

第 4 章

システムティックレビュー

Minds

4章 Key Words

| |
|---|
| ◎ AGREE II (The Appraisal of Guidelines for Research & Evaluation II) |
| ◎ AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews 2) |
| ◎ CINAHL (Cumulative Index to Nursing and Allied Health Literature) |
| ◎ Embase |
| ◎ GRADE (The Grading of Recommendations Assessment, Development and Evaluation) |
| ◎ GRADEpro GDT |
| ◎ I^2 統計値 |
| ◎ Institute of Medicine (IOM) |
| ◎ IOM の SR の基準 (ガイドライン評価における) |
| ◎ IOM の基準 (ガイドライン評価における) |
| ◎ ITT (intention-to-treat) 解析 |
| ◎ JMEDPlus |
| ◎ MEDLINE |
| ◎ MeSH (Medical Subject Headings) |
| ◎ NNT (number needed to treat) |
| ◎ PICO |
| ◎ PRESS (Peer Review of Electronic Search Strategies) |
| ◎ PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) |
| ◎ PROSPERO (International Prospective Register of Systematic Reviews) |
| ◎ PsycINFO |
| ◎ PubMed |
| ◎ Q 統計値 |
| ◎ Rayyan |
| ◎ RevMan (Review Manager) |
| ◎ ROBINS-I (Risk of Bias In Non-randomized Studies for Interventions) |
| ◎ SoF 表 (summary of findings table) |
| ◎ The Cochrane Library |
| ◎ アウトカムの重要性 |
| ◎ アブストラクト |
| ◎ 医中誌 Web |

| |
|---|
| ◎ 益と害 (benefit and harm) |
| ◎ 益と害のバランス |
| ◎ エビデンスの確実性 |
| ◎ エビデンスの確実性 (強さ) |
| ◎ エビデンスの確実性 (質) |
| ◎ エビデンス総体 (body of evidence) |
| ◎ エビデンス総体の総括 (overall evidence) |
| ◎ オッズ比 (odds ratio : OR) |
| ◎ 介入研究 (interventional study) |
| ◎ 観察研究 (observational study) |
| ◎ 感度分析 (sensitivity analysis) |
| ◎ クリニカルクエスチョン (clinical question : CQ) |
| ◎ グレイリテラチャー (grey literature) |
| ◎ 傾向スコア分析 |
| ◎ 検出バイアス (detection bias) |
| ◎ 交絡因子 (confounding factor) |
| ◎ 固定効果モデル (メタアナリシスにおける) |
| ◎ コホート研究 (cohort study) |
| ◎ 今後の研究 (future research) |
| ◎ コンシールメント (concealment) |
| ◎ サンプルサイズ (sample size) |
| ◎ システマティックレビュー (systematic review : SR, 系統的レビューと同義) |
| ◎ システマティックレビューチーム (systematic review team : SR チーム) |
| ◎ システマティックレビューの評価 |
| システマティックレビューレポート (サマリーレポートともいう) |
| ◎ 実行バイアス (performance bias) |
| ◎ 重要臨床課題 (key clinical issues) |
| ◎ 出版 (報告) バイアス (publication bias, reporting bias) |
| ◎ 症例減少バイアス (attrition bias) |
| ◎ 症例集積 (case series) |
| ◎ 症例対照研究 (case-control study) |
| ◎ 症例報告 (case reports) |
| ◎ 診断精度 (diagnostic test accuracy : DTA) 研究 |
| ◎ 診療ガイドライン (clinical practice guideline : CPG) |

| |
|---|
| ◎ 診療ガイドラインのためのシステマティックレビュー |
| ◎ 診療ガイドライン作成グループ (guideline development group : GDG, 診療ガイドライン作成パネルともいう) |
| ◎ 推奨 |
| ◎ スコープ (scope) |
| ◎ 選択バイアス (selection bias) |
| ◎ 早期試験中止バイアス |
| ◎ 多基準決断分析 (multi-criteria decision analysis : MCDA) |
| ◎ ネットワークメタアナリシス (network meta-analysis : NMA) |
| ◎ バイアスリスク (risk of bias) |
| ◎ ハザード比 (hazard ratio : HR) |
| ◎ 非一貫性 (inconsistency) |
| ◎ 非直接性 (indirectness) |
| ◎ 評価シート (evidence profile) |
| ◎ 標準化平均値差 (standardized mean difference : SMD) |
| ◎ 非ランダム化比較試験 (non-randomized controlled trial : non-RCT) |
| ◎ ファンネルプロット (funnel plot) |
| ◎ フォレストプロット (forest plot) |
| ◎ 不精確性 (imprecision) |
| ◎ 米国医療研究・品質庁 (Agency for Healthcare Research and Quality : AHRQ) |
| ◎ メタアナリシス (meta-analysis, メタ解析と同義) |
| ◎ 盲検化 (blinding) |
| ◎ 用量 - 反応勾配 (dose-dependent gradient) |
| ◎ ランダム化 (randomization) |
| ◎ ランダム化比較試験 (randomized controlled trial : RCT) |
| ◎ ランダム効果モデル (メタアナリシスにおける) |
| ◎ 利益相反 (conflict of interest : COI) |
| ◎ リスク差 (risk difference : RD) |
| ◎ リスク比 (risk ratio : RR) |

◎：巻末「重要用語集」の収録語

テンプレートの頭文字の略称

PR : preparation 【第2章 準備】 SC : scope 【第3章 スコープ】

SR : systematic review 【第4章 システマティックレビュー】 EC : economic 【第5章 医療経済評価】

RC : recommendation 【第6章 推奨】 FN : finalization 【第7章 公開に向けた最終調整】

PO : post 【第8章 診療ガイドライン公開後の取り組み】

4.0 概要

診療ガイドライン（clinical practice guideline：CPG）は推奨を作成するために、存在する全てのエビデンスを科学的に活用することが望まれる。そのため、システマティックレビュー（systematic review：SR、系統的レビューと同義）が行われる。システマティックレビューでは包括的に研究の検索を行い、一定の基準で選定された研究論文を一定の基準で評価し、アウトカムごとにエビデンス総体（body of evidence）のエビデンスの強さを決定することが望ましい。1つのクリニカルクエスション（clinical question：CQ）には複数のアウトカムが設定されるので、アウトカムごとに選定された研究集合ができ、それらがさらに研究デザインごとにまとめられることになる。

なお、エビデンスに対する評価について、質、レベル、限界、強さ、確信度、確実性などの言葉が同じ意味で用いられてきている。本マニュアルでは、広義のエビデンスに対する評価を「エビデンスの確実性」とし、推奨を考慮しない段階で行われるシステマティックレビューでは効果指標の確実性に対する確信という意味で「エビデンスの確実性（質）」、もしくは、「エビデンスの質」という言葉を用い、推奨作成の段階では推奨を支持する強さに対する確信という意味で「エビデンスの確実性（強さ）」、もしくは、「（エビデンス総体の）エビデンスの強さ」という言葉を用いる。この分類は、GRADE アプローチにおいて、システマティックレビューのエビデンスの確実性と推奨作成におけるエビデンスの確実性では異なる定義を用いることが明確化されている（Hultcrantz M 2017, Turner RM 2009）ことにも対応している。

診療ガイドライン作成のためのシステマティックレビューは、科学的作業であるとともに医療へ直接影響を与えるものであることを認識し、科学性、包括性、透明性、不偏性の確保に努めることが望ましい。

システマティックレビューにより、それぞれのアウトカムに対する効果の大きさとその不確実性が、効果指標の統合値と信頼区間として示される。それらが、益と害（benefit and harm）の評価の科学的根拠となる。

包括性を確保するため、データベース検索の際の検索式を工夫するとともに、必要に応じて検索外での文献追加も実行することが望ましい。透明性を高めるため、用いたデータベース、検索方法、検索式を記録する。可能であれば、論文タイトル、アブストラクトによる一次スクリーニング、全文に基づく二次スクリーニングをあらかじめ定めた採用基準、除外基準に基づいて行い、採用論文のリストを作成するとともに、不採用論文のリストも作成することが望ましい。

また、可能であれば、アウトカムごとに作成された研究集合の個別論文、およびそれらを統合したエビデンス総体のエビデンス評価を一定の基準で行う。定性的な評価と定量的な評価の両者に基づき最終的にエビデンスの強さを評価し、最後にアウトカム横断的にエビデンス総体の総括（overall evidence）の評価を行い、推奨作成を行うことが望ましい。

エビデンス評価の基準については、国際的に共通の認識があるので（AHRQ 2012, Berkman ND 2014, Higgins JP 2011a, Higgins JPT 2011b, Deeks JJ web, Schünemann H 2013）、それらに準じて行うことが望ましい。

4.0.1 システマティックレビューの定義

システマティックレビューはさまざまに定義付けが行われているのが現状である。「システマティックレビュー」という言葉が「メタアナリシス」と同義語で用いられていることもある。しかし、実際にはメタアナリシス（meta-analysis, メタ解析と同義）を伴わないシステマティックレビューも存在し、システマティックレビューではないメタアナリシスもあり得るので、完全な同義語ではない。

Institute of Medicine (IOM) はシステマティックレビューを以下のように定義している。「特定の問題に絞って、類似した、しかし別々の研究の知見を見つけ出し、選択し、評価し、まとめるために、明確で計画された科学的方法を用いる科学研究。別々の研究からの結果の定量的統合（メタアナリシス）を含むことも含まないこともある」（IOM 2011a: 21, IOM 2011b: 23）。診療ガイドライン作成のためのメタアナリシスとは、「研究標本から特定の集団に対する推定と適用を可能とすることを企図して、類似した研究の結果を定量的に結合するために統計学的方法を用いるシステマティックレビュー」である（IOM 2011a: 21）。いずれの定義も方法論的側面に着目した定義となっている。また、統合（synthesis）とは、「綿密な情報の照合、結合およびシステマティックレビューの結果のまとめのこと」と定義されている（IOM 2011a: 173）。

方法論的な面からの定義として IOM の定義に賛同するものであるが、ゴールに基づく基準^{*1}として考えると、システマティックレビューと呼べるための条件としては、1. 参照した研究に漏れがない、2. 採択された研究に偏りがなく、3. 中立の立場で一定の基準に基づき各研究を評価：①アウトカムに及ぼす効果の大きさ、②効果の確実性、4. 結論に評価の結果が反映されている、を提案する（**図 4-1** 参照）。

実際の作業の面から定義すると、システマティックレビューとは、「CQ に対して、研究を網羅的に調査し、研究デザインごとに同質の研究をまとめ、バイアスを評価しながら分析・統合を行うこと」である。

Cochrane review をはじめとし、数多くのシステマティックレビューあるいはメタアナリシスが発表されている。現在、システマティックレビューを計画時点で登録するウェブサイト PROSPERO^{*2}が運用されているので、診療ガイドラインのための CQ に基づくシステマティックレビューも、論文として発表を計画する場合には、個々に PROSPERO に登録することを検討すべきである。

システマティックレビューの事前登録をするときには、システマティックレビューのプロトコルが必要である。登録をしないときでも、スコープ（scope）に記載されたシステマティックレビューの方法に加えて、プロトコルを作成することが望ましい。

*1 システマティックレビューの要件を目標としてどれだけ達成しているかという観点から設定する基準である。いわゆるチェックリストによって研究を評価する際に、各項目が合致するかどうかを見ていく評価基準は分類評価基準（criteria-based criteria）と呼ばれる。これに対して、理想的な状態を目標として想定して、それをどれだけ達成できたかを見ていく評価基準は目標評価基準（goal-based criteria）と呼ぶ。

*2 PROSPERO (International Prospective Register of Systematic Reviews) <http://www.crd.york.ac.uk/PROSPERO>

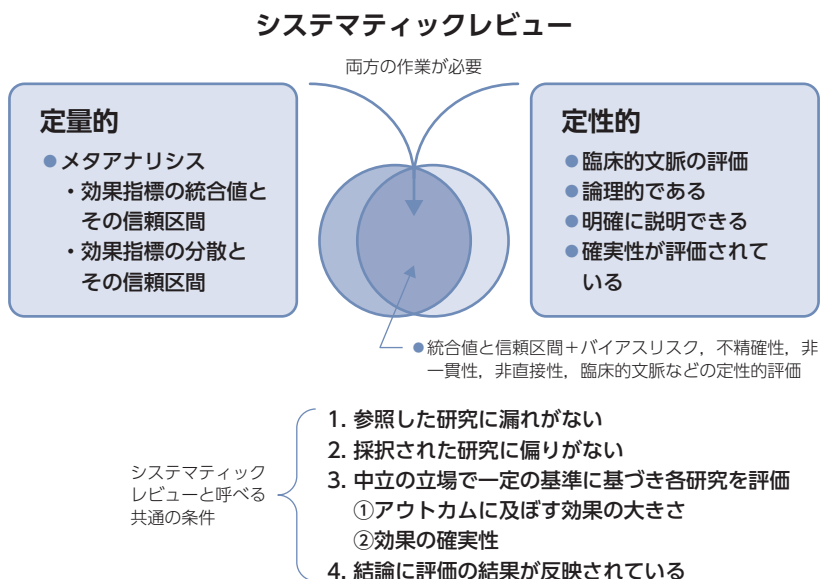


図 4-1 定量的システマティックレビューと定性的システマティックレビュー

4.0.2 定性的システマティックレビューおよび定量的システマティックレビュー

(1) 定性的システマティックレビュー

採択された研究や除外された研究の数、対象の特性と数、比較と介入、バイアスリスク (risk of bias) の評価などを記述し、深い理解を与えるために定性的にまとめたものをいう。定性的な評価の結果を効果指標の値の調整に用いる手法が開発されているが一般的ではないので、定性的システマティックレビューの結果は最終的にエビデンスの強さの判定に反映させる。

参考として、IOM の定性的統合の主要な目的と本マニュアルの考える定性的システマティックレビューにおける実際の作業について表 4-1 に示す。IOM は採択された研究や除外された研究の数、サンプルサイズ (sample size)、比較された介入、バイアスリスクの評価などの記述に留まらず、介入がどのように役立つか、誰のためか、どのような状況で用いられるかなどについてより深い理解を与えるものとして定性的システマティックレビューを位置付けている*³。

* 3 定量的統合あるいは定性的統合という場合は、必ずしもシステマティックという概念を含んでいない。得られる全てのエビデンスを統合する場合は、定量的システマティックレビューと定量的統合、また、定性的システマティックレビューと定性的統合は同じものになる。システマティックレビューで行われるメタアナリシスと単なるメタアナリシスも同様の関係にあり、メタアナリシスは一部の研究だけを定量的に統合する場合にも適用される。

表 4-1 定性的システマティックレビューの主要な目的

| 目的 | 該当する実際の作業 |
|--|--|
| 1. 臨床的展望の中で利用者に方向付けを与える。 | 臨床課題あるいは CQ を臨床的文脈において位置付ける。 |
| 2. 研究中に参加者に実際に起きたことを記述する。 | 害、脱落の頻度と理由も評価する。 |
| 3. エビデンス総体を論評する。 | バイアスリスク、非直接性 (indirectness)、不精確性 (imprecision)、非一貫性 (inconsistency)、などを評価する。 |
| 4. なぜ結果が異なるかを説明する個々の研究デザインおよび実行の相違点を明らかにする。 | 非一貫性、研究間の異質性を評価する。 |
| 5. 個々の研究のデザインおよび実行が実世界の臨床的状況との関連にどのように影響するかを述べる。 | 非直接性を評価する。 |
| 6. エビデンスの全般的まとめとセッティングと患者集団に基づくサブグループ解析の結果を取り込む。 | 非直接性を評価する。 |
| 7. 研究が不十分あるいは結果が異なる患者集団に対する注意を促す。 | 非直接性を評価する。 |
| 8. メタアナリシスの結果の堅牢さを解釈し評価する。 | 感度分析 (sensitivity analysis) の結果、出版 (報告) バイアス (publication bias, reporting bias) を評価する。 |
| 9. システマティックレビューの結果が医療業界や社会の通念と異なるかを記述する。 | 専門医、一般医、医療者、患者・市民代表の評価を取り込む。 |

(2) 定量的システマティックレビュー (メタアナリシス)

定量的システマティックレビューでも、いわゆるメタアナリシスの前にバイアスの評価など定性的な評価を行い、定量的に統合できるか、異質性 (heterogeneity) を検討する必要がある。メタアナリシスとは、効果指標の値が統計学的に統合され、統合値と信頼区間を計算し、定量的統合を行うことである。

メタアナリシスの原理を簡単に解説すると、いくつかの集団のサンプル数と平均値が分かっているが、個々のサンプルの測定値が分からない場合に、全体の平均値を求めることに類似している。分散 σ^2 の正規分布に従う母集団からの n 個のランダムサンプルの分散は、中心極限定理より σ^2/n である。したがって、このランダムサンプルの分散の逆数 n/σ^2 はサンプル数 n に比例することになり、分散の逆数で各集団の平均値を重み付けして算出した平均値は全体の平均値となる、というのが原理である。

(3) 定性的システマティックレビューと定量的システマティックレビューの関係

システマティックレビューとメタアナリシスを同じものと見なすと、複数の類似した研究の効果指標の値を統計学的手法で統合することがシステマティックレビューであるという考えが生じてしまう。個別の研究のサンプルサイズはさまざまで研究の実行の厳密さもさまざまなので、バイアスリスクなどによる研究の質は研究により異なり、得られた結果の確実性はさまざまである。各研究の質をなんらかのチェックリストで評価し、その結果をスコア化し効果指標の値の重

み付けに用いるメタアナリシスが提案されたが、その後その手法の問題点が指摘され、そのような研究の質による効果指標の調整は、現在では用いないことが推奨されている (Jüni P 1999)。一方、バイアスリスクの各項目を定量的に評価し、それを効果指標の値の調整に用いる方法が提案され (Turner RM 2009)、正しく機能することが示されているが、高度なスキルが要求されるため、一般化するに至っていない。

このような状況で、診療ガイドライン作成のためのシステマティックレビューでは、効果指標の値をメタアナリシスの手法で統合し、エビデンス総体の定性的評価の結果によって、エビデンスの確実性の評価をする方法がとられている。例えば、生存をアウトカムとしてランダム化比較試験 (randomized controlled trial : RCT) のメタアナリシスによりハザード比 (hazard ratio : HR) 0.5, 95%信頼区間 0.41 ~ 0.61 という結果が得られていて、効果が大きいと見なされる場合でも、各研究のバイアスリスクが高く、研究間の非一貫性も高く、非直接性も高いと判定されれば、エビデンス総体の強さは A「強い」ではなく、B「中」あるいは C「弱い」、場合によっては D「非常に弱い」と判定する。これは、もし定性的評価を効果指標の統合値と信頼区間に反映させることが可能であれば、真のハザード比が 0.6 で信頼区間が 0.38 ~ 0.94 であると推定するという事に相当する。すなわち、効果がより小さめで、確実性が低いと見なすことになる。

複数の研究をエビデンス総体としてまとめる場合に、研究デザイン、対象、介入、対照、アウトカム (PICO) や効果指標の類似性が十分な場合には、定量的システマティックレビューが可能である。しかしそうでない場合には、個々の研究が効果指標の値を提示していても、定量的システマティックレビューができない場合もある。また、定量的な効果指標の値が得られない研究の場合もあり得る。これら全ての研究をエビデンス総体としてまとめ、確実性を評価し、最終的にエビデンスの確実性 (強さ) を評価することが望ましい。

定性的システマティックレビューは定量的システマティックレビューと並行して行われるものと、定量的システマティックレビューすなわちメタアナリシスが適用できない複数の研究に適用されるものがある。前者では、バイアスリスク、不精確性、非一貫性、非直接性、出版 (報告) バイアス、臨床的文脈などの評価を行うことが定性的システマティックレビューに該当する。一方、後者では対象となる研究で、研究デザインが異なったり、PICO の違いがあったりするため、バイアスリスク (Guyatt GH 2011c)、不精確性 (Guyatt GH 2011e)、非一貫性 (Guyatt GH 2011f)、非直接性 (Guyatt GH 2011g)、出版 (報告) バイアス (Guyatt GH 2011d) の評価に加え、臨床的文脈を明確にし、論理的で明確な説明をし、確実性を評価することが重要となる。

例えば、RCT が 1 件、症例対照研究 (case-control study) が 1 件しかないような場合でも、それぞれの効果指標の評価と、定性的な評価は可能であり、それらをまとめて結論を導き出すことはシステマティックレビューと呼ぶことができる。

4.0.3 診療ガイドラインのためのシステマティックレビューと通常のシステマティックレビュー

Cochrane reviewをはじめとし、数多くのシステマティックレビューあるいはメタアナリシスが発表されている。これらのシステマティックレビューの多くは診療ガイドライン作成を目的として行われたものではない。これら通常のシステマティックレビューと診療ガイドラインのためのシステマティックレビューにはいくつかの相違点がある（表4-2参照）。通常のシステマティックレビューはエビデンスの確実性（質）を明らかにすることを目的としており、診療ガイドライン作成のためのシステマティックレビューは推奨と関連したエビデンスの確実性（強さ）を明らかにすることを目的としている。

表4-2 診療ガイドラインのためのシステマティックレビューと通常のシステマティックレビュー

| 項目 | 診療ガイドライン作成のためのシステマティックレビュー | 通常のシステマティックレビュー |
|-------------|---|---|
| 目的 | 益と害の判定のために効果の大きさとエビデンスの強さを明らかにする。 | 効果の大きさ（効果指標の統合値と信頼区間）と確実性を明らかにする。 |
| 研究の対象者 | 属性が限定的でより特異的なものから広範囲まで。 | 同一疾患で属性が広範囲。 |
| アウトカム | 複数の益のアウトカムと害のアウトカムに対する効果が評価される。 | 1～3つ程度の益の主要アウトカムに対する効果の評価が中心になることが多い。 |
| 害のアウトカム | <ul style="list-style-type: none"> ・ 益のアウトカムと同じように重要性が評価される。 ・ 必要な場合観察研究も対象とされ、包括的に解析される。 | <ul style="list-style-type: none"> ・ 副次的に扱われる可能性がある。 ・ RCTの統合では頻度の低い害は解析されない可能性が高い。 |
| 統合対象研究のデザイン | 異なる研究デザインを含む。 | 1つの研究デザインだけを取り扱っても成立する。 |
| 非直接性 | PICOの各項目についてCQとの類似性が評価される。 | 統合される研究間の類似性が評価される。 |
| 研究の質の評価 | 同じ研究でもアウトカムごとにあるいは非直接性により異なる可能性がある。 | 研究ごとに一意的に決めることが可能。 |
| 文献の管理 | 1つの文献が複数のアウトカム、あるいはCQに関連し、複雑になりやすい。 | 比較的単純。 |
| 文献の選定 | アウトカムごとに行うと同じ文献を何回もチェックしなければならない。 | アウトカムが1つであれば比較的単純。 |
| 研究の集合 | 1つの介入でアウトカムごとに複数の集合が必要。 | 1つの介入で1つの集合がある。 |

これらの相違点で特に注意すべき点は、診療ガイドラインのためのシステマティックレビューでは益と害のアウトカムの両方が重要視されること、同じ研究でもアウトカムが異なると質の評価が異なること、同じ研究が異なるアウトカムに対して適用される可能性があるため文献管理が複雑になることである。

4.0.4 既存のシステマティックレビューの利用

診療ガイドライン作成を目的としたシステマティックレビューは、エビデンス総体の強さを評価し推奨を決定することを目的としている。特に注意すべき点は、明確なCQに対する網羅的文献検索 (McDonagh M 2013) と文献採用基準が求められること、益と害のアウトカムの両方が重要視されること、同じ研究でもアウトカムが異なると質の評価が異なること、同じ研究が異なるアウトカムに対して適用される可能性があるため文献管理を一元的に行う工夫が必要となることである。

