

本ガイドライン作成方法の概略

J-SSCG 2024 は、GRADE (The Grading of Recommendations Assessment, Development and Evaluation) システムに準拠し、1. クリニカルクエスション (clinical question, CQ) の立案、2. 系統的レビューによるエビデンスの検索・収集・統合とその確実性の評価、3. 推奨の策定、の3つの工程を経て作成した。

1. CQ の立案

本診療ガイドラインは、診療の基礎的知識を網羅し、標準的な診療体系の構築を目指した。今回の改訂では、すでに広く普及している臨床課題やエビデンスの確実性が低く今後の推奨が変わりうる臨床課題は CQ として採用せず、敗血症診療において推奨を作成する意義が高い CQ に限定した。重要臨床課題は、推奨を出した CQ だけでなく、情報提示を目的としたバックグラウンドクエスション (background question, BQ)、重要課題だが推奨作成に十分なエビデンスがないフューチャーリサーチクエスション (future research question, FRQ) が含まれる (本文の表1参照)。CQ にはエビデンスの確実性は低いものの、介入による有益性がその有害性を上回る可能性が高く、強い推奨を出すべきと委員会が判断したグッドプラクティスステートメント (good practice statement, GPS) と、系統的レビューを実施して得られたエビデンスを基に推奨を出す GRADE 推奨に分類される。

まず J-SSCG 2020 の全 118 題の重要臨床課題を精査し、エビデンスの質が低く、追加の研究がないものを CQ から除外し、必要な場合は FRQ とした。その後、新規の重要臨床課題を追加し、各領域の委員とワーキンググループメンバーが協力して重要臨床課題 (CQ、BQ、FRQ) の原案を作成し、委員会による承認を経て確定した。

本ガイドラインにおける重要臨床課題の立案ルール

- 1) 質の高いエビデンスがなくても CQ として採用する。エビデンスがない、またはエビデンスの質が低く今後の研究結果によって推奨が変わる可能性が高い場合、これらの課題は FRQ として扱う。
- 2) 過去の J-SSCG や SSCG で採用された CQ の中から、臨床上重要な臨床課題を選択した。
- 3) CQ、FRQ は質問形式とし、PICO [Patient (対象患者); Intervention (評価する介入); Control (対照); Outcome (評価項目)] を決定した。本診療ガイドラインは敗血症患者を対象とするため、対象患者 (P) は敗血症または敗血症性ショックの患者とした。しかし、敗血症患者に限定とした臨床試験の報告が少なく、集中治療患者や人工呼吸患者を対象とした臨床試験の結果を敗血症患者に適用することができる関連領域においては、集中治療患者や人工呼吸患者を対象と設定した。

BQ の策定

BQ は、敗血症の診断や治療など、敗血症診療に必要な基本的知識に関する内容である。BQ に対する回答は推奨ではない。対象とする患者集団と本診療ガイドラインの使用者を考慮し、各ワーキンググループで必要と判断した BQ 案を検討し、委員会による承認により確定した。BQ に対する回答は各ワーキンググループで作成し、委員会の投票により合意形成した。

GPS の策定

GPS は、“飛行機から飛び降りる際にパラシュートを利用するか?” といった極めて常識的で、かつ RCT が倫理的に不可能な CQ に対して提示される。各ワーキンググループが対象とする患者集団と本診療ガイドラインの使用者を考慮したうえで GPS としての提示が相応しいと考える CQ を立案した。GPS に関しては早い時期に CQ に対する推奨を作成し、委員会での投票による合意形成を実施した。GPS として合意できなかった CQ に関しては、再度各ワーキンググループで検討した。

敗血症の定義と診断の提示

敗血症の定義と診断は、世界的に広く使用されている sepsis-3 の基準を採用することとした[1]。担当したワーキンググループで作成し、委員会での合議により決定した。

2. 系統的レビューによるエビデンスの検索・収集・統合

GPS を除く各 CQ に関して網羅的な文献検索を実施し、得られた文献から RCT および必要に応じて観察研究を抽出した。原則として GRADE の方法論に基づいて統合した。

