレベルは 2 とした。

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、

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。
決定不能	グレード U	上記の 4 項目を評価しても、現在までのエビデンスでは「行うこと」も「行わないこと」も推奨するまでにはいたらないと考えられる。

益が害を上回るか、患者一様に価値があるか、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。また、上記の 4 項目を評価しても、現在までのエビデンスでは「行うこと」も「行わないこと」も推奨するまでにはいたらない場合、グレード U とした。これらの評価結果と根拠は、各 CQ 末尾にステートメントごとに掲載した。

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、推奨度などを見直し、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(今回は該当事例なし)。