すでに同じCQに対応するシステマティックレビューが出版されている場合には、それを利用することが可能な場合もある。図4-2にその際の方針を示す (White CM 2009)。すでに同じCQに対応する診療ガイドラインが出版されている場合には、診療ガイドラインについてはAGREE II (The Appraisal of Guidelines for Research & Evaluation II), AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews 2) (Shea BJ 2017) [AMSTAR (Shea BJ 2007) の改訂版] などのツールを用いて評価し、質の高いものを選択し、その中に含まれるシステマティックレビューの利用については、上記の既存のシステマティックレビューの場合と同じ方針を採用する。

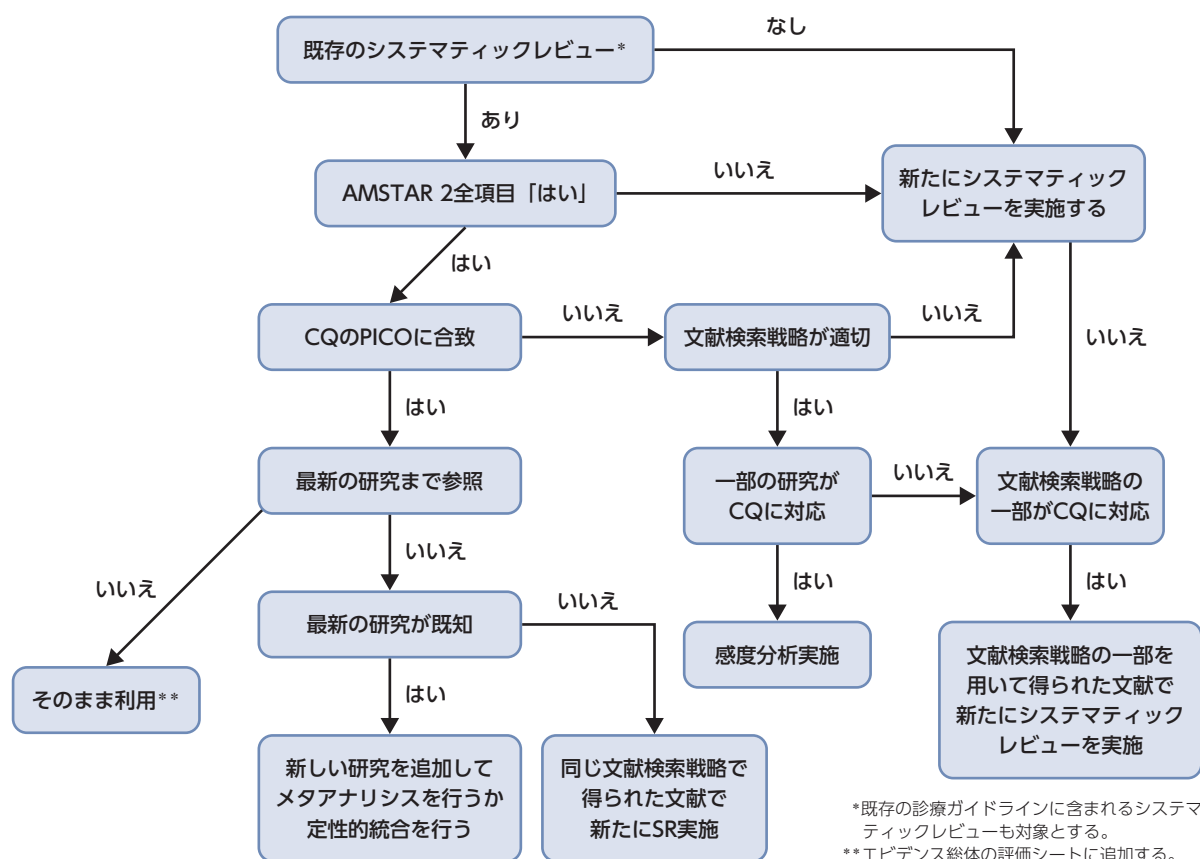


図 4-2 既存のシステマティックレビューを利用する場合の方針

出典：White CM, et al. (2009) Using existing systematic reviews to replace de novo processes in conducting Comparative Effectiveness Reviews. In: Agency for Healthcare Research and Quality (AHRQ). Methods Guide for Comparative Effectiveness Reviews. By courtesy of Dr. White https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/methods-guidance-de-novo-processes_methods.pdf (参照 2021/02/24) より作成

- ① 統合結果をそのまま利用する。
- ② 論文で採択されている元の研究データを再評価し、元の研究から CQ に適合するものを選択してシステマティックレビューを行う。
- ③ 同じ文献検索戦略を用いて得られた文献で新たにシステマティックレビューを行う。
- ④ 新しい研究を追加してメタアナリシスを行うか定性的な統合を行う。
- ⑤ 文献検索戦略の一部を用いて得られた文献で新たにシステマティックレビューを行う。

どの方針を用いるかは、そのシステマティックレビューそのものの質 [これは、以下の表 4-3 などの評価で行うものであり、エビデンスの確実性（質）や研究の質（バイアスリスク）ではない] が高いか低いか、また、そのシステマティックレビュー出版後の新たな研究がどれくらいあるか、それらの結果を含めると結果が変わる可能性があるかなどを評価して決める。

表 4-3 AMSTAR 2：ヘルスケア介入のランダム化または非ランダム化研究あるいは両方を含むシステマティックレビューのための厳密な評価ツール

項目 1 レビューの際、リサーチ・クエスチョンと組入れ基準に PICO の要素を含めたか？

| | | |
|---|---|---|
| Yes の場合： <input type="checkbox"/> 参加者 <input type="checkbox"/> 介入 <input type="checkbox"/> コントロール群 <input type="checkbox"/> アウトカム | 選択項目（推奨） <input type="checkbox"/> フォロアップ期間 | <input type="checkbox"/> Yes <input type="checkbox"/> No |
|---|---|---|

項目 2 レビューの報告に、レビュー実施前にその方法を決定したことを明示し、重大な計画変更があれば正当な理由を示したか？

| | | |
|--|--|---|
| 部分的な Yes の場合： 著者は以下のすべてを含むプロトコルや基準があることを明示する： <input type="checkbox"/> レビュー・クエスチョン <input type="checkbox"/> 検索戦略 <input type="checkbox"/> 組入れ／除外基準 <input type="checkbox"/> バイアスリスク評価 | Yes： 部分的な Yes の場合に加え、プロトコルが事前に登録され、また特定されるべき： <input type="checkbox"/> 必要に応じて、メタ分析／併合の計画、および <input type="checkbox"/> 異質性の原因を探る計画 <input type="checkbox"/> プロトコルからの何らかの逸脱に対する妥当性 | <input type="checkbox"/> Yes <input type="checkbox"/> Partial Yes <input type="checkbox"/> No |
|--|--|---|

項目 3 レビュー著者は、レビューに組入れる際の研究デザインの選択基準について説明したか？

| | |
|---|---|
| Yes の場合、そのレビューは以下の 1 つを満たすべきである： <input type="checkbox"/> RCT のみを組入れたことの説明 <input type="checkbox"/> または、NRSI *のみを組入れたことの説明 <input type="checkbox"/> または、RCT と NRSI の両方を組入れたことの説明 | <input type="checkbox"/> Yes <input type="checkbox"/> No |
|---|---|

*NRSI：非ランダム化ヘルスケア介入研究 (Non-Randomised Studies of healthcare Interventions)

項目4 レビュー著者は、網羅的な文献検索の手法を用いたか？

| | | |
|---|---|---|
| <p>部分的な Yes の場合（以下のすべて）：</p> <ul style="list-style-type: none"> <input type="checkbox"/> 少なくとも2つのデータベースで検索した（リサーチ・クエスチョンに関して） <input type="checkbox"/> キーワードと検索戦略を示した <input type="checkbox"/> 出版形態の制限の正当性を示した（例：言語） | <p>Yes の場合、次も必要（以下のすべて）：</p> <ul style="list-style-type: none"> <input type="checkbox"/> 検索した参考文献リスト／組入れた研究の文献一覧 <input type="checkbox"/> 検索した試験／研究登録データベース <input type="checkbox"/> 検索で組入れた／相談したその分野の専門家 <input type="checkbox"/> 関連する場合は、検索した灰色文献 <input type="checkbox"/> レビュー完成の24ヵ月以内に検索を実施した | <ul style="list-style-type: none"> <input type="checkbox"/> Yes <input type="checkbox"/> Partial Yes <input type="checkbox"/> No |
|---|---|---|

項目5 レビュー著者は、複数人で研究選択を行ったか？

| | |
|--|---|
| <p>Yes の場合、以下のいずれか1つを含む：</p> <ul style="list-style-type: none"> <input type="checkbox"/> 少なくとも2人のレビュアーが、基準を満たす研究の選択について独立して合意し、組入れる研究について合意形成した <input type="checkbox"/> または、2人のレビュアーが基準を満たす研究のサンプルを選択し、良好な一致度（少なくとも80%）に至り、残りは1人のレビュアーによって選択された | <ul style="list-style-type: none"> <input type="checkbox"/> Yes <input type="checkbox"/> No |
|--|---|

項目6 レビュー著者は、複数人でデータ抽出を行ったか？

| | |
|---|---|
| <p>Yes の場合、以下のいずれか1つを含む：</p> <ul style="list-style-type: none"> <input type="checkbox"/> 少なくとも2人のレビュアーが、組入れられた研究から抽出するデータについて合意形成した <input type="checkbox"/> または、2人のレビュアーが基準を満たす研究のサンプルからデータを抽出し、良好な一致度（少なくとも80%）に至り、残りは1人のレビュアーによって抽出された | <ul style="list-style-type: none"> <input type="checkbox"/> Yes <input type="checkbox"/> No |
|---|---|

項目7 レビュー著者は、除外した研究のリストを提示し、その除外が正当である理由を示したか？

| | | |
|---|--|---|
| <p>部分的な Yes の場合：</p> <ul style="list-style-type: none"> <input type="checkbox"/> 除外されたがフルペーパーとして十分に読み込まれたすべての潜在的に関連する研究のリストを提供した | <p>Yes の場合、次も含まなければならない：</p> <ul style="list-style-type: none"> <input type="checkbox"/> 潜在的に関連する研究についてレビューから除外したことの正当性を示した | <ul style="list-style-type: none"> <input type="checkbox"/> Yes <input type="checkbox"/> Partial Yes <input type="checkbox"/> No |
|---|--|---|

項目8 レビュー著者は、組入れた研究を十分詳細に記述したか？

| | | |
|---|---|---|
| <p>部分的な Yes の場合（以下のすべて）：</p> <ul style="list-style-type: none"> <input type="checkbox"/> 参加者の記載 <input type="checkbox"/> 介入の記載 <input type="checkbox"/> 比較対照の記載 <input type="checkbox"/> アウトカムの記載 <input type="checkbox"/> 研究デザインの記載 | <p>Yes の場合、次も必要（以下のすべて）：</p> <ul style="list-style-type: none"> <input type="checkbox"/> 詳細な参加者の記載 <input type="checkbox"/> 詳細な介入の記載（関連する投与量も含む） <input type="checkbox"/> 詳細な比較対照の記載（関連する投与量も含む） <input type="checkbox"/> 研究の場所の記載 <input type="checkbox"/> フォローアップ期間の記載 | <ul style="list-style-type: none"> <input type="checkbox"/> Yes <input type="checkbox"/> Partial Yes <input type="checkbox"/> No |
|---|---|---|

項目9 レビュー著者は、レビューに組入れた各研究のバイアスリスク評価に際し、満足のいく手法を用いたか？

| | | |
|---|--|--|
| <p>RCT</p> <p>部分的な Yes の場合、以下の状態のバイアスリスクが評価されていなければならない：</p> <ul style="list-style-type: none"> <input type="checkbox"/> 隠蔽されていない割振り、および <input type="checkbox"/> アウトカムを評価する際の、参加者と評価者のブラインディングの欠如（すべての死亡原因のような客観的なアウトカムの場合には不要） | <p>Yes の場合、以下の状態のバイアスリスクも評価されていなければならない：</p> <ul style="list-style-type: none"> <input type="checkbox"/> 本当にランダムではない割振り手順、および <input type="checkbox"/> ある特定のアウトカムについて多重測定や多重解析から報告された結果の選択 | <ul style="list-style-type: none"> <input type="checkbox"/> Yes <input type="checkbox"/> Partial Yes <input type="checkbox"/> No <input type="checkbox"/> Includes only NRSI |
|---|--|--|

| | | |
|--|---|--|
| <p>NRSI</p> <p>部分的な Yes の場合、以下の状態のバイアスリスクが評価されていない： <input type="checkbox"/> 交絡因子、および <input type="checkbox"/> 選択バイアス</p> | <p>Yes の場合、以下のバイアスリスクも評価されていない： <input type="checkbox"/> 曝露とアウトカムを確認するために使用された方法、および <input type="checkbox"/> ある特定のアウトカムについて多重測定や多重解析から報告された結果の選択</p> | <input type="checkbox"/> Yes <input type="checkbox"/> Partial Yes <input type="checkbox"/> No <input type="checkbox"/> Includes only NRSI |
|--|---|--|

項目 10 レビュー著者は、レビューに組入れた研究に対する資金源を報告したか？

| | |
|--|---|
| <p>Yes の場合： <input type="checkbox"/> レビューに組入れられた個々の研究の資金源が報告されなければならない(注: レビューがこの情報を探したが、著者によって報告されていない場合は、そのように報告する)</p> | <input type="checkbox"/> Yes <input type="checkbox"/> No |
|--|---|

項目 11 メタ分析が行われた場合、レビュー著者は、結果の統計的併合において適正な方法を用いたか？

| | |
|---|--|
| <p>RCT</p> <p>Yes の場合： <input type="checkbox"/> 著者はメタ分析でデータ併合を正しく行った <input type="checkbox"/> そして、研究結果の併合に適した重みづけ法を用い、存在する場合には異質性を調整した <input type="checkbox"/> そして、異質性の原因について調べた</p> | <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> No meta-analysis conducted |
| <p>NRSI</p> <p>Yes の場合： <input type="checkbox"/> 著者はメタ分析でデータ併合を正しく行った <input type="checkbox"/> そして、研究結果の併合に適した重みづけ法を用い、存在する場合には異質性を調整した <input type="checkbox"/> そして、生データを併合するのではなく、交絡因子を調整した NRSI からの効果量推定値を統計的に併合した、または調整された効果量推定値が利用できなかった場合に生データを併合することの正当性を示した <input type="checkbox"/> そして、RCT と NRSI の両方をレビューに組入れた場合、それぞれに分けて効果量を報告した</p> | <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> No meta-analysis conducted |

項目 12 メタ分析が行われた場合、レビュー著者は、各研究のバイアスリスクがメタ分析や他の統合結果に与える影響について評価したか？

| | |
|---|--|
| <p>Yes の場合： <input type="checkbox"/> 低バイアスリスクの RCT のみを用いた <input type="checkbox"/> または、プールされた効果量がさまざまなバイアスリスクでの RCT または NRSI に基づく場合、要約効果量におけるバイアスリスクの影響を調べる分析を行った</p> | <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> No meta-analysis conducted |
|---|--|

項目 13 レビュー著者は、レビュー結果の解釈や考察をする際にバイアスリスクを考慮したか？

| | |
|---|---|
| <p>Yes の場合： <input type="checkbox"/> 低バイアスリスクの RCT のみを用いた <input type="checkbox"/> または、中から高バイアスリスクの RCT や NRSI をレビューに組入れた場合、結果におけるバイアスリスクの影響を考察した</p> | <input type="checkbox"/> Yes <input type="checkbox"/> No |
|---|---|

項目 14 レビュー著者は、レビュー結果にみられるいかなる異質性に対しても、満足のいく説明や考察を行ったか？

| | |
|---|---|
| <p>Yes の場合： <input type="checkbox"/> 結果において有意な異質性はなかった <input type="checkbox"/> または、異質性があった場合、結果における異質性の原因を調べ、レビュー結果に与える異質性の影響を考察した</p> | <input type="checkbox"/> Yes <input type="checkbox"/> No |
|---|---|

項目 15 量的な統合がなされた場合、レビュー著者は出版バイアス（小規模研究バイアス）について十分な検討、およびレビュー結果に与える影響の可能性を考察したか？

| | |
|---|---|
| Yes の場合： | <input type="checkbox"/> Yes |
| <input type="checkbox"/> 出版バイアスに対するグラフィック的または統計的検定を行い、出版バイアスの影響の程度を考察した | <input type="checkbox"/> No |
| | <input type="checkbox"/> No meta-analysis conducted |

項目 16 レビュー著者は、レビュー実施に際して受領した財源を含め、考えられる利益相反について報告したか？

| | |
|--|------------------------------|
| Yes の場合： | <input type="checkbox"/> Yes |
| <input type="checkbox"/> レビュー著者は利益相反がないことを報告した、または | <input type="checkbox"/> No |
| <input type="checkbox"/> レビュー著者は資金源を記述し、潜在的な利益相反にどのように対処したかを記載した | |

出典：上岡洋晴 他. (2018) 「AMSTAR 2：ヘルスケア介入のランダム化または非ランダム化研究あるいは両方を含むシステマティック・レビューのための厳密な評価ツール」の紹介と解. 薬理と治療 46:111-113.

(1) ペア比較の研究のシステマティックレビュー

2つの介入を直接比較する RCT の結果を統合するシステマティックレビューであり、用いられるメタアナリシスの手法については、米国医療研究・品質庁（Agency for Healthcare Research and Quality：AHRQ）の見解によれば（Morton SC 2018）、ランダム効果モデルを用い、効果指標としては、リスク比（risk ratio：RR）あるいはリスク差（risk difference：RD）を用い、イベント率が1%以下の場合には Peto 法を用いること、ゼロイベントが1アームの場合には + 0.5 などで補正する（この場合は Mantel-Haenszel 法は用いない）ことが推奨されている。

RD は医療者に理解されやすく決断分析に適しているが、RR に比べるとばらつきが大きくなりやすく、イベント率が低い場合やベースラインリスクのばらつきが大きい場合は検出力が低くなることに注意する必要がある。GRADE アプローチでも、非一貫性の評価の際には RR、生存分析の場合はハザード比によるフォレストプロット（forest plot）を参照することが勧められている。

ペア比較のメタアナリシスについては、**4.4.4 項 (2)** を参照されたい。

(2) ネットワークメタアナリシス

ネットワークメタアナリシス（network meta-analysis：NMA）または multi-treatment meta-analysis あるいは mixed treatment comparisons（MTC）（Dias S 2010）は、3つ以上の介入・要因曝露の効果を直接的な比較（head-to-head comparison）だけでなく、間接的な比較も含めてペアでの比較や順位付けを可能にするメタアナリシスである（Dias S 2018）。前提としては、介入の効果を变化させる因子が研究間で差がないこと、すなわち移行性（transitivity）が成立することが必要である。移行性が十分成立する場合、直接比較の情報に間接比較の情報加味され、より堅牢な結果が得られると考えられている。

例えば、**図 4-3** の例 1 の場合のように治療法が A、B、C と 3 種類あり、A と B を直接比較した RCT と、A と C を直接比較した RCT があり、B と C の比較も行って、いずれの治療効果が最も高いといえるかを知りたいような場合に用いることができる。この例の場合、効果推定値

のパラメータ δ について A と B の差 δ^{AB} , A と C の差 δ^{AC} , B と C の差 δ^{BC} の関係は, $\delta^{AB} + \delta^{BC} = \delta^{AC}$ が成立し, 移項すると $\delta^{BC} = \delta^{AC} - \delta^{AB}$ となり, 治療 B と C を間接比較した場合の効果推定値が得られる。分散 V については, $V^{BC} = V^{AB} + V^{AC}$ が成立する。

図 4-3 の例 2 の場合には, 治療 B と C は直接比較と間接比較の両方が可能となる。この場合は, 直接比較 δ_d^{BC} と間接比較 δ_i^{BC} をそれぞれの分散の逆数で重み付けした平均値を算出する。直接比較の分散 V_d^{BC} , 間接比較の分散 V_i^{BC} となり, 前者は B と C の直接比較の研究データから, 後者は上記のように A と B, A と C の直接比較の研究データから算出された分散から算出される。

ネットワークメタアナリシスでは比較される介入の数が多くなるに従い, 比較されるペアの数が増加する。 S 個の介入を比較する場合, 最大 $S(S-1)/2$ の比較が必要となる。

また, ネットワークメタアナリシスでは, 効果の大きさに順位をつけ, それぞれの治療法が最善, すなわち 1 位になる確率を算出することができる。

ネットワークメタアナリシスに関する執筆ガイダンスとして, PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) 声明の拡張版が発表されており (Hutton B 2015), 可能な限り順守することが望ましい。

R では特定の解析を目的とした「パッケージ」と呼ばれるプログラムが世界中の開発者により提供されている。頻度論派の手法による解析用 R パッケージとして, netmeta パッケージ (Rücker G 2019), ベイジアン手法による R のパッケージとして, gemtc パッケージ (van Valkenhoef G 2016) がある。gemtc は JAGS (just another gibbs sampler) (Plummer M 2017) を, rjags (Plummer M 2019) を介して動作させるプログラムである。ベイジアン手法によるネットワークメタアナリシスの多くは OpenBUGS を用いて行われており, そのためのコードが数多く発表されている (Dias S web, Salanti G web) (R については 4.4.4 項 (2) を参照)。

ネットワークメタアナリシスの解析結果としてネットワークグラフ, フォレストプロット, 順位グラフ (rankogram), ノード分割モデルプロット (node-splitting model plot) (van Valkenhoef G 2016), SUCRA プロット (surface under the cumulative ranking curves plot) (Salanti G 2011) などが得られる。

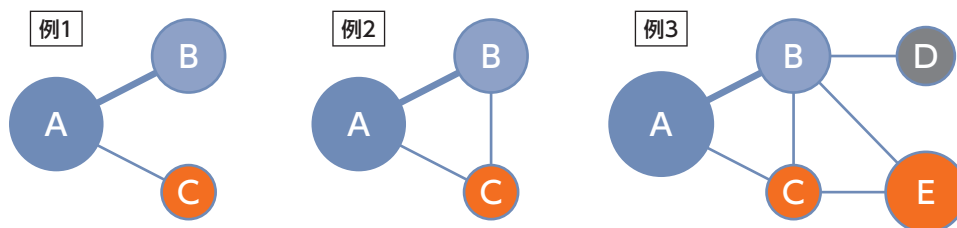


図 4-3 ネットワークグラフ。例 1 では治療 B と治療 C の間接比較が可能になる。例 2 では治療 B と治療 C の直接比較と間接比較も可能になる。例 3 はさらに複雑なモデルである。

ネットワークグラフは各ノードが1つの介入を示し、ノードの大きさは症例数を反映する。ノードを結ぶ直線は直接比較が可能なペアを示し、直線の太さは研究数を反映する。ネットワークグラフが示す概念はネットワークジオメトリ (network geometry) であり、それは介入比較の構造を示す。比較される介入、それぞれの介入にランダム割り付けられた症例数、特定の介入や比較が選好されたあるいは忌避されたかを見て取ることができる。ネットワークはさまざまな形をとり、結合線が少ない場合は、直接比較が少ないことを表し、そのようなネットワークメタアナリシスは信頼性が低いと見なされる。ネットワークジオメトリの定性的な記述が必要とされ、ネットワークグラフに付随されるべきであるとされている (Mills EJ 2013)。確立したネットワークジオメトリの見方はまだないので、それぞれの解釈については可能な限り付随する説明を用意することが望ましい。