Step 1: 文献検索

文献検索は、CENTRAL、PubMed、医学中央雑誌を含む複数のデータベースから検索を行い、必要に応じて EMBASE、CINAHL、PsycINFO などを追加した。検索式は、Medical Subject Headings (MeSH) terms と自由検索語を用いて、系統的レビューチームの 2 名以上のメンバーにより作成し、ワーキンググループメンバーにより確認された。PubMed の検索では、Cochrane の作成した Sensitive-maximizing version の Search Strategies を用いて RCT を限定した。出版時期や言語に制限は設けず、文献検索式は事前に特定した Key RCTs を含むことを確認したうえで最終決定した。各データベースの文献検索実施日と検索された論文数を記録した。J-SSCG2020 と同じ CQ を採用する場合は、前回文献検索を実施した期間以降の系統的レビューを実施した。

Step 2: 一次スクリーニング

Step 1 で特定された文献のタイトルと抄録をダウンロードし、文献管理ソフトウェア EndNote (Clarivate Analytics、USA) または Mendeley (Mendeley Ltd.、UK) を用いて重複除去を行った後、手作業でさらに重複を除去した。独立した 2 名以上の系統的レビューチームのメンバーが、Rayyan (<https://rayyan.qcri.org/welcome>) を用いて文献タイトルと抄録をレビューし、研究方法と PICO により文献の選別を行った。

Step 3: 二次スクリーニング

Step 2 で残った論文のフルテキストを入手し、独立した 2 名以上の系統的レビューチームのメンバーによって、研究デザインおよび PICO が CQ に一致する論文を選択し、対象論文として確定した。意見の不一致がある場合は、3 人目のメンバーによるレビューと議論を行った。除外した論文については理由を記録し、PRISMA フローダイアグラムでプロセスをまとめた。

※ ネットワークメタ解析に関する CQ の扱い

ネットワークメタ解析に関する CQ は、頻度論的解析方法を用いて解析を行った。エビデンスの質の評価は GRADE working group のアプローチに基づいて行った。本ガイドラインにおいては、CQ 2-11、CQ6-3 でネットワークメタ解析を行った。

Step 4: データ抽出、バイアスリスクの評価

データ抽出は、2 名以上の独立したレビューにより行い、標準化したデータ抽出フォームを用いた。研究ごとに、以下の項目について詳細情報を抽出した(研究デザイン、研究対象者参入基準、除外基準、症例数、介入とコントロール、アウトカム)。文献に十分な情報の記載がない場合は、その旨を記載し、著者への問い合わせは行わなかった。

Step 4: バイアスリスクの評価

2 名の独立したレビューにより、研究・アウトカムごとに、以下のドメインにおけるバイアスリスクを評価し

た。

(RCT)

RoB 2.0 の手法に則って評価した[2]。

- ランダム割り付けで生じるバイアス bias arising from the randomization process
- 意図した介入からの逸脱によるバイアス bias due to deviations from intended interventions
- 欠測アウトカムによるバイアス bias due to missing outcome data
- アウトカム報告によるバイアス bias in measurement of the outcome
- 報告する結果の選択によるバイアス bias in selection of the reported result

(観察研究)

ROBINS-I の手法に則って評価した[3]。

- 交絡因子によるバイアス
- 患者選択に伴うバイアス
- 群分けに関するバイアス
- 想定する介入とのずれ
- データ欠損に伴うバイアス
- アウトカム評価段階のバイアス
- 選択アウトカム報告

バイアスの潜在性の程度は、high、low、または unclear と判断し、データ抽出フォームと同様に標準化されたバイアスリスクの表に記載した。バイアスの潜在性が high と判断された場合は、その判断理由も記載した。バイアスリスクが high であることを理由に、文献を除外することはしなかった。バイアスリスクの評価は、CQごとに「risk of bias グラフ」および「risk of bias サマリー」に視覚的にまとめた。

Step 5:メタ解析とエビデンスの確実性の評価

採用する文献の定性的評価、定量的評価を行った。定量的評価では、可能な場合、Revman 5 を用いて、メタ解析を行った。それらを総括してエビデンスの確実性の評価を各領域班で作成した。

系統的レビューを行った CQ について、各班が担当 CQ におけるエビデンスの確実性 (A~D) を GRADE アプローチに基づいて行い、Risk of bias/非直接性/非一貫性/不精確さ/出版バイアス/その他の 6 項目をもとに評価した。Risk of bias 評価は RoB 2.0 に基づいて行った。エビデンスプロファイルの作成には、ガイドライン作成ツール (GDT) オンラインソフト (<http://gdt.guidelinedevelopment.org>) を使用した。RCT のエビデンスの確実性は高から開始し、深刻な問題があった場合ダウングレードされた。観察研究は低から開始し RCT と同様に評価されたが、方法論的に妥当で、効果が大きく、用量反応勾配があるか、またはもっともらしい交絡により提示されている効果が減少していることが確認されているならば、「中」または「高」の確実性のエビデンスにグレードを上げることを考慮された。