ペア比較の結果で、直接比較の場合は各研究間の異質性の評価は通常メタアナリシスと同様で、統計学的異質性 (statistical heterogeneity) は Q 統計値、 I^2 統計値などで検定されるが、その限界もよく知られているところであり (Gagnier JJ 2012)、臨床的異質性と研究デザインや研究実施の異質性も評価する必要がある。Mills EJら (2013) は後者を概念的異質性 (conceptual heterogeneity) という言葉で呼んでいる。

ネットワークメタアナリシスの場合は、さらに直接比較の結果と間接比較の結果を比べて不一致がないかを検討し、エビデンスの確実性を評価する必要がある。これは、間接比較も可能なクローズドループ内で行われる。不一致がある場合、インコヒレンス (incoherence) と呼ばれる。具体的には、直接比較と間接比較の結果の統合値と 95% 信頼区間を一覧表にまとめるなどして、一つひとつ検討することになる。gemtc では (van Valkenhoef G 2016)、ノード分割モデルの方法で自動的に直接比較と間接比較の統合値と信頼区間のグラフを作成でき、このような作業に用いることができる。

ネットワークメタアナリシスエビデンスに関する従来の GRADE ワーキンググループ (GRADE Working Group) の提言では、次のように述べられている。

- ① ネットワーク内のそれぞれのペア比較のエビデンスの確実性は別々に評価する。
- ② 直接比較と間接比較を両方考慮する (Puhan MA 2014)。
- ③ ネットワークメタアナリシスの効果推定値のレーティングには不精確性を考慮する必要はない。
- ④ 直接比較のエビデンスが間接比較のエビデンスよりも確実性が高く、ネットワーク推定値への直接比較の貢献が少なくとも間接比較と同じ程度であれば、間接比較のエビデンス評価は必要ない。
- ⑤ ペア比較のレベルでのインコヒレンスの評価においては、ネットワーク全体のインコヒレンスの統計学的検定を信頼すべきでない。
- ⑥ 直接比較と間接比較のエビデンスにインコヒレンスがある場合、それぞれの推定値のエビデンスの確実性により、どちらのエビデンスを信じるべきかを決める。

したがって、それぞれのペアの比較について、通常のメタアナリシスの場合と同様の評価ドメイン・項目を評価し、エビデンスの確実性を評価する必要がある。

また、順位については、1つの介入が1位になる確率が圧倒的に高いことはまれであり、不確実性を伴っていることを前提に解釈すべきである。

また、GRADE ワーキンググループは図 4-4 のようなインコヒレンスの評価の手順について提言している (Brignardello-Petersen R 2019)。優越的な推定値がなく、バイアスあるいは非直接性が認められる場合はレートダウンし、非移行性のみ場合はレートダウンしない。バイアスと非直接性の両方が影響している場合は、2レベルのレートダウンを考慮する。

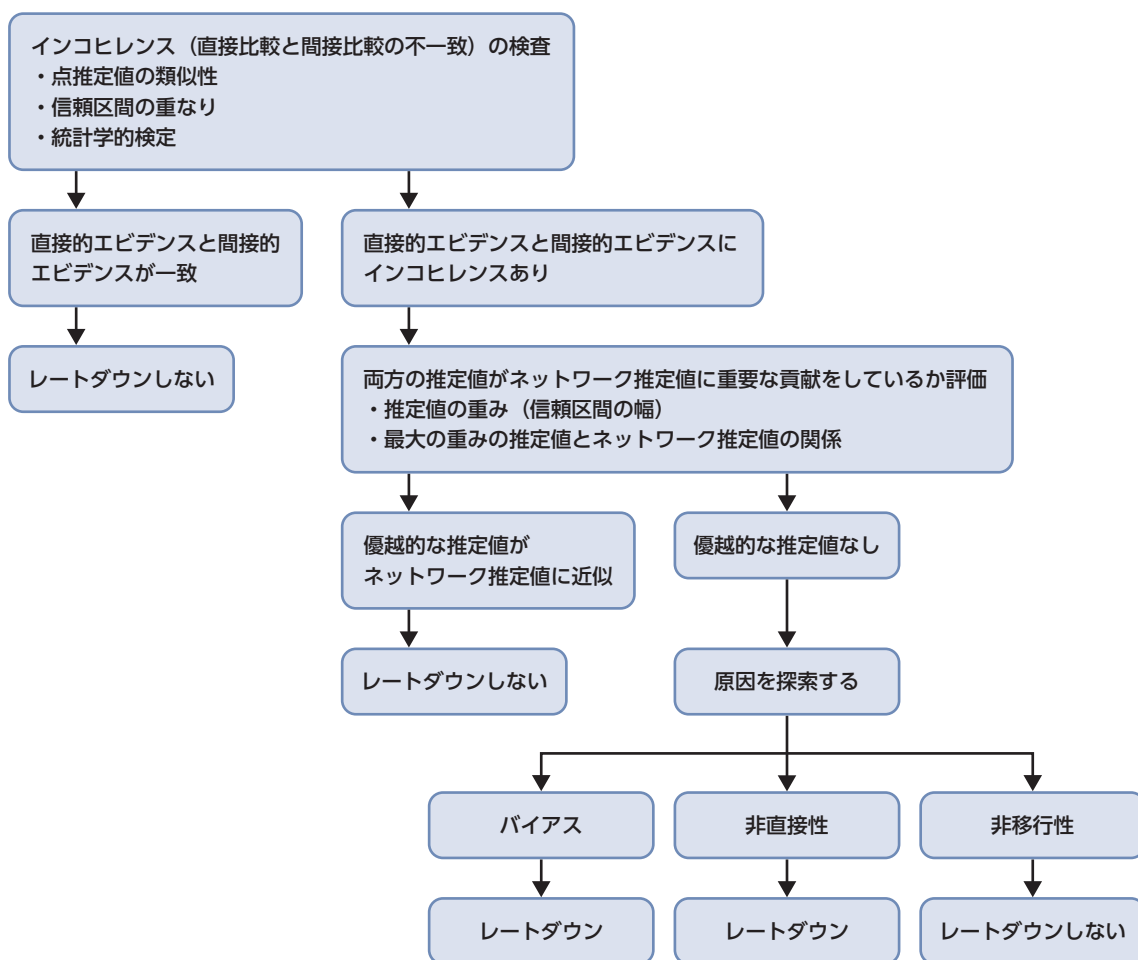


図 4-4 ネットワークメタアナリシスのインコヒレンス評価手順 (GRADE)

出典：Reprinted from Journal of Clinical Epidemiology, Vol 108, Romina Brignardello-Petersen et al., GRADE approach to rate the certainty from a network meta-analysis: addressing incoherence, p.77-p.85, Copyright (2018), with permission from Elsevier より作成

また GRADE ワーキンググループからは、ネットワークメタアナリシスに対応した結果のまとめ表、SoF 表 (summary of findings table) のフォーマットが発表されている (Yepes-Nuñez JJ 2019)。

(3) 診断精度研究のシステマティックレビュー

診断法の臨床的アウトカムへの直接的な効果を測定することは困難なことが多く、癌のスクリーニングなどにほとんど限定されている。そのような研究、特にRCTが行われている場合には、通常の治療的介入の場合と同じ手法でシステマティックレビューが可能となる。CQはPICO形式で設定可能であり、研究の評価、エビデンス総体の評価も通常のRCTの場合と同じように行うことができる。益と害の評価についても、同じ手法が適用できる。

診断法に関する研究の多くは診断能を明らかにする目的で行われており、診断精度(diagnostic test accuracy : DTA) 研究と呼ばれている。その目的は、診断法の感度・特異度を明らかにすることであり、何らかの参照基準で診断標的の有無を決定した対象者において診断法を実施し、診断標的を有する者における陽性率(感度)ならびに診断標的を有さない者における陰性率(特異度)を明らかにする。

診断精度研究のシステマティックレビューではCQの設定も異なってくる。WHOの診療ガイドライン作成に関するハンドブック(WHO Handbook for Guideline Development)では、PIRT [参加者(Participants), インデックス検査(Index test(s)), 参照基準(Reference standard), 標的状态(Target condition)] 形式でCQを設定することを述べている(WHO 2019)。AHRQは臨床的なアウトカムの場合と同じように、PICOTS (Patient population, Intervention, Comparator, Outcomes, Timing, Setting) 形式でCQ設定を用いることを提唱しており、Outcomesに患者健康アウトカムまたは中間アウトカムを設定し、中間アウトカムに(診断法の)技術的仕様、精度、決断や治療へのインパクトを設定することを提案している(Chang SM 2012)。

Mindsでは特別寄稿(森實敏夫 2015)で述べているように、CQの形式はPICOを適用することを提言している。診断精度研究の場合は、Oに診断標的を設定し、診断標的を検出する方法として参照基準を併記すること、インデックス診断法が1つの場合は、Iにそれを設定し、Cは空欄とすること、直接的な比較で2つの診断法を比較する場合は、Cに直接比較される診断法を設定することを提言している。また、参照基準が不完全な場合や、研究により異なる参照基準が用いられている場合の方法についても述べている。

メタアナリシスの手法としては、検査閾値を考慮したReitsmaまたはRutter-Gatsonisの二変量モデル、あるいはそれに該当する方法を用い、HSROC (hierarchical summary receiver operating characteristic) 解析を用いることが一般的である。頻度論派の手法によるRパッケージであるmadaパッケージ、あるいはベイジアン手法によるOpenBUGSを用いた解析のコードから、感度・特異度のフォレストプロットおよび統合値とSROC曲線が得られる。

診断精度研究のネットワークメタアナリシスについてはPRISMA声明が発表されており(McInnes MDF 2018)、今後活用が一般化する可能性がある。

感度・特異度の統合値から、一定の事前確率に対して、真陽性(true positive : TP)、偽陰性(false negative : FN)、偽陽性(false positive : FP)、真陰性(true negative : TN)の割合を算出することができる。これら4つに分類された対象者は、その後、一定の治療的介入を受け、その治療による益と害を受けることになる。その診断法を実施した場合を実施しない場合と比較し

て、あるいは別の診断法を実施した場合と比較して、益と害に差があるかどうかを検討する作業が必要になるが、診断法実施後の転帰について実際のデータが不十分な場合が多く、この作業は困難なことが多いと考えられる。また、事前確率が異なると TP, FN, FP, TN の割合も異なるので、益と害の大きさとバランスも異なってくる。さらに、診断法実施後の臨床的アウトカムの生起には時間が関与し、マルコフモデル (Markov model)、マイクロシミュレーション (microsimulation) (Knudsen AB 2016) などの決断モデルが必要になると考えられる。

したがって、診断精度研究のシステマティックレビューは感度・特異度の統合値をメタアナリシスによって明らかにすることが中心にならざるを得ない。

GRADE ワーキンググループは、診断精度とその後治療を含めた転帰のリンクエビデンスの確実性を評価する必要性およびその際のエビデンス総体の総括の確実性としては一番低いレベルを採用することを提案しているが (Schünemann HJ 2019b)、益と害の評価には今後、決断モデルの適用が必要であると述べている。

また、診断法実施後の治療の益とコスト (害と負担) を効用値として設定可能な場合は、決定木を作成して (Sox HC 2013)、無治療診断閾値 (診断法を実施せず治療もしない場合の効用値が、診断法を実施して陽性なら治療する場合の効用値を上回る疾患確率)、治療診断閾値 (診断法を実施せず治療する場合の効用値が、診断法を実施して陽性なら治療する場合の効用値を上回る疾患確率) を明らかにすることが可能である。無治療診断閾値と治療診断閾値の間の疾患確率は、診断法を実施して陽性なら治療をする場合の効用値が高い範囲となるので、これらの情報を提供することは、有用であろう。

診断精度研究のバイアスリスクは QUADAS-2 (Whiting PF 2011) に基づき、以下のドメインおよび評価項目 (括弧で示したもの) を評価する。

- ・選択バイアス (臨床に即したランダム選択)
- ・インデックス検査 (盲検化)
- ・参照基準 (盲検化, 不完全な参照基準)
- ・症例減少バイアス (不完全な検査実施)
- ・フローとタイミング (同時期に実施)
- ・その他 (データ欠損など)

個別研究について、これらは「高 (-2)」、「中/疑い (-1)」、「低 (0)」の3段階で評価する。非直接性は CQ の対象、インデックス検査、参照基準、アウトカムと各研究のこれら項目との一致性を評価する。また、上記のように3段階評価を行う。

エビデンス総体については介入研究 (interventional study) の場合と同様であり、「高」からスタートして各ドメインに問題があればダウングレードの方向で検討する。

4.0.5 エビデンス総体の評価

(1) エビデンス総体

ある臨床上的問題（CQ）に対して収集し得た全ての研究報告を、アウトカムごと、研究デザインごとに評価し、その結果をまとめたものをエビデンス総体と呼ぶ。

このアウトカムは、スコープ作成時に、1つのCQに対して、「患者にとって重要、重大な各アウトカム」として列挙されたものである。

(2) エビデンス総体の総括

エビデンス総体をアウトカム横断的に統合した全体を、「エビデンス総体の総括」と呼ぶ。診療ガイドライン作成過程において、CQに対する推奨を提示する場合（第6章参照）に、エビデンス総体の総括としてのエビデンスの確実性（強さ）を1つだけ決定する（Guyatt G 2013a）。

論理的には、推奨への影響が最も大きいアウトカムに対するエビデンス総体のエビデンスの確実性（強さ）を、エビデンス総体の総括の確実性（強さ）として設定すべきである。例えば、益と害のバランスが、益に傾き、益の大きさが十分であると判断した場合、その益のアウトカムに対するエビデンス総体のエビデンスの確実性（強さ）を、総括のエビデンスの確実性（強さ）として設定することが妥当である。

なお、GRADEアプローチでは、推奨にとって重大なアウトカムに対するエビデンス総体の確実性のうち一番低いものに決定する。

(3) 研究デザインの評価とエビデンスレベル

このように、エビデンスに関する評価は、エビデンス総体に関する評価、および、後述するように複数のアウトカムに対するエビデンス総体を見渡して、エビデンス総体の全体に関する評価を行うことで、推奨に関するエビデンスの強さについての評価とする。

「研究報告単位の研究デザインのみに基づいたエビデンスレベル付け」は、行わなくてもよい。

(4) エビデンス総体の評価

診療ガイドラインの作成においてはシステマティックレビューを実施し、エビデンスの評価、その強さと限界について明確に提示することが望ましい。なお、エビデンスの強さは研究デザインのみで決定されるのではなく、報告内容を詳細に評価、検討し、場合によってはさらにメタアナリシスを行って評価することが望ましい。

エビデンス評価はアウトカムごとに行い、エビデンス総体の強さと限界を明確に記載することが望ましい（表4-4参照）。

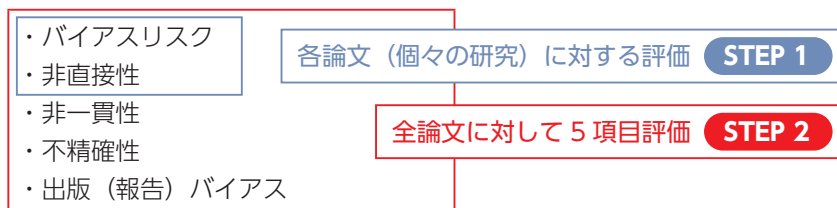
表 4-4 エビデンスの確実性の評価と判定方法の具体的作業

(1) CQ に対し、アウトカムを列挙し、重大、重要なアウトカムを採用

(2) 採用された一つひとつのアウトカムに対し以下の作業を行う

- 1) 研究デザインによる分類 (RCT 群, 観察研究群としてまとめる)
- 2) RCT に対するエビデンスの確実性の評価

エビデンスの確実性の評価を下げる項目



3) 観察研究に対するエビデンスの確実性の評価

エビデンスの確実性の評価を上げる項目

- ・介入による大きな効果 (large effect)
- ・用量 - 反応勾配 (dose-dependent gradient)
- ・可能性のある交絡因子による効果の減弱 (plausible confounder)

注) 各項目の重み付けは CQ によって異なる

出典 : Schünemann H, et al. (2013) GRADE Handbook. Permission by the editors. <http://gdt.guidelinedevelopment.org/app/handbook/handbook.html> (参照 2021/02/24) より作成

4.0.6 エビデンスの強さの考え方

診療ガイドラインにおけるエビデンスの強さは、その治療効果推定値に対する確信が、ある特定の推奨を支持する上で、どの程度適切かを反映している (Balshem H 2011)。その決定の際に、複数の益と害のアウトカムに対するそれぞれのエビデンスの確実性を臨床的文脈全体の中で相互の判断への影響を考慮すべきであることが、GRADE ワーキンググループから提言されており (Hultcrantz M 2017)、システマティックレビューの段階のエビデンス総体の確実性、すなわちエビデンスの確実性 (質) と、推奨作成段階のエビデンス総体の確実性、すなわちエビデンスの確実性 (強さ) が異なる場合があり得る。システマティックレビュー段階のエビデンス総体の確実性はエビデンスの質の評価として確定しておき、推奨作成の時点で評価が変わる場合は、その説明を記録することが望ましい。

エビデンスの確実性は、個々の研究報告の研究デザインのみに基づいて決定されるのではなく、① 個々の報告内容を詳細に評価検討した上で、② 場合によってはさらに統合解析を行って、アウトカムごとにまとめられた研究報告の「総体」としてのエビデンスが決定される。

システマティックレビュー段階でのエビデンス総体のエビデンスの確実性の基準については、診療ガイドライン作成グループ (guideline development group : GDG, 診療ガイドライン作成パネルともいう) において採用する基準をあらかじめ決めておくことが望ましい。表 4-5 に本マニュアルで推奨するエビデンスの確実性の例を示す。これは GRADE アプローチと同じ分類である。

表 4-5 エビデンス総体のエビデンスの確実性 (質)

| | |
|-------------|---------------------|
| A (高) : | 効果の推定値に強く確信がある |
| B (中) : | 効果の推定値に中程度の確信がある |
| C (低) : | 効果の推定値に対する確信は限定的である |
| D (とても低い) : | 効果の推定値がほとんど確信できない |

○ ステップ

- (1) エビデンスの収集
- (2) スクリーニング
- (3) エビデンスの評価：個々の報告評価 **STEP 1**
- (4) エビデンス総体の評価 **STEP 2**
- (5) システマティックレビューレポートの作成

4.1 ステップ1：エビデンスの収集

4.1.1 文献検索データベース

包括的な検索を行うため、複数のデータベース検索を行う（表4-6参照）。MEDLINE/PubMed^{*4}、The Cochrane Library^{*5}、医中誌 Web の3つのデータベースは必ず検索する。医学領域の Embase^{*6} や JMEDPlus、社会・心理学的研究の PsycINFO[®]、看護研究の CINAHL（Cumulative Index to Nursing and Allied Health Literature）など、診療ガイドラインのテーマにより重要なデータベースも併せて検索する。必要と考えられる場合は、PROSPERO、臨床試験登録、会議録、医薬品副作用情報なども検索することを検討する。

データベースによって検索式の記述法は異なるが、多くの場合互いに変換することも可能である。

表 4-6 主な文献検索データベース

| 重要度 | データベース名 | URL | 閲覧要件 |
|-----------|---|--|---------------------|
| 必須 | PubMed | https://pubmed.ncbi.nlm.nih.gov/ | なし |
| | MEDLINE | https://ovidsp.dc2.ovid.com/ovid-a/ovidweb.cgi など | 要契約 |
| | The Cochrane Library | https://www.cochranelibrary.com/ | 検索はフリーだが検索結果の保存は要契約 |
| | 医中誌 Web | https://search.jamas.or.jp/ | 要契約 |
| 重要 | Embase | https://www.embase.com/ | 要契約 |
| | PsycINFO [®] | https://www.apa.org/pubs/databases/psycinfo | 要契約 |
| | CINAHL | http://web.b.ebscohost.com/ | 要契約 |
| | JMEDPlus | https://jdream3.com/ | 要契約 |
| | Open Gray | http://www.opengrey.eu/ | なし |
| | Clinical Trials.gov | https://www.clinicaltrials.gov/ | なし |
| | World Health Organization (WHO) | https://apps.who.int/trialsearch/ | なし |
| | UMIN 臨床試験登録システム | https://www.umin.ac.jp/ctr/index-j.htm | なし |
| 国立保健医療科学院 | https://rctportal.niph.go.jp/ | なし | |

* 4 MEDLINE には PubMed 以外に Ovid などの商用データベースもある。

* 5 The Cochrane Library には CDSR (The Cochrane Database of Systematic Review)、CENTRAL (Cochrane Central Register of Controlled Trials) などのデータベースがある。例えば CENTRAL だけを検索した場合は CENTRAL と、指定していない場合は The Cochrane Library と記載する。

* 6 RCT に関しては、The Cochrane Library の CENTRAL から検索可能だがタイムラグがある。チーム内で利用環境が整っていれば検索対象とする。

4.1.2 文献検索戦略

システマティックレビューチーム (systematic review team : SR チーム) は, CQ ごとにキーワード, シソーラス (MeSH など) を組み合わせた検索式を2名 (1名は図書館員など医学文献検索専門家などであることが望ましい) が確認して最適な検索式を作成し, データベースごとに検索式, 検索期間, 検索日を記録する。検索者が1名しか得られない場合は検索式の査読を受けることも検討する。システマティックレビューの検索式査読のためのガイドライン PRESS (Peer Review of Electronic Search Strategies) が開発されており (McGowan J 2016), チェックリストが日本語に訳されている (河合富士美 2017)。

適切な検索を行うためには, 診療ガイドライン作成グループと検索担当者との間での重要臨床課題 (key clinical issues) と CQ に対する理解の共有が重要となるため, 両者の共同作業が必要である。いつ共同作業が必要かは, 次のようにいくつかの時点に分けて考えるとよい。

- ① 検索語句 (キーワード) を選出する時点
- ② 検索式を作成する時点
- ③ 検索実施後, 検索結果が得られた時点 (例えば, 文献数が多すぎたり少なすぎたりした場合)
- ④ 文献選定後に漏れやデータベースの追加などで検索式の見直しが必要になった時点

それぞれの時点で独立作業と, 照合とピアレビュー・修正の共同作業を行うことがあり得る。また, 複数の検索式が用いられ, 文献検索と文献の選定作業が同時進行するような場合には, すでに選定作業の終了した文献集合を除くなど, 選定作業に重複が生じないように工夫が必要である。

通常のデータベース検索では不十分と考えられる場合は, 引用文献, 会議録などのグレイリテラチャー (grey literature), 教科書の参照など, 検索外の追加を検討する。その場合も手段と件数を記録しておく。ただし, 検索外の追加については委員会での十分な検討が必要である。

全ての検索式, 文献選択の経過は**テンプレート【SR-1】**に記録し, さらに**テンプレート【SR-2】**に示すように, CQ ごとに PRISMA 声明のフローダイアグラムを改変した文献検索フローチャートに記載する。

(1) 検索の進め方

システマティックレビュー, 臨床研究は, PICO の P, I, C を用いて系統的に検索するが, 害と不利益を含んだ幅広いアウトカムを拾うため, 検索式に O は含めない。プロトコルにあらかじめ記載されている場合にはさらに研究デザインでの絞り込みを行う。**テンプレート【SR-1】**を用いた益のアウトカムに関する検索例を以下に示す。

【PubMed 検索例 1 〈益の検索〉】

| |
|---|
| タイトル： Stroke and rt-PA |
| CQ： 発症6時間以内の80歳未満の急性脳塞栓患者（patients）に対して、発症3時間以降のrt-PA投与（intervention）は推奨されるか？ |
| データベース： PubMed |
| 日付： 2019/08/25 |
| 検索者： NK / FK |

| # | 検索式 | 文献数 |
|----|--|------------|
| 1 | "acute ischemic stroke" [TIAB] | 12946 |
| 2 | "acute ischaemic stroke" [TIAB] | 1874 |
| 3 | "Brain Ischemia" [Mesh] | 103113 |
| 4 | "Stroke" [Mesh] | 124695 |
| 5 | #1 OR #2 OR #3 OR #4 | 175665 |
| 6 | "Tissue Plasminogen Activator" [Mesh] | 17046 |
| 7 | rt-pa [TIAB] | 2554 |
| 8 | alteplase [TW] | 1559 |
| 9 | #6 OR #7 OR #8 | 19457 |
| 10 | #5 AND #9 | 5914 |
| 11 | <u>#10 AND (Meta-Analysis [PT] OR systematic [SB])</u> | <u>144</u> |
| 12 | <u>#10 AND "Randomized Controlled Trial" [PT]</u> | <u>328</u> |

【PubMed 検索例 2 〈益の検索〉】

| # | 検索式 | 文献数 |
|----|---|-------------|
| 1 | "Cerebrovascular Disorders" [Mesh:NoExp] | 45882 |
| 2 | "Brain Ischemia" [Mesh] | 103113 |
| 3 | "Carotid Artery Diseases" [Mesh:NoExp] OR "Carotid Artery Thrombosis" [Mesh] | 24050 |
| 4 | "Stroke" [Mesh:NoExp] OR "Brain Infarction" [Mesh] | 124325 |
| 5 | "Hypoxia-Ischemia, Brain" [Mesh] | 5319 |
| 6 | "Cerebral Arterial Diseases" [Mesh:NoExp] OR "Intracranial Arterial Diseases" [Mesh:NoExp] | 2921 |
| 7 | "Intracranial Embolism and Thrombosis" [Mesh:NoExp] | 8680 |
| 8 | stroke* [TW] OR apoplex* [TW] OR "cerebral vascular" [TW] OR cerebrovasc* [TW] OR cva [TW] OR "transient ischemic attack*" [TW] OR "transient ischaemic attack*" [TW] OR tia [TW] | 393220 |
| 9 | brain [TW] OR cerebr* [TW] OR cerebell* [TW] OR vertebrobasil* [TW] OR hemispher* [TW] OR intracran* [TW] OR intracerebral [TW] OR infratentorial [TW] OR supratentorial [TW] OR "middle cerebral" [TW] OR mca* [TW] OR "anterior circulation" [TW] | 1851347 |
| 10 | ischemi* [TW] OR ischaemi* [TW] OR infarct* [TW] OR thrombos* [TW] OR emboli* [TW] OR occlus* [TW] OR hypoxi* [TW] | 1172113 |
| 11 | #9 AND #10 | 238225 |
| 12 | #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #11 | 531896 |
| 13 | "Tissue Plasminogen Activator" [Mesh] | 18149 |
| 14 | "PLAT protein, human" [Supplementary Concept] | 149 |
| 15 | plasminogen [TW] OR plasmin [TW] OR tPA [TW] OR t-PA [TW] OR rtPA [TW] OR rt-PA [TW] | 78965 |
| 16 | alteplase [TW] | 1890 |
| 17 | #13 OR #14 OR #15 OR #16 | 79409 |
| 18 | #12 AND #17 | 11379 |
| 19 | <u>#18 AND (Meta-Analysis [PT] OR systematic [SB])</u> | <u>245</u> |
| 20 | ("Randomized Controlled Trial" [PT] OR "Controlled Clinical Trial" [PT] OR randomized [TIAB] OR placebo [TIAB] OR "Clinical Trials as Topic" [Mesh: noexp] OR randomly [TIAB] OR trial [TI]) NOT (Animals [MH] NOT Humans [MH]) | 1144087 |
| 21 | <u>#18 AND #20</u> | <u>2007</u> |

(2) PubMedにおける研究デザインの絞り込み方法

1) システマティックレビュー

初めに、先行するシステマティックレビューを検索する。最新で質の高いものが見つかった場合は個別研究の検索が省略できる、または掲載されている検索式を利用して続く検索を行い文献を追加することができる可能性がある。具体的には、systematic [SB] というサブセットを検索結果に掛け合わせる。ただし、別に Meta-Analysis [PT] という Publication Type もあるので、両者を OR で結ぶとより包括性が高まる。なお、Systematic Review [PT] という Publication Type も 2019 年に新設されている。他にも BMJ^{*7} や CADTH^{*8} などのフィルターがある。

2) RCT

スクーピングサーチの結果、RCT でシステマティックレビューを行うことが予定されている場合は検索結果を RCT に絞る。方法としては、PubMed の "Randomized Controlled Trial" [PT]、またはその上位概念である Controlled Clinical Trial [PT] を掛け合わせる。さらに広く拾う場合は Cochrane が用いているフィルター^{*9} を掛け合わせることを検討する。

Cochrane には PubMed 検索用の RCT フィルター 2 種類が用意されている。

感度最大化バージョン (2008 revision), PubMed フォーマット

randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized [tiab] OR placebo [tiab] OR drug therapy [sh] OR randomly [tiab] OR trial [tiab] OR groups [tiab]

感度・正確度最大化バージョン (2008 revision), PubMed フォーマット

randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized [tiab] OR placebo [tiab] OR clinical trials as topic [mesh: noexp] OR randomly [tiab] OR trial [ti] NOT(animals [mh] NOT humans [mh])

Cochrane の検索戦略の考え方は以下のとおりである。

① 検索式は 3 つの検索語句のセットからなる。

- ・ 対象 (健康状態を表す語句) P
- ・ 介入を表す語句 I
- ・ 研究デザイン

* 7 BMJ (<https://bestpractice.bmj.com/info/us/toolkit/learn-ebm/study-design-search-filters/>) ※ Ovid 用

* 8 CADTH (<https://www.cadth.ca/resources/finding-evidence/strings-attached-cadth-database-search-filters#svst>)

* 9 <https://work.cochrane.org/pubmed>

- ② 検索戦略の開発は、すでに引き出された文献集合に基づいて、検索語句が修正・改変される繰り返しプロセスである。
- ③ 検索式の修正と再検索を繰り返していくと追加で見つかる文献の数は減少していき、それ以上の繰り返しのリターンが労力に見合わない点に達する。
- ④ 多すぎる異なる検索コンセプトは避ける。しかし、それぞれのコンセプトの中で広範な同義語と関連語句（フリーテキストと制御された語彙の用語の両方を含む）を OR で結合して用いる。
- ⑤ 異なるコンセプトを AND で結合する。
- ⑥ RCT のための感度最大化検索フィルターを最初に使用し、多すぎたら感度・正確度最大化フィルターを用いる。
- ⑦ 新しい文献の検索にはインデックスされていない文献用に別の検索を行う（可能な場合）。

(3) 害の検索の進め方

観察期間が短いため RCT に含まれる害のアウトカムでは不十分なことが多い。積極的に害についての研究を探す場合は、研究デザインを RCT ではなくコホート研究（cohort study）などの観察研究（observational study）とし、“adverse effects”, “chemically induced”, “complications” などのサブヘディングを掛け合わせる方法がある。害に関する検索フィルター^{*10}はいくつか開発されている^{*11}。

(4) 費用対効果検索の進め方

医療経済評価研究の検索には、近年の研究から MEDLINE (PubMed), Embase, HTA Database^{*12} の3つのデータベース (Arber M 2017) が必要とされ、そのうち PubMed, HTA Database は無料で利用できる。より完全な医療経済評価研究を含むデータベースの一覧リストは SuRe Info^{*13} から入手可能である。費用対効果の検索をする場合は、可能な限り医療経済の専門家に相談する。医療経済評価の文献検索、システマティックレビュー、新規の医療経済評価の方法についての詳細は第5章を参照のこと。

また、費用対効果を推奨作成において考慮することを決定している場合、治療効果、診断精度などの文献検索の際に、医療経済評価研究の包含基準に当てはまるものが見つかる可能性もあるため、医療経済評価研究のシステマティックレビューの包含基準を考慮したスクリーニングを行うことが望ましい。

* 10 害の検索フィルター紹介サイト：ISSG (Search Filters Resource: Filters to Identify Studies of Adverse Effects) <https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/adverse-events-filters>

* 11 なお、益の検索の結果、十分な害に関するエビデンスが収集される場合は、害の検索は省略可能である。

* 12 CADTH Evidence Driven (<https://www.cadth.ca/resources/hta-database-canadian-search-interface>)

* 13 HTAi Vortal SuRe Info (<http://vortal.htai.org/index.php?q=home>)

(5) 診断精度研究の検索

検索式は原則として、「インデックス検査 (OR 参照基準) AND 標的診断」から構成され、さらに必要に応じて「AND 検索フィルター」が組み合わされる。

それぞれのデータベースごとに多数の診断用検索フィルターが提案されているが、システマティックレビュー作成にはどれも用いるべきでないとの Cochrane review も報告されている。しかし、検索結果の論文数が非常に多く、NNR (number needed to read) が非常に大きい場合、フィルターの感度と特異度を認識し、漏れ (必要な文献の見逃し) のリスクと作業効率のバランスを考慮した上で、いずれかのフィルターの使用を検討してもよい。

また、すでに診断精度に関するシステマティックレビュー/メタアナリシスが発表されている場合には、非直接性の評価を含めた CQ への適合度、AMSTAR による妥当性評価結果に基づいて採用を考える。

PubMed の Clinical Queries では、診断に関する検索のためのフィルター (Sensitive/Broad) が用いられている。その感度は 98%、特異度は 74%と報告されている (Haynes RB 2004)。

```
(sensitiv* [Title/Abstract] OR sensitivity and specificity [MeSH Terms] OR diagnos* [Title/Abstract]
OR diagnosis [MeSH:noexp] OR diagnostic* [MeSH:noexp] OR diagnosis,differential [MeSH:noexp]
OR diagnosis [Subheading:noexp])
```

(6) 比較効果研究の検索

PubMed Special Queries: Directory of Topic-Specific PubMed Queries には比較効果研究 (comparative effectiveness research : CER) の検索式が用意されている。

Research Category として、

- Randomized controlled trials
- Observational Studies (cohort, administrative data, registries, and electronic health records)
- Systematic Reviews, Simulations, Models

Selected Topics として、

- Health Disparities
- Costs and Cost Analysis
- Comparative Effectiveness Research as Subject

などがある。これらのいずれか1つあるいは全てを検索語句と組み合わせて検索ができる。

(7) 価値観, 意向, 効用, QOL 研究の検索

患者・市民が診療ガイドライン作成に関わる方法のひとつとして、患者・市民の価値観や希望について調査された文献のレビューがある。この検索は主に推奨作成で必要となっており、通常のシステマティックレビューとは別に行う。

主なキーワードとしては Patient Preference [MeSH] があるが、この MeSH は 2010 年に採用されている。また、最近では患者・市民の価値観や希望に関する検索フィルター (van Hoorn R 2016, Selva A 2017) もいくつか提案されている。

健康効用 (health utility) や QOL について検索する場合の主なキーワードは Quality of Life [MeSH] であるが、広く探す場合はその上位の Life [MeSH] を用いるか、フィルターの利用を検討する。

4.1.3 文献の再検索

検索時期から 1 年以上経過して診療ガイドラインがまだ完成していない場合は新しい文献が出ていないか確認のため、再検索を行う。未発表の大規模な研究が把握された場合、または重要な研究が見つかった場合は、レビューをアップデートすることを検討する。しかし時間的に難しい場合は追加情報として文献情報を記載しておく。

● 手順

- (1) 検索式の作成
- (2) 検索の実行

● テンプレート

- 【SR-1 データベース検索結果】
- 【SR-2 文献検索フローチャート】 (PRISMA 声明を改変)

● テンプレート 記入方法

- 【SR-1 データベース検索結果 記入方法】
- 【SR-2 文献検索フローチャート 記入方法】

● テンプレート 記入例

- 【SR-1 データベース検索結果 記入例】
- 【SR-2 文献検索フローチャート 記入例】

4.2 ステップ2：スクリーニング

4.2.1 一次スクリーニング

原則としてシステマティックレビューチームの2名が独立して一次スクリーニングを行う。一次スクリーニングでは、タイトル、アブストラクトからCQに合っていないもの、採用基準に合致しないものを除外する。例えば、明らかに臨床研究、疫学研究のデザインでないもの（Letter、系統的でない総説）、医薬品の臨床試験のうち phase I に相当するもの、法令、官庁からの通知なども、採用基準に合致しない場合は除外する。なお、ここではアウトカムは考慮しない。抄録で判断できないものは原則として残す。

2名の結果を照合し、二次スクリーニング用データセットを作成し、文献を収集する。

4.2.2 二次スクリーニング

原則としてシステマティックレビューチームの2名が独立して全文を読み、二次スクリーニングを行う。選択基準に合った論文を選び、2名の結果を照合するが、意見が異なる場合は第三者の意見を取り入れ、採用論文を決定する。研究の二次スクリーニング後の一覧表を**テンプレート【SR-3】**としてまとめる。

4.2.3 文献集合の作成

エビデンスの統合は研究デザインごとにまとめられたアウトカムについて行われ、さらにそれらがまとめられてエビデンス総体を形成する。**テンプレート【SR-3】**にまとめられた文献集合をアウトカムごと、研究デザインごとに並べ替え、エビデンス総体の評価の準備をする。

4.2.4 文献集合の管理

電子的に収集した文献をインターネット上のクラウドなどで共有する、あるいは大量に印刷して配布する行為は、著作権侵害に当たる可能性があるので十分注意する。文献は、文献コードによって一元的に管理することが望ましい。文献管理の方法はいくつかあるが、異なるCQで何度も同じ文献が出てくる可能性があることから、著者名、発行年で本文中に挿入し、CQごとにまとめて引用文献を記載するハーバード方式（例：Smith J 2013）を推奨する。引用文献は、CQごとに採用論文、不採用論文、その他の引用論文に分けて筆頭著者のアルファベット順に列挙する。以下に本文中の記載例を示す。ハーバード方式では、同年に公表された論文は登場順にアルファベット（a, b, c, ……）を付して区別するが、診療ガイドライン作成のための引用には年内の発行順にこだわる必要はない。

- ① 1件のRCT（Smith J 2013a）では、……
- ② 米国の大規模研究（Smith J 2013b）では、……
- ③ Smith A 2013らの報告によれば、……

①, ②は同一人物の 2013 年の研究, ③は別の人物の研究であることが分かる。

また, 不採用論文についてはその理由を記録に残すことが重要である。テンプレート【SR-4】を用いてこれらを記録する。

● 手順

- (1) 一次スクリーニング
- (2) 二次スクリーニング
- (3) 文献の管理方針を決定する
- (4) 方針に従って管理する
- (5) 引用文献を記載する

● テンプレート

【SR-3 二次スクリーニング後の一覧表】

【SR-4 引用文献リスト】

● テンプレート 記入方法

【SR-3 二次スクリーニング後の一覧表 記入方法】

【SR-4 引用文献リスト 記入方法】

● テンプレート 記入例

【SR-3 二次スクリーニング後の一覧表 記入例】

【SR-4 引用文献リスト 記入例】

4.3 ステップ3：エビデンスの評価：個々の研究に対する評価 **STEP 1**

エビデンスを評価し、強さと限界については明確に提示する。エビデンスの質は研究デザインのみで決定されるのではなく、報告内容を詳細に評価、検討し、場合によってはさらにメタアナリシスを行って評価することが望ましい。エビデンス評価は、アウトカムごとに行い、エビデンス総体の質あるいは限界を明確に記載することを心掛ける。

4.3.1 論文の選出・選択

系統的検索によって収集された文献（4.1節、4.2節参照）を基に、以下の論文選出・選択作業を行う。

① 診療ガイドラインやシステマティックレビューが抽出された場合

- ・ 現在作成中の CQ と比較して、P（対象は同一か）、I（介入は同一か）、C（対照は同一か）、O（アウトカムは同一か）などの詳細な設定を詳しく評価し、さらに CQ に関する全てのアウトカムが診療ガイドラインやシステマティックレビューに含まれているかを確認する。
- ・ 診療ガイドラインやシステマティックレビューが検索対象とした期間を確認し、同じ検索式を用いてアップデートする。その後新たな RCT などが出ている場合は、これを含めて再度評価する。場合によってはメタアナリシスを行う。

② システマティックレビューが複数抽出された場合

- ・ 上記の「診療ガイドラインやシステマティックレビューが抽出された場合」と同様の評価を行い、場合によってはメタアナリシスを行う。
- ・ 益と害の評価に十分な情報があると判定される場合は、観察研究の評価が省略できる場合がある。

③ 除外論文の取り扱い

抽出された論文のうち、表面的には適格基準を満たしているように見えても、さらに調べてみると満たしていなかった研究は、その論文名と除外理由をリストにしておく。

④ 益の評価における観察研究の取り扱い

益の評価でシステマティックレビューや十分な RCT がある場合は、観察研究の評価が省略できる場合がある。

⑤ 害の評価における観察研究の取り扱い

- ・ RCT が複数抽出されて、観察期間が十分長い場合は、観察研究の評価を省略できる場合があるが、特に害の評価においては RCT のみでは限界があることを認識すべきである。これは、RCT の研究対象症例数が益の視点から決定されているからであり、害については、観察研究も評価対象とすることが望ましい。
- ・ 薬剤副作用情報、添付文書も重要な情報である。

4.3.2 個々の研究に対する評価 STEP 1

アウトカムごとにまとめられた文献集合の個々での論文について、研究デザインごとに**テンプレート【SR-5】、【SR-6】**を用いて、バイアスリスク、非直接性を評価し、対象人数を抽出する。その際、効果指標の提示方法が異なる場合は、リスク比、リスク差などに統一して記載する。RevMan (Review Manager) などを用いてメタアナリシスと併せて計算すると簡便である。RCT でもバイアスの影響を受けている可能性があるため (Chess LE 2013, Hamm MP 2010)、個々の研究についていくつかのドメイン・項目別にバイアスリスクの評価を行う。

(1) バイアスリスク：個々の研究についての評価 (Higgins JPT 2011b: 8.8-9 を引用改変)

1) 原則

以下の Cochrane バイアスリスクツール (Cochrane Risk of Bias Tool) による原則 (Higgins JP 2011a) を参考とする。

1. 質評価用のチェックリストあるいは尺度は用いない。
2. 内的妥当性は、非直接性とは分けて評価する。
3. 方法の記載や報告の不備ではなく、研究結果から推測して評価する (Mhaskar R 2012)。
4. 評価者の判断が必要であり、機械的には決められない。
5. バイアスリスクのうち、重要視すべきドメインあるいは項目は一律ではない。
6. 研究対象は患者全体を代表しているものとして評価し、症例を除外することによってバイアスが小さくなる操作が行われていないか注意する。
7. アウトカムごとに分けて評価を行う。

出典：Higgins JP, et al. (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 343:d5928. By courtesy of Dr. Higgins より作成

また、評価者間の一致率は項目によって異なり、全体として中等度であることが知られているので (Hartling L 2011, Hartling L 2013, Vale CL 2013)、2名の評価者で異なる結果の場合は、コンセンサスを得て決定する。

2) バイアスリスクのドメイン・項目と定義

バイアスリスクはドメインによって分類され、さらに各ドメインに評価項目が含まれる。現時点では、各バイアスリスクのドメインには1つまたは2つの評価項目が含まれている (表 4-7 参照)。

表 4-7 RCT の場合のバイアスリスク

| ドメイン | 評価項目 |
|---------------------------|--------------------------------|
| 選択バイアス (selection bias) | ランダム化 (ランダム配列の生成, ベースライン不均衡) |
| | 割り付けの隠蔽 (コンシールメント) |
| 実行バイアス (performance bias) | 参加者と医療提供者の盲検化 |
| 検出バイアス (detection bias) | アウトカム測定者の盲検化 |
| 症例減少バイアス (attrition bias) | 不完全アウトカムデータ |
| | ITT (intention-to-treat) 解析非実施 |
| その他のバイアス | 選択的アウトカム報告 |
| | 早期試験中止バイアス |
| | その他のバイアスの可能性 |

出典 : Higgins JP, et al. (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 343:d5928. By courtesy of Dr. Higgins より作成

各ドメインの概念は RCT だけでなく観察研究にも適用される。

臨床研究におけるバイアスは数多く報告されており (Berkman ND 2014, Dechartres A 2011, Viswanathan M 2012), 研究ごとに異なるバイアスが異なる影響を与えている可能性がある。可能であれば, バイアスがアウトカムに影響する大きさ (magnitude), 方向 (direction) を評価すべきであるが, 臨床的な知識・経験がないと判定が難しいことが多いので, 大きなバイアスのリスクの程度, すなわちバイアスリスクとして判定することが一般的である。バイアスリスクが高い研究は過大評価になっているという報告が多いが (van Tulder MW 2009), 疾患, 介入, アウトカムによって異なる可能性があり (Fenwick J 2008), 個別の判断が重要である。

① 選択バイアス

研究対象の選択の偏りにより生じるバイアス。特に, 比較される群の研究対象が介入や危険因子への曝露以外の点で異なることによってアウトカムに影響を受けるバイアス (Armijo-Olivo S 2015, Britton A 1999, Savović J 2012)。

例) 比較試験でランダム割り付けが行われていないため, 比較される群の年齢構成が異なり, 平均年齢が高い方のアウトカムが悪くなる。

- ・アウトカムの差が介入あるいはリスクファクターへの曝露によるものかどうかを考える。
- ・非ランダム化比較試験 (non-randomized controlled trial : non-RCT) や歴史的対照群を用いる場合など, 比較される群のさまざまな特性がもともと異なる場合には, 選択バイアスが生じる。

評価項目：ランダム化**・ランダム配列の生成**

ランダム配列生成：患者の割り付けがランダム化されているか、さらにランダム化の方法が乱数表やコンピューターランダム化など適正なものかについて、詳細に記載されているかを検討する (Kjaergard LL 2001)。

Chalmers TC ら (1983) の 145 件の急性心筋炎の RCT を検討した報告では、割り付けを盲検化した RCT では死亡率が 8.8 %であったのに対し、患者を盲検化しなかった RCT では 24.4%、ランダム化が行われなかった試験では 58.1%であったとされている。

・ベースライン不均衡

症例登録時に、アウトカムに影響を与える患者属性（年齢、重症度、病期など）に比較する群間で差がないかを検討する。これらの情報は、患者背景 (patients' profile) の表にまとめられていることが多い。

ランダム配列の生成とベースライン不均衡はランダム化に関わるバイアスリスクとして、区別して評価してもよい。

評価項目：割り付けの隠蔽（コンシールメント）

患者を組み入れる担当者に組み入れる患者の隠蔽化がなされているかを検討する。介入現場での割り付けではなく登録センターや中央化などの方法が有用である (Kunz R 2007, Herbison P 2011, Nuesch E 2009a, Wood L 2008, Pildal J 2007, Schulz KF 1995)。

Herbison F ら (2011) は、389 件の RCT を解析した結果、「double blind」の記載があっても、割り付けの隠蔽化が不明瞭な場合はバイアスリスクが高くなり、RORs (pooled ratios of odds ratios) は 0.86 (95%信頼区間 0.78 ~ 0.96) であったと報告している。