3. 推奨の策定

推奨の決定に先立ち、委員とワーキンググループは協働して Evidence to Decision テーブル(EtD テーブル)を作成した。その後、エビデンスの確実性・効果のバランス・価値観、そして、コストや資源の利用の 4 要因を加味し、委員会での合議により推奨を策定した。

「介入による望ましい効果が望ましくない効果を上回るか下回るか」について、どの程度確信できるかを示すために、一部の CQ において完全コンテキスト化アプローチを用いた。その具体的な方法については、

GRADE ワーキンググループが公表したコンセプトに準拠した[4]。詳細な方法は割愛するが、系統的レビューに採用された重大なアウトカムのうち、介入の総合的な効果を判断するのに必要であると思われたアウトカムを選択し、それらのアウトカムの効果推定値をその相対的価値も考慮して統合して算出される、総合的な効果推定値 (net effect estimate) を用いる方法である。そのコンセプトを実際に採用したガイドラインはまだ報告がなかったため、アカデミック班で事前に J-SSCG 2020 のデータを用いて実際に検証をした[5]。

GRADE システムで示される推奨の強さは、推奨・弱い推奨・弱い非推奨・非推奨の 4 つのカテゴリーに分類される(本編の表 2 を参照)。

PICO に沿った文献の網羅的検索によって、該当する RCT がなかった CQ は FRQ に変更された。また、いくつかの RCT によりメタ解析を行なったものの十分なエビデンスが得られなかった CQ では、委員とワーキンググループは協働して EtD テーブルを作成し、FRQ に変更するかどうか委員会で議論した。その結果、エビデンスの確実性が低く今後の研究成果によって推奨が反対方向に変わる懸念があると判断された場合は FRQ として扱った。作成された EtD テーブルは付録資料として掲載した。

Step 1: 投票プロセス

GRADE に基づく CQ における合意形成は、修正 Delphi 法を用いた。委員は各推奨案に対し 1 から 9 点 (1 点:同意できない、9 点:同意できる) を独立して投票し、7 点未満の場合はその理由も記載した。投票は Google フォームを用いてオンラインかつ匿名で実施した。投票に関与しない独立メンバーが集計し、得点の中央値 (Median)、パーセンタイル範囲 (Interpercentile Range、IPR)、対称を補正したパーセンタイル範囲 (Interpercentile Range Adjusted for Symmetry、IPRAS)、見解不一致指数 (Disagreement Index、DI) を算出した。該当する CQ に COI の懸念がある委員は投票を棄権した。

Step 2: パネル会議

集計結果に基づき、以下のようにパネル会議で合意形成を行った。

1. Median < 7 かつ DI \geq 0.3 の場合

担当委員により EtD と推奨文の改訂を行い、委員会で議論後に再投票を実施した。合意基準を満たすまで、修正と投票を繰り返した。

2. Median \geq 7 あるいは DI < 0.3 の場合

A) 投票時に委員より推奨に対する重大な意見が提出された場合

委員会で議論を行い、合意形成を図った。該当する CQ を担当した委員から修正依頼があった場合は、EtD と推奨の改訂を行い、再投票を実施した。

B) 推奨に対する重大な意見が提出されなかった場合

投票結果を確認し、合意形成とした。CQ の微修正に関して担当委員から申請があった場合は、委員会で議論を経て合意形成とした。

※BQ については GRADE 基準による CQ と同じ閾値を用い、GPS については Median \geq 8、DI < 0.20 を合意形成の閾値とした。

参考文献

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315:801–10.

2. <https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool?authuser=0https://methods.cochrane.org/robins-i>

3. <https://methods.cochrane.org/robins-i>

4. Alper BS, Oettgen P, Kunnamo I, Iorio A, Ansari MT, Murad MH, et al. Defining certainty of net benefit: a GRADE concept paper. *BMJ Open*. 2019;9:e027445.

5. Terayama T, Okano H, Kawakami S, Kano K, Sakuraya M, Aoki Y, et al. Comparison of certainty of evidence between the net benefit approach and the traditional GRADE method using the data of Japanese Clinical Practice Guidelines for Management of Sepsis and Septic Shock 2020. *J Intensive Care*. 2023;11:33.