- ・ 観察研究の場合は、比較される群のアウトカムに影響を与え得る背景因子がそろっているかを評価する。また、傾向係数を用いた解析（プロペンシティー解析）が行われているかを評価する。
- ・ 診断法の研究の場合は、実臨床でその診断法が実施される群が対象になっているか (single-gate study)、任意に集められた群が比較されているか (two-gate study) を評価する。

② 実行バイアス

評価項目：参加者と医療提供者の盲検化

比較される群で、介入・ケアの実行に系統的な差がある場合に生じるバイアス（Armijo-Olivo S 2017, Krogh J 2015, Gurusamy KS 2009）。

例) RCT で割り付けが分かっしまい、医療提供者がケアを変えてしまう、あるいは患者が他の治療を受けてしまうなどによってアウトカムに差が出る。

- ・ 介入群か非介入群か、あるいはどの介入が行われているのかを、患者から分からなくする（単盲検）ことに加えて、医療提供者にも分からなくしているか（二重盲検）を評価する。患者のプラセボ効果や医療提供者のバイアスを排除することを目的とする。
- ・ 盲検化されていない場合は、それが結果に及ぼす影響を評価する。

③ 検出バイアス

評価項目：アウトカム測定者の盲検化

比較される群で、アウトカム測定に系統的な差がある場合に生じるバイアス（Hróbjartsson A 2012, Liu CJ 2011）。

例) RCT で測定者が割り付けを知っしまい、新しい治療法に有利な測定結果を出してしまう。

- ・ アウトカム測定者が盲検化されているかどうかを評価する。
- ・ 盲検化されていない場合は、それが結果に及ぼす影響を評価する。
- ・ 観察研究の場合は、アウトカム測定が正確で、適切なタイミングで行われているか、測定記録が正確かなどを評価する。

④ 症例減少バイアス

比較される群で、解析対象となる症例の減少に系統的な差がある場合に生じるバイアス。

- ・ それぞれの主アウトカムに対するデータが完全に報告されているか（解析における採用および除外データを含めて）、アウトカムのデータが不完全なため、症例を除外していないかを評価する（Nüesch E 2009b）。
- ・ 症例の除外が結果に影響するほど大きいかを評価する。

評価項目：不完全アウトカムデータ

症例の減少はアウトカムが不明であることに根拠があると考えられるので、不完全アウトカムデータとして捉えられる（Riley RD 2004）。

例) 患者が通院を止めてしまい脱落症例となる、副作用により治療を中断し脱落症例となるような場合アウトカムのデータが不完全となるが、その程度が比較する群で異なり、それが群間のアウトカム評価の差に影響する。

評価項目：ITT 解析非実施

ITT 解析とは治療企図分析のことで、RCT の統計解析において、脱落例やプロトコル非合致例を無効例あるいは何らかの方法で補完されたアウトカムの例として割り付けどおりに解析することである。評価統合する報告が RCT であり ITT 解析が行われていない場合には、バイアスが生じている可能性を疑う。

⑤ その他のバイアス**評価項目：選択的アウトカム報告**

測定された複数のアウトカムのうち一部しか報告されていない場合、効果の大きい都合のいい結果だけが報告されるというバイアス（報告バイアス）を生じる可能性がある（Bialy L 2014, Norris SL 2012, Ioannidis JP 2011, Dwan K 2010, Kirkham JJ 2010, Kyzas PA 2005, Chan AW 2004a, Chan AW 2004b）。

- ・登録された研究プロトコル（研究計画書）に記載されたアウトカムが全て報告されているかどうかを評価する。

評価項目：早期試験中止バイアス

中間解析が計画されたデザインでない、あるいは適切に計画された研究（adaptive study design：適応的研究デザイン）でないにもかかわらず、当初計画されたサンプルサイズを満たす前に効果が証明されたとして中止された臨床試験の場合、効果が過大評価されるバイアスが生じる可能性がある（Bassler D 2013, Schou IM 2013, Hughes MD 1992）。

- ・あらかじめ多段階の試験が計画されたかどうかを評価する。
- ・O'Brien-Flemming 法、ベジアン解析などの方法が採用されているかどうか評価する。

評価項目：その他のバイアスの可能性

上記のバイアス以外のバイアスの可能性。

- ・利益相反（conflict of interest：COI）とは、教育・研究に携わる専門家としての社会的責任と、産学連携活動に伴い生じる利益などが衝突・相反する状態である。COIの開示・管理などの記載から、結果がCOIから影響を受けていないか評価する。
- ・多変量解析により交絡因子（confounding factor）の調整が行われているかを評価する。
- ・その他のバイアスがあり得るか評価する。例えば、単施設での臨床試験は多施設の臨床試験と異なり（Inaba Y 2009）、前者でより過大評価になるバイアスも報告されている（Unverzagt S 2013, Bafeta A 2012）。また、営利組織から資金提供を受けた臨床試験はポジティブな結果が多いことが報告されている（Als-Nielsen B 2003）。

3) Cochrane バイアスリスクツール v.2.0

RCTのバイアスリスク評価ツールとして発表されたCochraneバイアスリスクツール*¹⁴は、アップデートにより従来のバイアスドメイン名から、より直接的表現を用いた長い名称に変更されたが、概念的には従来のバイアスドメインと同じである。ドメインの数も5個に限定された。表4-8に、従来のドメイン名とAHRQの分類を比較して示す。

もうひとつの変更点は、シグナリングクエスチョンにYes / Probably Yes / Probably No / No / No information または Not applicable で答えることによって、ドメインごとの判定が決定できるようになったことである。Higgins JPTら（2019）は、その答えの組み合わせで判定が可能なExcelのマクロが付随したファイルをウェブサイトで提供しているが、使用者の判定アルゴリズムを優先することを認めているため、シグナリングクエスチョンに対する回答の組み合わせの結果によるドメインの評価と評価者の評価が異なる場合は、評価者の評価を記録できるようになっている。

また、判定の分類がHigh, Low, Unclear から High, Some concerns, Low とされた。Unclear がなくなり、3段階の判定であることがより明確になっている。

* 14 2018年にVersion 2.0が発表され、2019年8月に若干のアップデートが発表された（Higgins JPT 2019）。

表 4-8 バイアスの名称

| 従来* 15 のドメイン名/現在の Minds のドメイン (評価項目) | Cochrane バイアスリスクツール v.2.0 の分類 | AHRQ の分類 |
|--|---|--|
| Selection bias 選択バイアス (ランダム化, コンシールメント) | Bias arising from the randomization process ランダム化の過程から生じるバイアス | Bias arising in the randomization process or due to confounding ランダム化の過程あるいは交絡で生じるバイアス |
| Performance bias 実行バイアス (盲検化) | Bias due to deviations from intended interventions 治療企図からの乖離のため生じるバイアス | Bias due to departures from intended interventions 治療企図からの乖離によるバイアス |
| Attrition bias 症例減少バイアス (ITT 解析, 不完全アウトカムデータ) | Bias in measurement of the outcome アウトカムデータの欠損のため生じるバイアス | Bias from missing data データ欠損によるバイアス |
| Detection bias 検出バイアス (盲検化) | Bias in measurement of the outcome アウトカム測定におけるバイアス | Bias in measurement of outcomes アウトカム測定におけるバイアス |
| Reporting bias 報告バイアスまたは出版バイアス (出版バイアス) | Bias in selection of the reported result 報告結果の選択におけるバイアス | Bias in reporting results selectively 結果を選択的に報告することによるバイアス |
| Other bias その他のバイアス (選択的アウトカム報告, 早期試験中止, COI によるバイアス, その他) | なし | 観察研究の場合以下の 2 ドメインが追加: Bias in selecting participants into the study 研究への参加者の選択におけるバイアス Bias in classifying interventions 介入の分類におけるバイアス |

Cochrane バイアスリスクツールのような方法では、評価対象の多面的な属性をドメインおよびドメインの下位の概念である項目として設定し、それぞれの属性を評価し、全体をまとめて判定し、対象の評価を決める。判定の方法には、ドメインベース (domain-based) と目標評価基準の 2 つの手法があり、Cochrane バイアスリスクツールのような方法は、ドメインベースとされている (IOM 2011: 136)。例えば、選択バイアスのランダム配列の生成の判定では、コンピューター生成ランダム配列、封筒シャッフル法は低バイアスリスク、それ以外の方法、例えば、ID 番号や誕生日の末尾の奇数・偶数での割り付けは高バイアスリスクと判定することになっている。これは、項目が該当するかどうかで判定するものに当たるので、ドメインベースと考えられる。一方で、ベースライン不均衡という項目については、介入群と対照群の年齢や重症度にバランスがとれていない場合、それが結果に大きな影響を及ぼすかどうかを判断し、大きな影響を及ぼすと判断したら、高バイアスリスク、その可能性が低いと判断したら低バイアスリスクと判定することになる。この場合は、そのバイアス項目のアウトカムへの影響を判断するものに当たるため、目標評価基準と考えられる。すなわち、評価項目が本来の目標を達成しているかどうかを判断している。

また、Unclear の判定項目が 3 つ以上のときにその研究のバイアスリスクのまとめを高リスクバイアスと判定するというルールを設定すれば、目標評価基準になる。このように、

* 15 Cochrane バイアスリスクツール v.1.0

Cochrane バイアスリスクツールを用いる場合、評価者が意識するしないにかかわらず、ドメインベースと目標評価基準が混在して判定が行われているのが現状である。

Minds の方法は原則的に目標評価基準であり、各バイアスの項目のアウトカムへの影響の大きさの判断に基づいて、高、中、低の3段階の判定を行い、研究ごとのバイアスリスクのまとめも同じく3段階で行う。

Cochrane バイアスリスクツール v.2.0 については、評価が定まるまで時間を要するが、診療ガイドライン作成グループによっては採用を検討してもよい。

4) 観察研究のバイアスリスク評価 (ROBINS-I 含む)

ROBINS-I (Risk of Bias in Non-randomised Studies for Interventions) は観察研究のバイアスリスク評価のツールである (Sterne JA 2016)。そのガイダンスによると観察研究の中でも non-RCT、コホート研究で介入の効果を比較する研究を主な対象としている。

ROBINS-I では target trial として理想の RCT を想定し、それと評価対象の観察研究を比較しながら、バイアスリスクの程度を評価する。ターゲットトライアル (target trial) は実行可能性がないものも想定可能である。したがって、エビデンス総体の評価の際には、観察研究であってもエビデンスの確実性は A (高) からスタートしてグレードダウンする方向で考えることになる。

評価の分類は Low, Moderate, High, Critical の4段階となっている。もし、複数の観察研究のバイアスリスク評価の結果、エビデンス総体としてバイアスリスクのドメインが High になった場合は、エビデンス総体の他の評価ドメイン、非直接性、不精確性、非一貫性、出版 (報告) バイアスと併せて評価しても A (高) から C (低) へとグレードダウンすることが考えられる。

GRADE ワーキンググループは従来の GRADE アプローチにおける観察研究のバイアスリスク評価およびエビデンス総体の評価方法と ROBINS-I を用いた場合を比較し、いずれの方法でも結論は同じになる可能性が高いと述べているが、交絡で生じるバイアスの取り扱いについて注意が必要としている (Schünemann HJ 2019a)。GRADE アプローチではエビデンス総体の評価は C (低) からスタートし、大きな効果、量反応関係、効果減弱交絡因子がある場合、グレードアップするルールになっている。C (低) からスタートする理由として、未知の交絡因子の影響を受けている可能性を排除することができないことが挙げられている。

バイアス効果の大きさと方向を評価せずにバイアスのリスクしか評価しない場合、次のような問題が生じる。高リスクと判定した場合、効果減弱に働く交絡因子と、効果増強に働く交絡因子のいずれの場合もエビデンス総体の確実性をグレードダウンの方向で評価してしまうことになる。しかし、効果減弱に働く交絡因子の場合は、確実性をグレードアップする方向で評価しなければならないので、バイアス効果の大きさと方向を評価せずにバイアスのリスクのみを評価すると、エビデンスの確実性の判定を間違えることになる。ROBINS-I を用いる場合は、この点に注意が必要である。それを防ぐためにはバイアス効果の大きさと方向を評価する必要があり、ROBINS-I ではそれがオプションとなっているので、GRADE アプ

ローチの枠組みで用いる場合は必須となるであろう。

また ROBINS-I では、どのような交絡因子があり得るかを最初に疾患専門家と議論して可不足なくリストアップする作業を勧めている。この段階の交絡因子は候補であって、研究結果に影響する交絡因子と決定されたわけではない。交絡因子は、どの介入を受けるかとアウトカムの両方に関連がある因子として同定される。したがって、介入群と対照群のベースラインで頻度あるいは値に差があることが前提となる。時間経過の中で変動する因子を取り上げるかどうかは評価者の方針によるが、観察研究で十分なデータが得られる可能性は低いと考えられる。

Minds では観察研究のバイアスの評価ドメインは RCT と同じ、選択バイアス、実行バイアス、検出バイアス、症例減少バイアス、その他のバイアスを取り上げており、それらの項目として、比較される群間の背景因子の差、ケアの差、不適切なアウトカム測定、不完全なフォローアップ、不十分な交絡の調整、その他のバイアスを取り上げている。その他のバイアスに選択的アウトカム報告を含めることによって、ROBINS-I あるいは AHRQ の方法と評価項目が同じになる。比較される群間の背景因子の差は上記のごとくベースラインで差があれば交絡因子の可能性が高いと考えられるので、交絡で生じるバイアスの評価と同じと考えられる。Minds では観察研究のエビデンス総体の評価の際、ROBINS-I のように、A（高）からスタートする方法と、GRADE のように C（低）からスタートする方法のいずれも対応できる。観察研究でもデザインにより、例えば傾向スコア分析法や変数操作法あるいはメンデルランダム化法のような RCT に近いものから、症例対照研究、症例集積（case series）、症例報告（case reports）までさまざまな確実性のものがある。未知の交絡因子の影響が大きいと考えられる場合は C（低）からスタートし、その可能性が低いと考えられる場合は A（高）からスタートするという方法も取り得る。

また、AHRQ は Cochrane バイアスリスクツール v.2.0 および ROBINS-I に対応し、さらに RCT と観察研究へ共通して適用できるバイアスリスク評価法を公表している（Viswanathan M 2017）。研究者が治療の割り付けを行う non-RCT と、前向きおよび後ろ向きコホート研究で比較されるアームのある観察研究、そして症例対照研究を対象としている。

ここまでに述べた観察研究のバイアスリスク評価方法について、表 4-9 にまとめた。なお、観察研究の評価については、Cochrane から 2011 年に発表された、RoBANS^{* 16} もある。

* 16 RoBANS (Risk of Bias Assessment tool for Non-randomized Studies)

<https://abstracts.cochrane.org/2011-madrid/risk-bias-assessment-tool-non-randomized-studies-robans-development-and-validation-new>

表 4-9 観察研究のバイアスリスク評価方法

| Minds 観察研究 (2017 年) | AHRQ ^{b)} (2017 年) | ROBINS-I ^{b)} (2016 年) |
|---|---|--|
| non-RCT, コホート研究, 症例対照研究, その他の観察研究 | RCT / 非ランダム化研究 ^{c)} / 症例対照研究 | non-RCT, コホート研究, その他の観察研究 ^{d)} |
| 選択バイアス | ・ランダム化の過程あるいは交絡で生じるバイアス ^{e)} ・研究への参加者の選択におけるバイアス | ・交絡で生じるバイアス ・研究への参加者の選択におけるバイアス |
| 比較される群間の背景因子の差 背景因子の割合や値に差がある場合、アウトカムへの影響が想定されれば交絡因子の可能性はある。 | アウトカムに影響する予後因子が介入の選択にも影響する場合。 介入 / 要因曝露後に観察された属性により選択された場合 (観察研究のみ)。 | ベースラインでアウトカムの予知因子で受ける介入の予知ができる場合。 適格例の一部の除外, 一部の参加者の最初の経過観察期間, いくつかのアウトカムの生起が介入とアウトカムの両方に関係するとき, 効果が同じ介入でも介入とアウトカムに関連が生じる。(例: すでにその治療を受けていた者が参加する。) |
| 実行バイアス | 治療企図からの乖離によるバイアス | 治療企図からの乖離のため生じるバイアス |
| ケアの差 比較する群はそれぞれの比較する介入が一定でなく, アドヒアランスが異なったり, 他の治療を受けたり, ケアに差がないか。 | 企図された介入と実際に受けた介入が異なる場合。 | 介入群と対照群で受けるケアに系統的な差があり, 企図された介入からの乖離を表す。このドメインのバイアス評価は, 効果のタイプ (介入への割り付けの効果か, 介入の開始とアドヒアランスの効果か) による。 |
| 検出バイアス | アウトカム測定におけるバイアス | アウトカム測定におけるバイアス |
| 不適切なアウトカム測定 アウトカム測定者が介入に気付いたり, アウトカム測定や評価の信頼性が低かったりしないか。 | 全体としての, あるいは系統的なアウトカム評価における群間の差。 | アウトカムデータ測定の差別的あるいは非差別的過誤によってもたらされる。そのようなバイアスは, もし, 異なる介入群で異なる方法が用いられたり, 測定過誤が介入の状態あるいは効果に関連していたりする場合, アウトカム評価者が介入の状態に気付くことで起き得る。 |
| 症例減少バイアス | データ欠損によるバイアス | データ欠損のため生じるバイアス |
| 不完全なフォローアップ フォローアップが不完全で, アウトカムデータの記録が不正確で, 不完全なアウトカムデータか。 | 全体としての, あるいは系統的な解析に含まれない参加者の減少による群間の差。 | 最初に対象に含めフォローされた個人の, その後のフォローアップができなかったとき (予後因子に影響されるフォローアップの差別的逸失のような場合) にバイアスが生じる。すなわち, 介入の状態あるいは交絡因子のようなその他の変数に関する情報の逸失のあった個人の除外によるバイアス。 |
| | 結果を選択的に報告することによるバイアス | 報告結果の選択におけるバイアス |
| | 結果に基づいてアウトカムを選択的に報告する。 | 知見に基づくような, あるいは, メタアナリシス (あるいはその他の統合) による推定値が得られなくなるような結果の選択的報告。 |
| その他のバイアス | 介入の分類におけるバイアス | 介入の分類におけるバイアス |
| ・不十分な交絡の調整 ・その他のバイアス 多変量解析, 傾向スコア法, 操作変数法, メンデルランダム化法などにより交絡の対処が十分か。 選択的アウトカム報告によるバイアスはその他のバイアスに含める。 | 介入開始時に, 介入の状態が有効で信頼できる方法で記録されず, 参加者の介入の状態が間違っ分る。 | 差別的な, あるいは非差別的な介入状態の誤った分類により引き起こされるバイアス。非差別的な誤った分類はアウトカムとは関連がなく, 通常介入の効果は無効果の方向へのバイアスとなる。差別的な誤った分類は介入の誤った分類がアウトカムあるいはアウトカムのリスクと関係があるときに引き起こす可能性がある。 |

太字: ドメイン名 (概念的な単位), 下線: 項目名

- a) バイアスリスクの評価の分類は, Low, Moderate, High, Unclear である。AHRQ ではバイアス効果の向き (過大評価・過小評価) と大きさの 2 つを評価することを求めている。
- b) バイアスリスクの評価の分類は, Low, Moderate, Serious, Critical, No information となっている。
- c) 研究者が治療の割り付けを行う non-RCT と前向きおよび後ろ向きコホート研究で比較されるアームのある観察研究。
- d) ROBINS-I は non-RCT, 準ランダム化比較試験, コホート研究, 症例対照研究に適用される。なお, ガイダンス^{*17} では研究への参加者が比較される異なる介入を受け, 経過観察されるコホート型研究を対象としている。
- e) アウトカムに影響を及ぼす予知変数が受ける介入とも関連している場合をいう。AHRQ の分類では, 非ランダム化研究 = 研究者が割り付けを行う non-RCT と比較対照がある前向きおよび後ろ向きコホート研究に適用されるが, RCT には適用されない。AHRQ の分類では, これらのバイアスの適用は, RCT, 非ランダム化研究, 症例対照研究の 3 つに分類されている。概念的には選択バイアスに属する。AHRQ はランダム化の過程あるいは交絡で生じるバイアスと 1 つにまとめて取り扱っている。

* 17 ガイダンス (ROBINS-I の解説) <https://sites.google.com/site/riskofbiastool/welcome/home/current-version-of-robins-i/robins-i-detailed-guidance-2016>

5) 定量的バイアス評価

Cochrane バイアスリスクツール (Higgins JP 2011a) では研究結果あるいは結論に顕著な影響を与えるバイアスを実質的なバイアスと定義し、実質的バイアスの可能性が高い場合にバイアスリスクは「高」、可能性がないか低い場合に「低」、可能性が不明か可能性はあってもその効果あるいは結果への影響が不明の場合は「不明」の判定をする。ただし、どの程度の大きさのバイアスを実質的バイアスとするかはレビュアーが決める必要がある。また、バイアスの方向と大きさの判定が困難な場合は、不明とする。研究ごとのバイアスリスクのまとめは推奨レベルになっており、今までの Cochrane review では必須事項になっていない。現在 Cochrane バイアスリスクツールは v2.0 が発表されており、それぞれのバイアスのドメインについてシグナリングクエスチョンに No, Probably No, No information, Probably Yes, Yes のいずれかで答えながら、最終的に高 (High)、いくらかの懸念 (Some concerns)、低 (Low) の3段階で評価する方法へ改訂された (Higgins JPT 2019)。

Cochrane バイアスリスクツールでは1つの研究でバイアスリスクの1つのドメインあるいは項目が「高」の判定をする場合、その研究のバイアスリスクのまとめは「高」と判定することを前提とする。複数の研究のまとめは各研究のサンプルサイズとイベント率で重み付けをして、エビデンス総体として評価する。

Minds の判定方法は以下のとおりである。

1. それぞれの論文についての評価

バイアスは、上記の各要素について、評価者の判断によって、高リスク (-2)、中/疑い (-1)、低リスク (0) の3段階で評価する。その判断は評価者の知識、経験、専門領域などの影響を受けるため、評価者によって異なる判定がなされることがあり得る。できるだけ、2名の評価者により、判定が異なる場合には、意見を調整し統一する。

バイアスのリスクがあるかないか、リスクが高いか低いか、リスクが高い場合どの程度高いかという評価が行われる。

2. バイアスリスク「まとめ」の判定 (STEP 1 表 4-7)

ほとんどが-2……「まとめ」⇒とても深刻なリスク (-2)

3種が混じる……「まとめ」⇒深刻なリスク (-1)

ほとんどが0……「まとめ」⇒リスクなし (0)

(2) 非直接性：個々の研究についての評価

非直接性にはさまざまな名称が用いられてきたが (Turner RM 2009, Atkins D 2011) *18、ある研究から得られた結果が、現在考えている CQ や臨床状況・集団・条件へ適用し得る程度を示す。CQ の PICO の各項目と研究の PICO とのずれが結果にどのように影響するかを推定し、CQ で問われている効果推定値を明らかにする。

* 18 非直接性のうち、対象者の相違について適用可能性として分離する考え方もある。AHRQ はそのような考え方を採用しており、非直接性の概念はアウトカムが代理アウトカムであるような場合と、介入の比較が直接的な比較ではなく間接的に行われている場合に限定して用い、適用可能性については臨床において医療者が判断することを前提としている。

非直接性の評価は疾患専門家でないとい困難である。日本人を対象とした診療ガイドラインで欧米人を対象にした研究を利用する場合、遺伝的背景、医療環境、生活習慣などの違いがアウトカムへ与える影響を推定する必要がある。個々の研究についての評価を行った上で、エビデンス総体の評価を行う。

以下の項目について検討する。

- ・研究対象集団の違い（適用可能性：applicability）

例）年齢が異なる。

CQ では 85 歳未満を対象にしているが、研究論文には年齢制限を設けていない。

- ・介入の違い（適用可能性）

例）薬剤の投与量、投与方法が異なる。

欧米の報告では 100mg の NSAIDs 坐薬を推奨しているが、日本には 50mg までの剤形しかない。

- ・比較の違い

例）設定した CQ と検索収集した研究報告で、対照か別の介入があるかどうか異なる。

介入の直接比較でなく、共通の対照を介した間接比較である。

- ・アウトカム測定の違い（surrogate outcomes）

例）設定した CQ と検索収集した研究報告で、アウトカム測定が異なる。

〈判定方法〉

非直接性が高（-2）

非直接性が中／疑い（-1）

非直接性が低（0）

なお、もし個々の研究については非直接性の評価を実施せず、エビデンス総体の評価の際に実施するのであれば、**テンプレート【SR-3】**を作成する際に、PICO の各項目に非直接性の判断に必要な情報を記載しておく必要がある。その情報は研究対象者の適格基準・除外基準ではなく、実際に解析対象となった症例の属性を含む必要がある。介入についても、日本で実行可能な介入との差を判断できる情報が必要である。

(3) 値の抽出

アウトカムが名義変数で二値変数の場合、例えば治癒と非治癒、発症と非発症、改善と非改善などのような場合は、それぞれの群の症例数（分母＝リスク人数）とアウトカム生起例数（分子＝イベント数）を抽出する。これらの値からは、リスク比、リスク差、オッズ比などの効果指標の統合値がメタアナリシスによって算出できる。

アウトカムが、連続変数、例えば入院日数、臨床検査の測定値などのような場合は、それぞれの群の平均値と標準偏差の値を抽出する。これらの値からは、平均値差または標準化平均値差が統合値として算出される。

生存分析の結果の場合は、各群の症例数とハザード比とその標準誤差を抽出する。ハザード比とその標準誤差からメタアナリシスが行われるが、これらの値が直接記載されていない場合には計算が必要となる（Tierney JF 2007）。ハザード比と95%信頼区間、ログランク検定の結果、ハザード比と総イベント数などさまざまな指標から計算することが可能である。

● 手順

- (1) 論文の選出・選択
- (2) 個々の研究に対する評価
 - (2)-1 バイアスリスクの評価
 - (2)-2 非直接性の評価
 - (2)-3 値の抽出

● テンプレート

- 【SR-5 評価シート 介入研究】（RCT 評価用）
- 【SR-6 評価シート 観察研究】（観察研究評価用）

● テンプレート 記入方法

（なし）

● テンプレート 記入例

- 【SR-5 評価シート 介入研究 記入例】（RCT 評価用）
- 【SR-6 評価シート 観察研究 記入例】（観察研究評価用）

4.4 ステップ4：エビデンス総体の評価 **STEP 2**

エビデンス総体とは、ある臨床上の問題（CQ）に対して収集し得た全ての研究報告を、アウトカムごと、研究デザインごとに評価し、その結果をまとめたものである。このアウトカムは、スコープ作成時に1つのCQに対して「患者にとって重要、重大な各アウトカム」として列挙されたものである（Guyatt GH 2011b）。

エビデンス総体をアウトカム横断的に統合した全体を「エビデンス総体の総括」と呼ぶ。診療ガイドライン作成過程においてCQに対する推奨を提示する場合（第6章参照）に、エビデンス総体の総括としての強さを1つだけ決定する（Guyatt G 2011a）。

4.4.1 エビデンス総体評価

1つのアウトカムで選択抽出された複数の論文をまとめて、エビデンス総体を評価する。RCT、観察研究など研究デザインごとにそれぞれの文献集合をまとめ直し、あらためてバイアスリスク、非直接性を評価し、これに加え、非一貫性、不精確性、出版（報告）バイアスなどを評価する。

システマティックレビューを疾患専門家以外が行った場合、非直接性の評価には限界がある。また、推奨の方向や強さが想定できない状態で、エビデンスの確実性（強さ）を決定するのも困難であろう。そのような場合は、システマティックレビューでのエビデンス総体の評価と推奨作成時の最終的なエビデンス総体の評価は異なる可能性があることを前提に、前者をエビデンスの確実性（質）、後者をエビデンスの確実性（強さ）として記録することが望ましい。

- ・ 現在目的としている1つのアウトカムが、評価項目（エンドポイント）として記載されている論文を全て抽出する。さらに、これらの論文を研究デザインごとにまとめる。以降、まとめられた論文を対象として作業を行う。
- ・ **4.3節 STEP 1** で行った「バイアスリスク評価」結果を、あるアウトカムに関連して抽出し、まとめられた「論文の全て」を対象として統合し、「バイアスリスク評価」のまとめを記載する。非直接性については、個別の論文に対して行われ、かつ、1つのアウトカムに関してまとめられた複数の論文間の評価にも用いられる。
- ・ 同様の作業を、非一貫性、不精確性、出版（報告）バイアスに対しても行い、バイアスリスク、非直接性と合わせて計5項目を評価する。
- ・ この5項目を用いて、1つのアウトカムに対する「エビデンスの総体」を評価し、記載する。

- ・アウトカムごとにまとめられたエビデンス総体の確実性（質）はシステマティックレビュー段階で、**表 4-5** で示す分類で評価される。推奨作成の段階では、エビデンスの確実性（強さ）として評価する。本マニュアルが推奨するエビデンス総体のエビデンスの確実性（強さ）は**表 4-10** のとおりであり、GRADE アプローチと同じ分類である。

表 4-10 エビデンス総体のエビデンスの確実性（強さ）

| | |
|-----------|--------------------------------|
| A（強）： | 効果の推定値が推奨を支持する適切さに強く確信がある |
| B（中）： | 効果の推定値が推奨を支持する適切さに中程度の確信がある |
| C（弱）： | 効果の推定値が推奨を支持する適切さに対する確信は限定的である |
| D（とても弱い）： | 効果の推定値が推奨を支持する適切さにほとんど確信できない |

- ・なお、この後、推奨を決定する作業として、各アウトカムに対するエビデンス総体の評価結果を見渡して、アウトカム全般に関する全体的なエビデンスの強さ（エビデンス総体の総括）を決定することになる（**第 6 章**参照）。この過程はいわば **STEP 3** である。この作業はシステマティックレビューには含まれない。
- ・決定されたエビデンスの強さは、CQ とその推奨に対するエビデンスの強さを意味し、「検索し統合されたエビデンス総体が、推奨を支えるものとしてどれほど適切か」を示している。決定にあたっては、益と害も考慮し、推奨文作成に最も重要な要素となったエビデンス総体のレベルを選択する。

4.4.2 エビデンスの確実性の評価を下げる 5 項目

エビデンス総体について次の 5 つの項目で評価を行う。

(1) バイアスリスク

4.3 節 STEP 1 の結果をまとめて、バイアスリスク評価として記載する。

(2) 非直接性

4.3 節 STEP 1 で評価された個々の研究に対する非直接性をまとめた上で、プールされた論文全体が現在検討中の CQ や臨床状況・集団・条件へ適用し得る程度を再度検討する。

- ・研究対象集団の違い

例) 年齢が異なる。

CQ では 85 歳未満を対象にしているが、研究論文には年齢制限を設けていない。

- ・介入の違い
- ・比較の違い

例) 設定した CQ と検索収集した研究報告で、対照か別の介入があるかどうか、異なる。

・アウトカム測定の違い

例) 設定したCQと検索収集した研究報告で、アウトカム測定が異なる。

〈判定方法〉

- とても深刻な非直接性あり (-2)
- 深刻な非直接性あり (-1)
- 非直接性なし (0)

(3) 非一貫性

アウトカムに関連して抽出された全て(複数)の研究を見ると、報告によって治療効果の推定値が大きく異なる(すなわち、結果に異質性または、ばらつきが存在する)ことを指し、根本的な治療効果に真の差異が存在することを意味する。

〈判定方法〉

- とても深刻な非一貫性あり (-2)
- 深刻な非一貫性あり (-1)
- 非一貫性なし (0)

研究間の結果のばらつきは、サンプリングエラーに伴う偶然の偏り、バイアスによる偏り、非直接性による偏り、説明できない偏りによって生じる。これらの偏りの程度が研究ごとに異なるため、結果にばらつきが生じる。ここでいう非一貫性は説明できない結果のばらつきのことである(Guyatt GH 2011f)。

非一貫性の判定は、メタアナリシスの結果として得られるフォレストプロットを見て、リスク比、ハザード比などの相対効果指標の点推定値の分布のばらつきと、95%信頼区間の重なりを評価して行う(ばらつきが大きく出やすいので、オッズ比、リスク差は用いるべきでない)。

統計学的な異質性(I^2 値、 τ^2 値、Q統計値のP値、 H^2 値を指標とする)は臨床的な異質性と研究方法の異質性が原因で生じるが、かならずそれらを検出できるとは限らない。あくまで非一貫性の根拠の1つとして使用すべきである。Cochraneハンドブック*¹⁹あるいはGRADEハンドブック(Schünemann H 2013)では、 I^2 値について以下のような基準が提示されているが、重なりが多いので I^2 値だけで非一貫性の判定をするのは困難であり、個別の事例に応じて考える。

| | |
|-----------|------------|
| < 40% | おそらく小さい |
| 30 ~ 60% | おそらく中等度 |
| 50 ~ 90% | おそらくかなり大きい |
| 75 ~ 100% | おそらく大きい |

* 19 Cochrane Handbook for Systematic Reviews of Interventions (<https://training.cochrane.org/handbook/current>)

臨床的な異質性と研究方法の異質性の判断を行った上で、統計学的な異質性の指標を参考にし、フォレストプロットで95%信頼区間が全ての研究で重なっているかを見て、全体として判断することが望ましい。

(4) 不精確性

サンプルサイズやイベント数が少なく、そのために効果推定値の信頼区間が幅広いことを指す。プロトコルに示された予定症例数が達成されていることが必要である。

〈判定方法〉

- とても深刻な不精確性あり (-2)
- 深刻な不精確性あり (-1)
- 不精確性なし (0)

不精確性の判断方法については議論の多いところである (Schünemann HJ 2016)。GRADE アプローチでは、システマティックレビューの段階と推奨作成の段階で異なる不精確性の定義を用いている (Schünemann H 2013, Guyatt GH 2011e)。

前者の場合、メタアナリシスの結果を1つの臨床試験の結果と見なした場合に必要なサンプルサイズである最適情報量 (optimal information size : OIS) を満たすかどうか、次に無効果の効果推定値 (例えばリスク比なら 1.0, リスク差なら 0) が95%信頼区間に含まれるかを判定する。OIS が満たされ、95 % 信頼区間に含まれない場合はレートダウンしない。もし OIS の基準が満たされ、95 % 信頼区間が無効果を含み (例えば、リスク比の信頼区間が 1.0 を含む)、重要な益あるいは重要な害を除外することができない場合には、不精確性によりレートダウンする。この場合でも、効果が大きく、重要な益あるいは害があると考えられる場合にはレートダウンしない。

推奨作成の段階では、信頼区間の上・下限値が決断閾値の片側にあるかどうかを考慮する。治療推奨と非推奨の臨床決断閾値を信頼区間がまたぐかどうかを考慮し、もしそうなら (すなわち、信頼区間が閾値と交差する)、点推定値と信頼区間がどこにあっても不精確性によりレートダウンする。もし、信頼区間が交差しなければ、OIS を満たすかどうかを見る。満たされない場合は、不精確性によりレートダウンする。

ただし、例外としてイベント率が低く、サンプルサイズが大きい場合は、OIS の基準は適用せず信頼区間の基準だけで判定する。

OIS の計算には R の TrialSize パッケージ (Zhang E 2017)、ウェブツールとして無料で提供されている CCRB (Centre for Clinical Research and Biostatistics) (CCRB 2017)、CRAB (Cancer Research And Biostatistics) : Statistical Tools (CRAB web) を利用することができる。

最近、Hultcrantz M ら (2017) は益のアウトカムに対するエビデンス総体の確実性の決定には、害のアウトカムに対するエビデンス総体の確実性が影響を与えるので、複数のアウトカム全体の文脈を織り込んだ確実性の決定が、推奨作成のためには必要だと述べている。そのためには、1つのアウトカムに対するエビデンス総体の確実性の評価のために臨床閾値を設定する際

に、他のアウトカムに対するその介入の効果の大きさと不確実性を考慮しながら決めた上で、不正確性の評価を決め、そのエビデンス総体の確実性を決める必要がある。臨床閾値の設定は価値観に影響されるので、エビデンスの確実性の概念に含めるべきかという議論も今後必要である。

(5) 出版（報告）バイアス：複数の研究の総合評価

研究が選択的に出版されることによって、根底にある益と害の効果が系統的に過小評価または過大評価されることをいう（Guyatt GH 2011d, Hedin RJ 2016, Driessen E 2015, Roest AM 2015, Niemeyer H 2012, Poorolajal J 2010, McGuire W 2010, Lee A 2006, McLeod BD 2004, Song F 2000, Song F 2010）。臨床試験計画時のサンプルサイズ設定で過小に設定するため（Clark T 2013）、研究結果が有意にならず出版されにくくなる可能性も報告されている（Aberegg SK 2010）。サブグループ解析として出版された研究は過大評価が多いことも報告されている（Hahn S 2000）。

なお、個別の論文評価としての、選択的アウトカム報告バイアスは、前述のバイアスリスクとして評価する。

〈判定方法〉

とても深刻な出版（報告）バイアスあり（-2）

深刻な出版（報告）バイアスあり（-1）

出版（報告）バイアスなし（0）

出版（報告）バイアスの評価は、メタアナリシスの際に作成するファンネルプロット（funnel plot）の目視、および Egger 検定、Begg 検定の結果で $P < 0.1$ を基準として総合的に判断する。ファンネルプロットではサンプルサイズの小さい、精度の低い研究の効果指標が左右均等に分布していない場合、出版（報告）バイアスを疑う。

出版（報告）バイアスへの対処法として trim and fill method（van Aert RC 2016, Watzlawick R 2014, Schwarzer G 2010, Sterne JA 2001）などが用いられているが、感度分析の一環として施行することが勧められている（Peters JL 2007）。

4.4.3 観察研究に対するエビデンスの確実性の評価

観察研究に対するエビデンス評価の考え方も、RCT と同様である。ただし、エビデンスとして対象となる論文は、前向きコホート研究や各国の国民規模の症例対照研究など、背景が十分調整された研究や大規模な研究である。

観察研究の場合は、エビデンス総体のエビデンスの強さは「弱」から評価を開始する。しかし、まれながら研究結果によっては、観察研究のエビデンスの強さの評価を「中」あるいは「強」に上げることもある。いろいろな理由により RCT を行うことが難しいが、その介入の効果が極めて大きく、明らかに有効で、日常診療で通常行われている場合などが該当する。基本原則は、エビデンスがある

推奨を支持するのにどれだけ確かで、確信できるかによって判定されるということであり、これは RCT と同様である。

観察研究のエビデンスの強さを評価する場合には、**4.4.2 項**で挙げた項目で強さを下げることとを考慮するほかに、以下の項目で強さを「中」に上げることも検討する (Guyatt GH 2011h)。

(1) 介入による大きな効果

飛行機から降下する場合、パラシュートを使うとほとんど救命され、使わないと死亡することが多いが、このように介入の効果が極めて大きい場合にはエビデンスの強さを上げることを考慮してもよい。介入（治療）を行うとほとんど救命され、行わないとほとんど死亡するような場合が相当する。

〈判定方法〉

効果が非常に大きい (very large) RR > 5 または < 0.2 (+2)
 効果が大きい (large) RR > 2 または < 0.5 (+1)
 上記以外 (0)

(2) 用量 - 反応勾配

報告では介入の量（回数、投与方法）が少ないが、より多くの量の介入を行えば有意差が出ていたと考えられる場合には、エビデンスの強さを上げることを考慮してもよい。

〈判定方法〉

非常に大きな用量 - 反応勾配が考えられる (+2)
 大きな用量 - 反応勾配が考えられる (+1)
 上記以外 (0)

(3) 可能性のある交絡因子による効果の減弱

報告の中に調整できていない交絡因子があり、その交絡因子が報告されている効果を減弱させていると想定される場合には、エビデンスの強さを上げることを考慮してもよい。例えば、介入を行った群には高齢者、糖尿病の患者が多かったため、効果としての死亡率がわずかしき改善しなかったが、もし、背景が均一化されていれば大きな有意差が出ていただろうと想定される場合などが相当する。

〈判定方法〉

可能性のある交絡因子が、効果を大きく減弱させていると考えられる (+2)
 可能性のある交絡因子が、効果を減弱させていると考えられる (+1)
 上記以外 (0)

なお、観察研究の評価を省略する場合はその理由を記載し、エビデンスの強さは「弱」のままとする。

4.4.4 エビデンスの統合

(1) 定性的システマティックレビュー

エビデンス総体を質的に統合することを定性的システマティックレビューという。定性的システマティックレビューは各研究のバイアスリスクの評価と非直接性の評価、それらを反映したエビデンス総体のバイアスリスクと非直接性の評価、エビデンス総体を構成する研究間の非一貫性、不精確性、報告（出版）バイアスなどの評価と臨床的文脈の評価も含まれる。**4.4.1**～**4.4.3**項までをまとめた作業のことである。

なお、定量的システマティックレビューを行った場合でも、効果指標の統合値と信頼区間だけでエビデンスの強さを決定せず、定性的システマティックレビューの結果もエビデンスの強さの評価に反映させるため、診療ガイドライン作成において必須の作業となる。

(2) 定量的システマティックレビュー（メタアナリシス）

研究結果をまとめる際に統計学的に効果指標の値を統合し、その信頼区間とともに提示するのが定量的システマティックレビュー、すなわちメタアナリシスである。

診療ガイドライン作成のためのシステマティックレビューで、研究デザインが同じで、PICOの各項目の類似性が高い場合には、効果指標を量的に統合するメタアナリシスが可能となる。メタアナリシスの結果、効果指標の統合値と信頼区間が得られるとともに、フォレストプロット、ファンネルプロットが得られる。これらのプロットは非一貫性、不精確性、出版（報告）バイアスの判定にも有用となる。診療ガイドライン作成におけるシステマティックレビューは定量的システマティックレビュー、すなわちメタアナリシスだけではなく、定性的システマティックレビューを同時に行う必要がある。

メタアナリシスが実行できる場合には、その結果はエビデンス総体の強さを検討する1つの項目となる。例えば、ある介入が統計学的有意に、そして顕著によいアウトカムを導くという結果が出ている場合には、エビデンスの強さを上げることを考慮してもよい。また、小さな差しかないという結果が出ている場合には、エビデンスの強さを上げる結果ではないと判断してもよい。そして、有意な差がない場合には、エビデンスの強さを下げることを考慮してもよい。

1) メタアナリシスが省略できる場合

メタアナリシスは必須ではない。次の場合にはメタアナリシスは省略できる可能性がある。

① 定性的にエビデンスの強さが保証できる場合

すなわち、エビデンスの強さが定性的評価から論理的に説明ができ、効果の確実性が保証されていると評価可能である場合は、メタアナリシスを省略してもよい。例えば、対象とした論文が全て同じ結論である場合などがこれに相当する。

② 同じ研究デザインの研究報告が1つしかない場合

あるアウトカム・介入群に関する報告のうち、同じ研究デザインについての報告が1つしかない場合には、その報告自体の値しか統合に利用することができないため、メタアナリシスの必要がない。

③ 診療ガイドライン作成グループによって定められたCQおよび方法と同様のメタアナリシスがある場合

例えばCochrane reviewや先行する診療ガイドラインのシステマティックレビューが診療ガイドライン作成グループの定めた方法とほぼ同様であり、その結果を利用できると判断する場合には、そのメタアナリシスの結果をエビデンスの評価に用いることが可能である。ただし、これらのレビュー報告後に新たな研究報告がなされている可能性があり、さらなる注意深い検索が必要である。

2) メタアナリシスのためのソフトウェア

メタアナリシスのためのソフトウェアは無料のものも含め多数存在するので、それぞれが使いやすいと思われるものを必要な機能に応じて使用すればよい。

Cochraneはシステマティックレビューを行うためのRevMan（現バージョンは5.4）と呼ばれるソフトウェアを無料で提供しており、その中にメタアナリシスのプログラムが含まれている。ウェブサイトからダウンロードして自由に使用することができる。RevManでは各研究の名称や介入などを順次入力し、テーブルを作成してからデータを入力し、メタアナリシスを実行する。同じデータからリスク比、オッズ比、率差などを指標としたメタアナリシスを行うことができ、フォレストプロットをWordなどに貼り付け可能な形で出力できる（4.6節参照）。

また、統計解析のオープンプラットフォームであるRでもメタアナリシス用のさまざまなパッケージが公開されており、これらを利用することも可能である。リスク比、リスク差、オッズ比、ハザード比、平均値差などの効果指標を用いるのであれば、metaforパッケージを用いることができる。Excelでデータを用意し、簡単な操作でメタアナリシスを実行し、フォレストプロット、ファンネルプロットなどの結果を得ることができる（4.7節参照）。

3) 効果指標

メタアナリシスでは複数の研究結果を統合するが、統合されるのは効果指標の値である。効果指標にはさまざまなものがあるが、リスク比[risk ratio:RR,または相対危険度(relative risk:RR)],オッズ比(odds ratio:OR),率差[rate difference:RD,またはリスク差(risk difference:RD)],平均値差(mean difference:MD),標準化平均値差(standardized mean difference:SMD),ハザード比(hazard ratio:HR),その他が用いられている。

効果指標の値は、効果の強さあるいは大きさを定量的に表すものであることから、効果サ

イズ（effect size）とも呼ばれる。また、効果指標は一定の分布に従うが正規分布に従う場合には、ばらつきの指標として標準偏差（standard deviation）を用いることができる。効果指標の分布の標準偏差は、標準偏差とも標準誤差（standard error）とも呼ばれるが、同じ意味で用いられている。

以下、表 4-11 の四分表を基として、これら効果指標の算出について解説する。

表 4-11 効果指標算出の基になる四分表

| | アウトカム (+) | アウトカム (-) | 合計 |
|-----------|-----------|-----------------|---------|
| 治療:介入 (+) | r_i^T* | $n_i^T - r_i^T$ | n_i^T |
| 対照:介入 (-) | r_i^C | $n_i^C - r_i^C$ | n_i^C |

*その研究の治療群におけるイベント数。 r : アウトカム陽性（イベント生起）例の人数、 n : 各群の総症例数、 i : 研究番号。2群の率（割合）から RR, OR などが算出される。

① リスク比

リスク比は $(r_i^T/n_i^T)/(r_i^C/n_i^C)$ で求められる。リスク比の自然対数は正規分布に従い、その分布の標準偏差（標準誤差）は $\sqrt{1/r_i^T + 1/r_i^C - 1/n_i^T - 1/n_i^C}$ で計算される。リスク比はそれぞれの群の率の比に相当するが、対数変換するとそれぞれの群の率の対数の差になるため、モデル化が容易である。対数化したリスク比、標準誤差を用いて、統合値および 95%信頼限界を算出し、指数変換（exponential transformation）して元のスケールに戻すことが行われる。その際には、標準誤差の平方 = 分散の逆数で重み付けした平均値を求める（固定効果モデル）。分散に研究間の分散を加算して重み付けするとランダム効果モデルとなる。なお、割り算の分母が 0 になる場合には、 r および $n - r$ に 0.5 を加算する。

② オッズ比

オッズ比は $[r_i^T(n_i^C - r_i^C)]/[r_i^C(n_i^T - r_i^T)]$ で求められる。オッズ比の自然対数は正規分布に従い、その分布の標準偏差（標準誤差）は $\sqrt{1/r_i^C + 1/(n_i^C - r_i^C) + 1/r_i^T + 1/(n_i^T - r_i^T)}$ で計算される。なお、割り算の分母が 0 になる場合には、 r および $n - r$ に 0.5 を加算する。

③ 率差（リスク差）

率差は $r_i^T/n_i^T - r_i^C/n_i^C$ で求められる。2群の率の差は正規分布に従い、その分布の標準偏差（標準誤差）は $\sqrt{[r_i^T(n_i^T - r_i^T)/(n_i^T)^2] + [r_i^C(n_i^C - r_i^C)/(n_i^C)^2]}$ で計算される。

④ 標準化平均値差

標準化平均値差として、Hedge's unbiased estimator が推奨されているが、Cohen's d, Hedge's g, Glass's Δ なども用いられている。連続変数の場合の効果指標の例を以下に示す。

・Cohen's $d = (M_1 - M_2)/SD_{\text{pooled}}$

$$SD_{\text{pooled}} = \sqrt{\frac{\sum (X_A - \bar{X}_A)^2 + \sum (X_B - \bar{X}_B)^2}{n_A + n_B - 2}}$$

・Glass's $\Delta = (M_1 - M_2)/SD_{\text{control}}$

・Hedge's $g = (M_1 - M_2)/SD_{\text{pooled}}^*$

$$SD_{\text{pooled}}^* = \sqrt{\frac{(n_A - 1)SD_A^2 + (n_B - 1)SD_B^2}{n_A + n_B - 2}}$$

連続変数に対する d ファミリーの効果量。pooled：プール，control：対照。

⑤ ハザード比

ハザード比はリスク比と類似した概念であるが、時間イベントアウトカムの場合、すなわち生存分析の場合に適用される。Cox の比例ハザード解析、カプラン・マイヤー生存解析、ログランク検定などの結果からハザード比と信頼区間を算出することが可能である。

⑥ その他

リスク比の信頼区間が 1.0 を含んでいる場合には、益に対する治療必要人数 (number needed to treat for benefit : NNTB) と害に対する治療必要人数 (number needed to treat for harm : NNTH) の値が ∞ を挟んだ形になる。生存分析に基づく必要人数 (number needed to treat : NNT) の計算も可能である。

アウトカムが害の場合には害のための必要人数 (number needed to harm : NNH) が、介入がスクリーニング検査の場合にはスクリーニングのための必要人数 (number needed to screen : NNS) が用いられる。

リスク比、オッズ比はログ変換 (自然対数) することによって、正規分布に従うので、ログ変換後に正規分布を前提とした統合を行うことが多い。

4) 統合のモデル：固定効果モデル／ランダム効果モデル

固定効果モデルでは統合の対象となった研究以外の研究は想定せずに、それらの研究の効果指標の平均値を求める。ランダム効果モデルでは、実際に統合の対象となった研究以外の研究が母集団として想定され、母集団からランダムに抜き出されたものが統合の対象となった研究であると考えられる。

したがって、研究間の異質性があることを理由にランダム効果モデルを用いるという考えは誤りであり、一方で、固定効果モデルによる効果指標の統合値は対象となった研究だけをまとめたものとしては全く正しいといえる。しかしながら、臨床研究は多くの異質性を生む要素があるので、ランダム効果モデルによる統合値を算出することが推奨されている (Morton SC 2018)。

固定効果モデルの計算法の一例を示す。分散に基づく方法である。分散の逆数 (inverse variance) を重み付けに用いて効果指標の平均値を算出する。固定効果モデルでは、各研究の分散だけが重み付けに用いられる。

$$M = \frac{\sum W_i M_i}{\sum W_i}, \quad W_i = \frac{1}{V_i} = \frac{N_i}{s_i^2}$$

$$V_M = \frac{1}{\sum W_i}, \quad M \pm 1.96 \times \sqrt{V_M}, \quad Z = \frac{|\ln(M)|}{\sqrt{V_M}}$$

Inverse variance-based method とも呼ばれる。 M : 統合値, W : 重みであり分散の逆数, V : 分散, Z : 有意差検定に用いられる値 (1.96 以上で有意), i : 研究番号, s : 標準偏差 (標準誤差)。

研究間の異質性の検出には Q 統計値や I^2 統計値が用いられる。検出力が低いため、 $P < 0.1$ で有意とする場合も多い。これら 2 つの統計値の算出法に関しても一般化 Q 統計値を標準的に用いるべきであるという考えもある。

ランダム効果モデルと Q 統計値, I^2 統計値について、以下に示す。

$$Q = \sum W_i M_i^2 - \frac{(\sum W_i M_i)^2}{\sum W_i}$$

$$Q > k - 1 = df \rightarrow \text{heterogeneity}(+)$$

$$I^2 = \frac{Q - df}{Q} \times 100\%, \quad T^2 = \frac{Q - df}{\sum W_i - \frac{\sum W_i^2}{\sum W_i}}$$

$$W_i^* = \frac{1}{V_i^*} = \frac{1}{V_i + T^2}$$

T^2 : τ^2 の推定値 (データに基づき算出された効果指標の研究間の分散), i : 研究番号, M : 効果指標の値, V : 分散, W : 重み。統合値は上記の式で W の代わりに W^* を用いて計算する。DerSimonian-Laird の方法に基づく。

また、統計学的異質性に基づいて固定効果モデル ($P > 0.10$), ランダム効果モデル ($P < 0.10$) とすることは推奨されない。メタアナリシスの対象とした研究以外の研究は想定せず、対象研究の効果指標の平均値を知りたいのであれば、固定効果モデルの統合値がそれに相当する。また、ランダム効果モデルで研究間の分散である τ^2 が 0 になる場合は、固定効果モデルと同じ結果となる。

また、ランダム効果モデルの結果は効果指標の平均値 μ とその信頼区間および効果指標の分布の標準偏差すなわち上記の τ の 2 つのパラメータの解釈が必要になる。多くの場合、 μ

の値とその95%信頼区間が評価の対象とされ、95%信頼区間が1を含んでいなければ有意であり、効果があるというように解釈されている。しかしながら、このような結果は全ての研究の効果指標の平均値のとり得る値についての解釈であり、個々の研究の効果指標のばらつきは別に考える必要がある。したがって、将来同じような研究が行われた場合には、得られる効果指標の値がどのように分布するかについては、 μ と τ の両者の値に基づいたベイジアン推測が必要になる。そのためには、ランダム効果モデルベイジアンメタアナリシスが必要となる。

5) 感度分析

メタアナリシスを行う際に、ある仮定を設定する必要があるが、任意の選定あるいは決断が必要であったりする場合、その仮定や選定を変えた場合に結果がどのように変動するかを見るのが感度分析である。

6) システマティックレビューチームとメタアナリシス

システマティックレビューの実行には文献検索のスキルおよびメタアナリシスのための統計学的知識とスキルが必要とされる。また、文献の選定、感度分析、メタリグレーションの妥当性を確保するためには実診療に精通している必要がある。したがって、多くの分野の専門家から構成されるチームでシステマティックレビューを実行することが望まれる。

しかしながら、わが国の現状では、システマティックレビューチームに医学統計学の専門家をそれぞれ配置することは多くの場合困難であり、メタアナリシスの実行は、その知識・スキルを持つメタアナリシス担当者と各領域の専門家の協力の下に実行することが望まれる。

4.4.5 エビデンスの強さに関する判定および表記方法

(1) 総合的評価 (cf. 算術和)

エビデンスの強さに関する各要素・項目は、「-2」、「-1」、「0」、「+1」、「+2」といった数値を用いて評価を行っているが、「-2」は「-1」の2倍低いという意味ではなく、「-2 (とても深刻な問題)」は「-1 (深刻な問題)」という程度を示す指標として用いる。つまり、「-1 - 2 - 2 = -5」になるという計算式ではない。それぞれのまとめの評価は、それぞれの評価項目の算術和ではなく、評価者の総合的判断によって決定する (Jüni P 1999, Linde K 1999)。例えば、観察研究のエビデンスの評価を上げる3項目とも「+1」であれば、エビデンスの強さが「弱」から「中 (+1)」に上昇判定される。

(2) エビデンス総体としてのエビデンスの強さの評価

エビデンス総体としてのエビデンスの強さの評価は、研究デザインをベースにして評価を開始するが、評価を下げる項目、評価を上げる項目を考慮し、最終的な強さを決定する。

具体的には、RCTのみでまとめられたエビデンス総体の評価は「強」、観察研究 (コホート、

症例対照研究)のみでまとめられたエビデンス総体の評価は「弱」、症例報告、症例集積研究のみでまとめられたエビデンス総体の評価は「とても弱」として初期評価を与える。次に、エビデンスの強さの評価を下げる5項目、上げる3項目の検討を行う。そして、メタアナリシスが実行可能である場合にはメタアナリシスを実行する。初期評価と項目評価とメタアナリシスの結果を総合して、エビデンスの強さをA「強」、B「中」、C「弱」、D「非常に弱」に分類する。

(3) エビデンスの強さの判定作業に関する注意事項

次のような場合には特別な対応を検討してもよい。

① RCTが1つしかない場合

原則としてエビデンス総体は「弱」と判定するが、その内容(バイアスリスク、非直接性やエビデンスの強さの評価を上げる項目)を吟味して総合判定する。

② Cochrane reviewで内容が評価されている論文を用いる場合

Cochrane reviewのエビデンス評価[バイアスリスク、非一貫性、不精確性、出版(報告)バイアスなど]を利用することもできる(=Cochrane reviewの多くは、GRADEアプローチで評価されている)。

③ 海外の診療ガイドラインを利用する場合

海外の診療ガイドラインが、診療ガイドライン作成グループによって決められた方法と同様の方法によって作成されている(例えばGRADEアプローチで作成されているなどの)場合は、それぞれのエビデンス評価[バイアスリスク、非一貫性、不精確性、出版(報告)バイアスなど]を利用してもよい。ただし、診療ガイドラインの推奨内容に関しては、コンセンサスや海外の医療事情が盛り込まれているので、そのままエビデンスとして利用することは難しい。

④ 1つのCQに対してRCTと観察研究がある場合

RCTと観察研究は別シートで評価し、最初から一括して評価しないことが重要である。作業としては、RCTのエビデンス総体と観察研究のエビデンス総体をそれぞれ評価し、必要がある場合にはそれぞれの評価結果を総合して評価する。

例えば、RCTでは追跡期間が短く有意差なしとされた内容が、コホート研究では長い追跡期間の結果、大きな有意差が示されることがある。このような場合は、コホート研究のバイアスが影響しているためなのか、追跡期間が十分長いために有意差が出たのかについては診療ガイドライン作成グループで検討を行う必要がある。その上で、エビデンスの強さを決定する。

⑤ 専門家の意見/コンセンサス

専門家の意見やコンセンサスはエビデンスとしては用いないが、重要なエビデンスの引用や、方向性、追試の必要な臨床研究の考え方など重要な情報源になる可能性がある。

⑥ 先行する診療ガイドライン

過去に出版された診療ガイドラインを取り上げる場合、診療ガイドライン全体（特に推奨文）としては、コンセンサスが含まれるのでエビデンスとしては用いない。一方、その診療ガイドライン作成で行われた、それぞれのCQに対する文献収集や選択、統合などによって作成されたシステマティックレビューの結果やエビデンス評価 [バイアスリスク、非一貫性、不精確性、出版（報告）バイアスなど] については、必要に応じて利用してもよい。

⑦ 害の評価と症例集積研究／症例報告

症例集積研究や症例報告は原則としてエビデンスの強さは「非常に低い」と判定される。しかし、前述のごとく「害」の評価では重要な情報となる場合があり、注意が必要である。

4.4.6 決断分析における効果推定値の取り扱い

臨床研究では介入あるいは要因曝露の効果を明らかにするため、対照群との相対的な効果を表す効果指標が用いられる。リスク差は絶対リスク減少（absolute risk reduction）とも呼ばれ、絶対効果指標と呼ばれるが、介入群と対照群のイベント率の差なので、相対的な効果を表す指標である。なお、単一群のイベント率や平均値、例えば介入群、対照群それぞれのイベント率や平均値は絶対リスク（absolute risk）と呼ばれる。

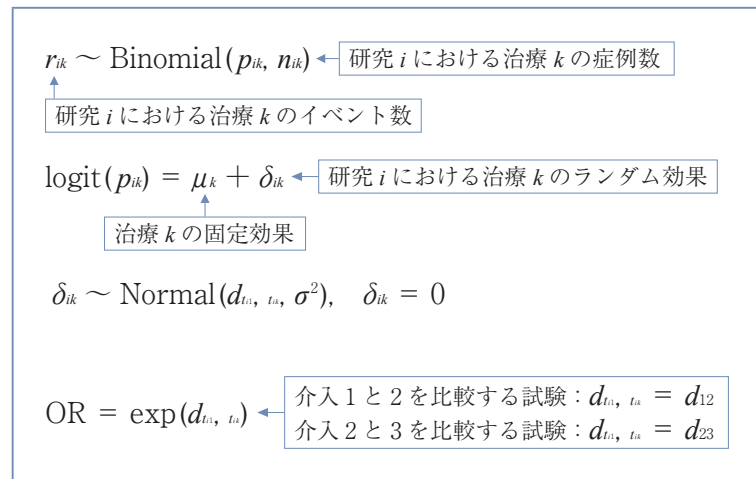
一方、多基準決断分析（multi-criteria decision analysis：MCDA）のうち、NCI/Gail法ではリスク差を用いて正味の益が計算されるが、swing-weighting その他の多くの多基準決断分析の方法で総スコア（aggregate score）を算出する場合は1つの群の絶対リスク、すなわち介入群と対照群、またネットワークメタアナリシスのような複数の介入を比較するような場合は、それぞれの介入群におけるイベント率や平均値の値が用いられる。

メタアナリシスで得られる効果推定値は通常、相対的な効果指標であり、各群のイベント率の統合値は得られない。単一群のイベント率の統合値を分散逆数法などで算出する必要がある。また、ネットワークメタアナリシスでも「アームベースモデル」（Zhang J 2014, Hong H 2015, Hong H 2016）のような手法を用いる必要がある（Hawkins N 2016）。アームベースモデルでは単一群の研究の結果を取り込む方法も開発されている（Zhang J 2019）。Rのpcnetmetaパッケージ（Lin L 2017）はベイジアン手法でアームベースネットワークメタアナリシスを実行するプログラムである。

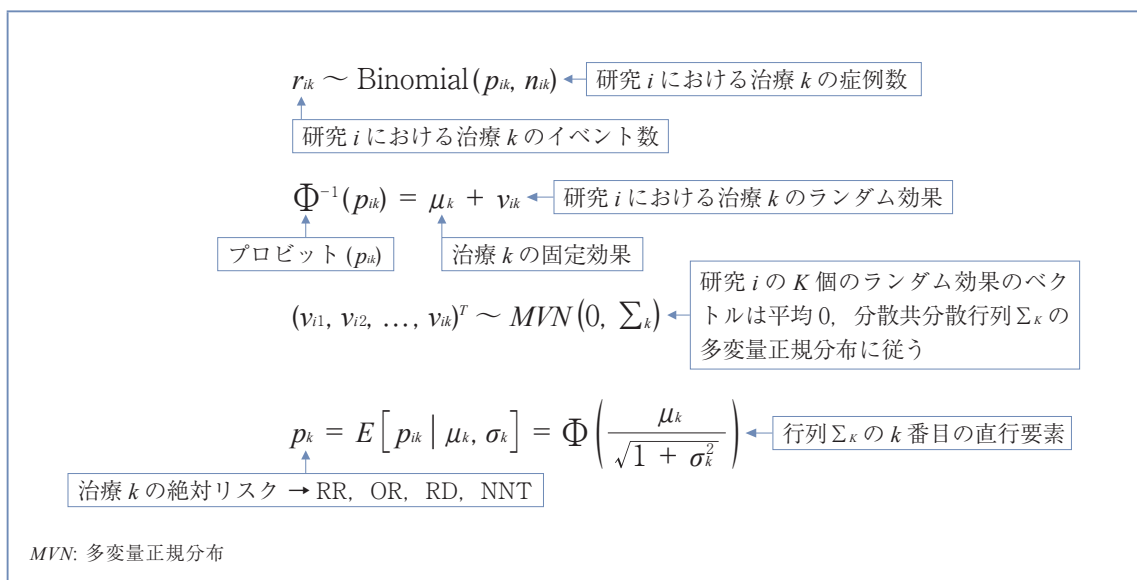
従来のメタアナリシスあるいはネットワークメタアナリシスでは「コントラストベースモデル」（contrast-based model）が用いられ、これは相対効果指標を統合するモデルであった。一方、アームベースモデルは、各治療群のイベント率や平均値が統合され、それらから相対効果指標を算出するモデルである。コントラストベースモデルでは、ベースラインリスクが変動しても相対効果指標はほとんど変動しないと見なされ、ランダム割り付けが崩されることはない。一方で、ベースラインリスクの推定値はコホート研究、疾患レジストリなどから得るか、個別に設定することが望ましいとされている。通常RCTは参加者の選択の時点で、リアルワールド（real world）の患者を必ずしも代表しなくなる場合が多いこと、治療も厳密に管理されており、リアルワールドの効果を必ずしも正確に表しているとはいえないと考えられ、RCTの対照群のデータからベースラインリスクの推定値

を求めることは必ずしも適切でないと言われてきた。一方、アームベースモデルはランダム割り付けが保証されなくなり、ばらつきが大きくなると批判されてきた。しかし、最近必ずしもそうではなく、いずれのモデルもそれぞれのバイアスの影響を受ける可能性が報告されている (White IR 2019)。いずれもモデルでも、それぞれの特性を理解して用いる必要がある。

以下に、メタアナリシスにおけるコントラストベースモデルを示す。



続いて、メタアナリシスにおけるアームベースモデルを示す。



また、マルコフモデル、discrete event simulation、マイクロシミュレーションなどで1つの健康状態から次の健康状態へ推移する確率の値を設定する場合、それらの値をメタアナリシスで直接算出することは困難である。ハザード率やイベント率から指数関数モデルなどを前提に確率を算出する必要がある。

推奨作成の際に、益と害の大きさとバランス、正味の益を定量的に評価する際にどのような効果推

定値を用いるかを十分検討しておく必要がある。確率的シミュレーションを実施する場合には、平均値だけでなく標準誤差などの不確実性の指標の値も必要となる。

● 手順

- (1) エビデンスの確実性の評価を下げる 5 項目を評価する。
 - (1)-1 観察研究の場合は、エビデンスの確実性の評価を上げる 3 項目も評価する。
- (2) 定性的システマティックレビューをまとめる。
- (3) 定量的システマティックレビュー（メタアナリシス）を行える場合はメタアナリシスを行い、結果をまとめる。
- (4) エビデンスの強さを評価する。

● テンプレート

[SR-7 評価シート エビデンス総体]

[SR-8 評価シート エビデンス総体（絶対効果指標の結果を記入する場合）]

[SR-9 定性的システマティックレビュー]

[SR-10 メタアナリシス]

● テンプレート 記入方法

[SR-9 定性的システマティックレビュー 記入方法]

● テンプレート 記入例

[SR-7 評価シート エビデンス総体 記入例（介入研究）]

[SR-7 評価シート エビデンス総体 記入例（観察研究）]

[SR-10 メタアナリシス 記入例]

4.5 ステップ5：システマティックレビューレポートの作成

4.5.1 システマティックレビューレポートのまとめ方

システマティックレビューチームは、定性的または定量的システマティックレビューの結果をエビデンス総体の強さとしてシステマティックレビューレポート（サマリーレポートともいう）にまとめ、推奨作成を行う診療ガイドライン作成グループに提出する。

CQのうち、適切な論文が検索されなかった場合、または検索された全ての論文の質が高くなかった場合など、将来的な研究が必要と判断される場合は、Future Research Questionとして記載する。

診療ガイドライン作成グループは、提出されたシステマティックレビューレポートを基に推奨作成のステップに進む。したがって、添付の資料も含めて、推奨決定の作業において必要・十分な情報を提示することに留意する。

システマティックレビューレポートにはGRADEアプローチで用いられている結果のまとめ表、SoF表（Guyatt G 2011a, Guyatt GH 2012, Guyatt GH 2013b, Carrasco-Labra A 2015, Carrasco-Labra A 2016）を作成してもよい。SoFは以下の7つの要素を含む。

- ① 全ての重要・重大な望ましい・望ましくないアウトカム
- ② これらのアウトカムのベースラインリスク
- ③ 介入群のリスク（イベント率）あるいは絶対リスク減少（絶対効果指標）
- ④ リスク比，ハザード比などの相対効果指標
- ⑤ 参加者の人数（総数）と研究数
- ⑥ アウトカムごとのエビデンス総体の確実性
- ⑦ コメントあるいは説明

4.5.2 SoF 表の作成

テンプレート【SR-12】を参考に、上記の7項目を含めるように作成する。

4.5.3 協働意思決定を考慮した結果の提示法

医療を受ける者のヘルスリテラシー，数量リテラシーを考慮し，理解しやすい表現を用いることやグラフを作成することを考慮する。AHRQのSHAREアプローチ^{*20}などを参考に資料を作成する。

* 20 <https://www.ahrq.gov/health-literacy/professional-training/shared-decision/index.html>

● テンプレート

- [SR-11 システマティックレビューレポートのまとめ]
- [SR-12 結果のまとめ (SoF 表) (ペア比較のメタアナリシス)]
- [SR-13 結果のまとめ (SoF 表) (ネットワークメタアナリシス：様式 1)]
- [SR-14 結果のまとめ (SoF 表) (ネットワークメタアナリシス：様式 2)]
- [SR-15 Future Research Question]

● テンプレート 記入方法

- [SR-11 システマティックレビューレポートのまとめ 記入方法]
- [SR-15 Future Research Question 記入方法]

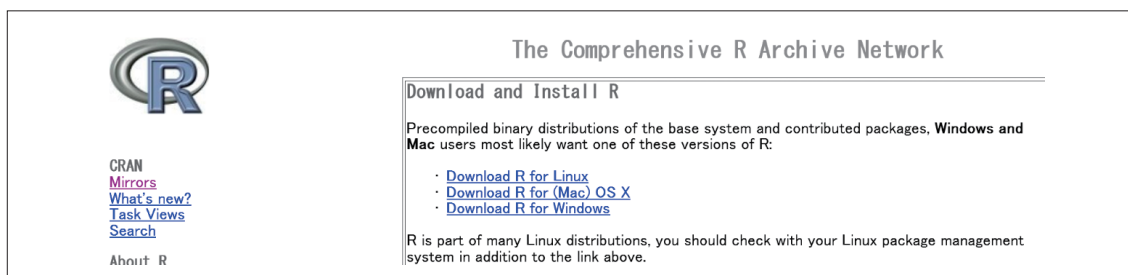
● テンプレート 記入例

- [SR-12 結果のまとめ (SoF 表) (ペア比較のメタアナリシス) 記入例]

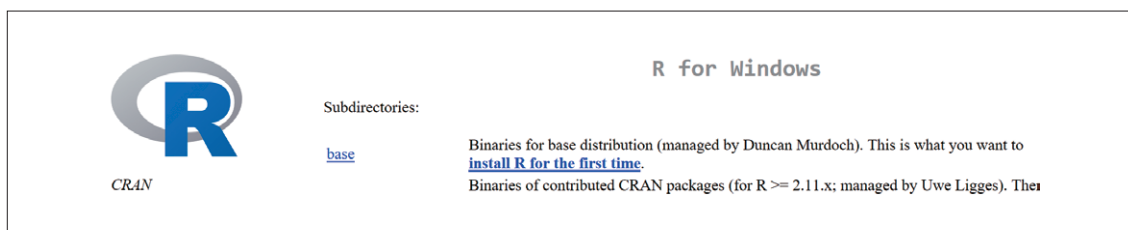
4.6 補足資料 1 : R と metafor パッケージを用いたメタアナリシス

(1) R のインストール

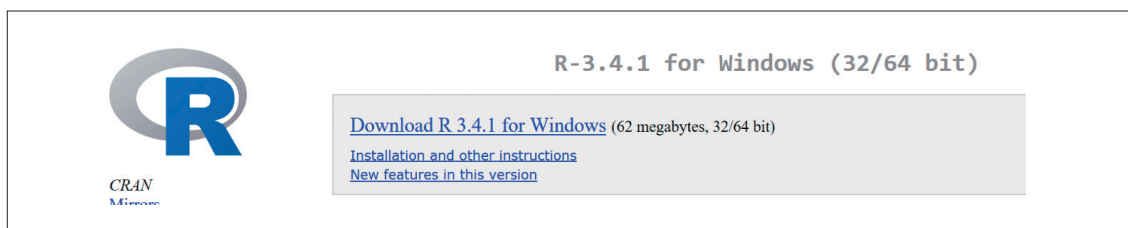
R のプログラムは The Comprehensive R Archive Network (CRAN) のウェブサイト*²¹ からダウンロード可能である。OS のプラットフォームとして Windows, Linux, Mac に対応したバージョンが用意されている。



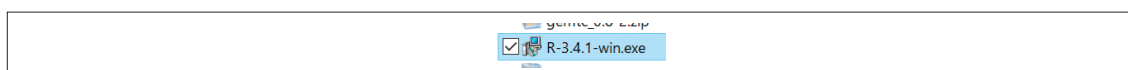
Windows の場合であれば、「Download R for Windows」をクリックし、次の画面で「install R for the first time」をクリックする。



次の画面で「Download R “バージョン No.” for Windows」をクリックして R-3.4.1-win.exe*²² ファイルをダウンロードし、任意のフォルダに保存する。



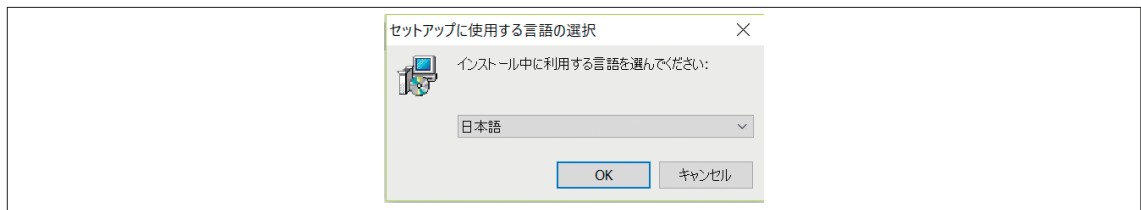
ダウンロードしたファイルをダブルクリックして、インストールを開始する。



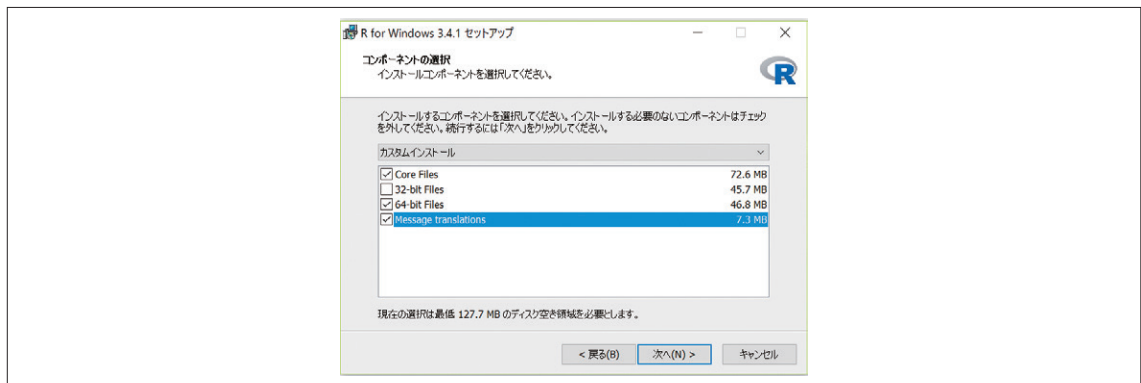
コンピューターへの変更を確認する画面が出るので [はい] をクリックする。言語を選択する画面となるので日本語を選択する。

* 21 CRAN (The Comprehensive R Archive Network) <https://cran.r-project.org/>

* 22 ファイル名の数字は参照時点のもの。



セットアップウィザードの指示に従ってインストールを進める。途中、「コンポーネントの選択」画面で「Message translations」にチェックを入れておくと、インストールされたRのメニューが日本語表示となる。

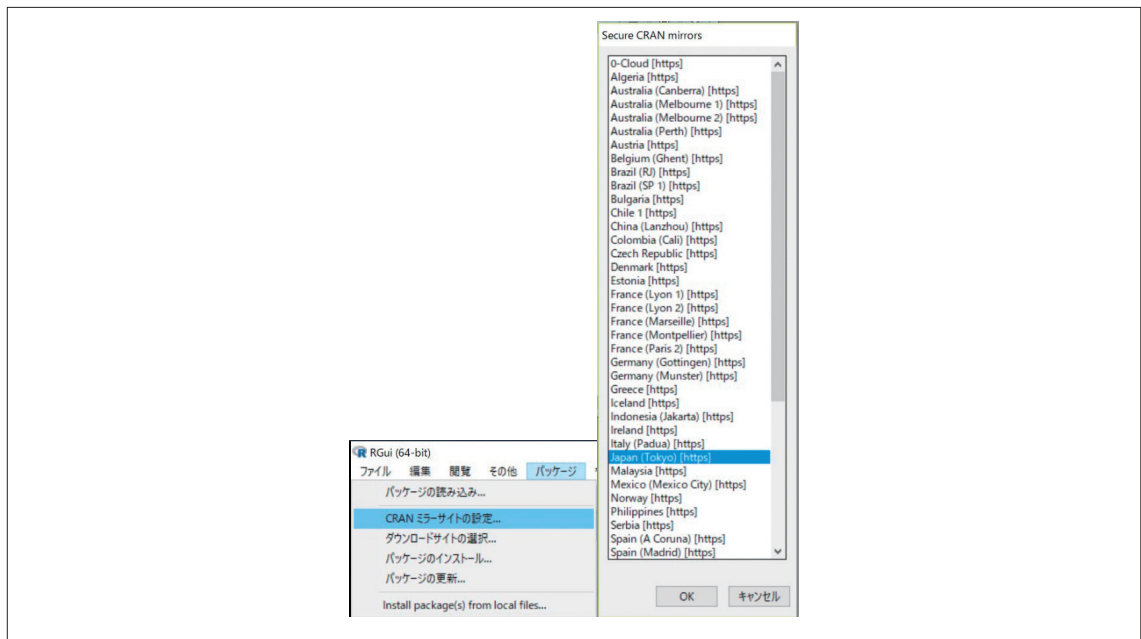


32ビット版と64ビット版が選択できるので、各自のパソコンに合わせていずれかをインストールすればよい。両方をインストールすることもできる。

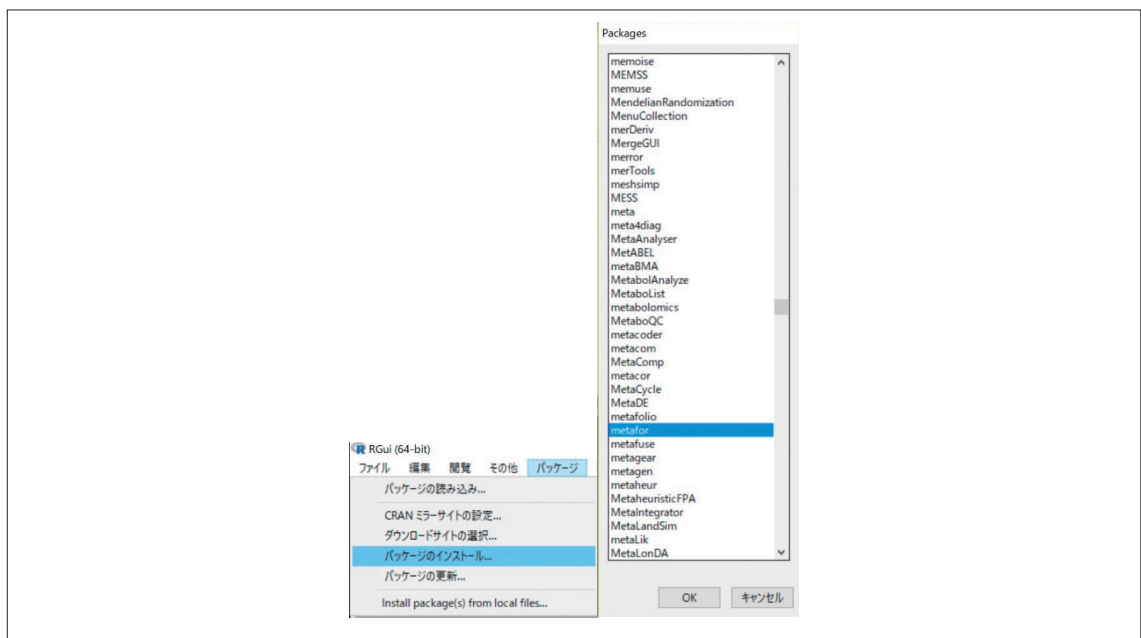
(2) パッケージのインストール

Rでは特定の解析を目的とした「パッケージ」と呼ばれるプログラムが世界中の開発者により提供されている。ここでは、メタアナリシスのために開発された metafor パッケージ (Viechtbauer W 2020) と、フォレストプロット解析のために開発された forestplot パッケージ (Gordon M 2020) を用いる。

Rを起動後、パッケージメニューから「CRAN ミラーサイトの設定」を選択し、一覧から「Japan (Tokyo) [https]」を選択して、[OK] をクリックする。



次に、「パッケージ」メニューから「パッケージのインストール」を選択し、表示されたパッケージの一覧から目的のパッケージを選択して、[OK] をクリックする。forestplot パッケージについても、同様の操作を行う。



通常、管理者権限で Windows を使用している場合は「Program Files」フォルダ>「R」フォルダ>「library」フォルダにパッケージがインストールされる。R の「パッケージ」メニューから「パッケージの読み込み」を選択するとインストール済みのパッケージを確認することができる。

(3) メタアナリシスのためのデータの準備

Excel で用意したデータが使用できるので、テンプレートなどの評価シートの一部を用いることができる。データのラベル名はそのまま、ラベルのカラムのテキストは内容に応じて変更する。効果指標の種類は略称によって判断してメタアナリシスを実行する。

アウトカムが二値変数で効果指標がリスク比 (RR) の場合の例を以下に示す。同じデータで「RR」を「RD」に、「Risk Ratio」を「Risk Difference」に変更すれば、リスク差 (RD) を効果指標としたメタアナリシスを行うことができる。同様に、それぞれ「OR」、「Odds Ratio」に変更すると、オッズ比 (OR) を効果指標としたメタアナリシスを行うことができる。

| アウトカム 治癒 | | | | | | |
|-----------------|------------|--------------|------------|--------------|-------------|-----------------------|
| 研究コード | 対照群 総例数 | アウトカム +例数 | 介入群 総例数 | アウトカム +例数 | ラベル | |
| author | nc | cpos | nt | tpos | label | この行のラベル名は変更不可 |
| Salminen P 2015 | 273 | 272 | 257 | 186 | Author Year | 研究ID |
| Vons C 2011 | 119 | 104 | 120 | 72 | Surgery | 対照 |
| Hansson J 2009 | 167 | 142 | 202 | 83 | Antibiotics | 介入 |
| Styrud J 2006 | 124 | 124 | 128 | 97 | Healing | アウトカム |
| Eriksson S 1995 | 20 | 20 | 20 | 13 | Risk Ratio | 効果指標 |
| | | | | | RR | 略称(RR, OR, RD, HR,MD) |
| | | | | | | 表示用の効果指標名 |
| | | | | | | 効果指標の種類を指定。略称のいずれか。 |

効果指標がハザード比 (HR) の場合の例を以下に示す。

| アウトカム 死亡(ハザード比) | | | | | | | |
|---------------------|------------|------------|-----------|---------|--------|--------------|-----------------------|
| 研究コード | 介入群 総例数 | 対照群 総例数 | 効果指標 値 | 標準誤差 | ラベル | | |
| author | nt | nc | yi | sei | label | | |
| Doffoel M 2008 | | 62 | 61 | -0.0726 | 0.1972 | Author Year | 研究 |
| Lo CM 2002 | | 40 | 40 | -0.6931 | 0.245 | BSC | 対照 |
| Llovet JM 2002 | | 40 | 35 | -0.755 | 0.4661 | TACE | 介入 |
| Pelletier G 1998 | | 37 | 36 | -0.0834 | 0.2659 | Death | アウトカム |
| Groupe d'Etude 1995 | | 50 | 46 | -0.3425 | 0.2303 | Hazard Ratio | 効果指標 |
| Pelletier G 1990 | | 21 | 21 | 0.3946 | 0.343 | HR | 略称(RR, OR, RD, HR,MD) |

アウトカムが連続変数で効果指標が平均値差 (MD) の場合の例を以下に示す。

| アウトカム 入院日数 | | | | | | | | |
|---------------|-----|------|------|-----|------|------|------------------|-----------------------|
| 研究コード | 対照群 | | | 介入群 | | | ラベル | |
| | 総例数 | 平均値 | 標準偏差 | 総例数 | 平均値 | 標準偏差 | | |
| author | n2i | m2i | sd2i | n1i | m1i | sd1i | label | |
| Eriksson 1995 | 20 | 3.4 | 1.9 | 20 | 3.1 | 3 | Author Year | 研究 |
| Styrud 2006 | 124 | 2.6 | 1.2 | 128 | 3 | 1.4 | Appendectomy | 対照 |
| Turhan 2009 | 183 | 2.4 | 0.14 | 107 | 3.14 | 0.1 | Antibiotics | 介入 |
| Vons 2011 | 119 | 3.04 | 1.5 | 120 | 3.96 | 4.87 | Days in hospital | アウトカム |
| | | | | | | | Mean Difference | 効果指標 |
| | | | | | | | MD | 略称(RR, OR, RD, HR,MD) |

(4) メタアナリシスの実行

インターネットに接続された状態で R を起動し、「ファイル」メニューから「新しいスクリプト」でエディター画面を開き、以下のスクリプトを書き込む、あるいは Excel のデータ画面からコピーして貼り付ける (Excel でデータを用意した場合) *²³。

* 23 あとで利用することを考えて、「ファイル」メニューから「保存」を選択して「metaclip.R」というファイル名で保存しておく。以降使用する際は「ファイル」メニューから「スクリプトを開く」で開いて用いる

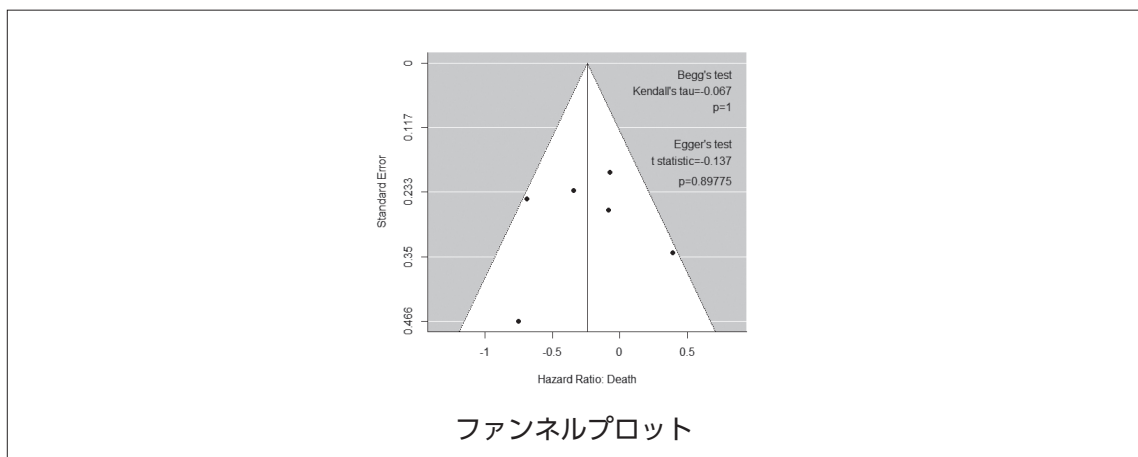
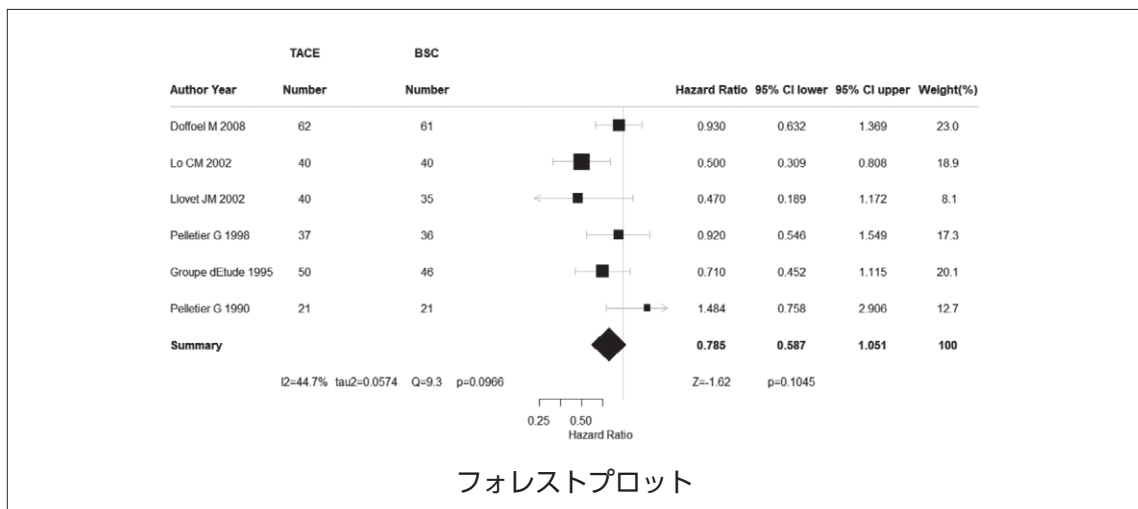
metaclip.R

```
#1. クリップボード経由でデータ読み込み：ヘッダーあり
exdat=read.delim("clipboard",sep="\t",header=TRUE) # Windows
exdat=read.delim(pipe("pbpaste"),sep="\t",header=TRUE) # Mac
#2. スクリプトを読み込んでメタアナリシスを実行。
source("http://zanet.biz/med/useRj/scripts/forest_metafor_b.R")
```

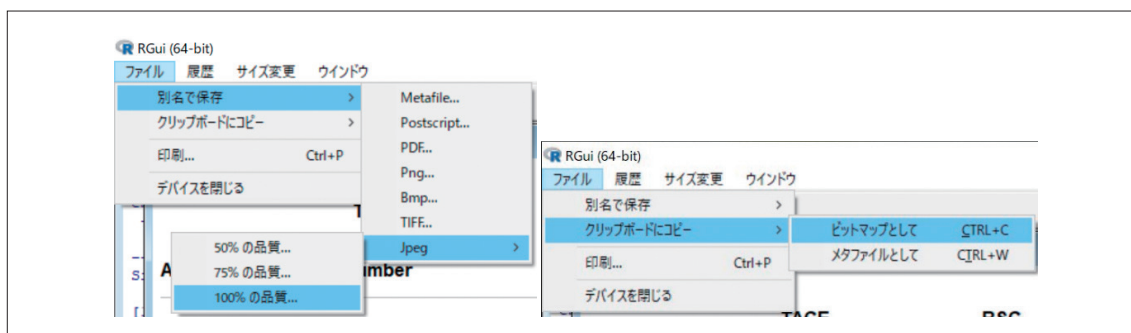
Excel を起動して、(3) **メタアナリシスのためのデータの準備** で用意したデータファイルを開き、青枠の範囲を選択してコピーする。

R に戻り、上記スクリプトの #1 を実行する。Windows の場合は、#Windows と書いてある行にカーソルを置き、実行ボタンをクリックする（または [Ctrl] + [R] キー）。Mac の場合は、#Mac と書いてある行にカーソルを置き、[Command] + [Enter] キーを押す。この操作により、データがクリップボードを経由して R に手渡される。

続いて、上記スクリプトの #2 を実行する。これは上記スクリプトの source(……) の行にカーソルを置くことで同様に実行でき、フォレストプロットとファンネルプロットが描画される。フォレストプロットはウィンドウの大きさを変更して、全体が見えるように調整する。



プロットを選択した状態で、「ファイル」メニューからファイルとして保存したり、「クリップボードにコピー」より Excel, Word, PowerPoint に貼り付けたりすることができる。



4.7 補足資料2：診断精度研究のメタアナリシス (RとmadaおよびOpenBUGS)

詳細は以下を参照されたい。

森實敏夫 他. (2015) 特別寄稿 5 診断に関する診療ガイドライン (CPG) の作成.
https://minds.jcqhc.or.jp/s/guidance_special_articles5_1 (参照 2021/01/28)

ここでは、参照基準が不完全で、インデックス診断法と相関があり、異なる参照基準が用いられている場合に対応する Dendukuri N らによる BUGS コードを提示する。

```
#Imperfect reference standard, covariance-adjusted (T1 dependent on T2).
#l: the number of studies, k: the number of references.
model {

  for(i in 1:l) {
    theta[i] ~ dnorm(THETA,prec[1])
    alpha[i] ~ dnorm(LAMBDA,prec[2])
    p[1,i] <- phi(-(theta[i] - 0.5*alpha[i])/exp(beta/2))
    p[2,i] <- phi(-(theta[i] + 0.5*alpha[i])*exp(beta/2))
    prob[i,1] <- pi[i]*( p[1,i] * s2[ref[i]] + covs12[i] ) + (1-pi[i])*( p[2,i] * (1-c2[ref[i]]) + covc12[i] )
    prob[i,2] <- pi[i]*( p[1,i] * (1-s2[ref[i]]) - covs12[i] ) + (1-pi[i])*( p[2,i] * c2[ref[i]] - covc12[i] )
    prob[i,3] <- pi[i]*( (1-p[1,i]) * s2[ref[i]] - covs12[i] ) + (1-pi[i])*( (1-p[2,i]) * (1-c2[ref[i]]) - covc12[i] )
    prob[i,4] <- pi[i]*( (1-p[1,i]) * (1-s2[ref[i]]) + covs12[i] ) + (1-pi[i])*( (1-p[2,i]) * c2[ref[i]] + covc12[i] )
    results[i,1:4] ~ dmulti(prob[i,1:4],n[i])
    n[i] <- sum(results[i,1:4])
    pi[i] ~ dbeta(1,1)
    se[i] <- p[1,i]
    sp[i] <- 1-p[2,i]
    alpha.covs12[i] <-1
    beta.covs12[i] <-1
    alpha.covc12[i] <-1
    beta.covc12[i] <-1
    us[i] <- min(se[i], s2[ref[i]]) - ( se[i]*s2[ref[i]] ) ;
    uc[i] <- min(sp[i], c2[ref[i]]) - ( sp[i]*c2[ref[i]] ) ;
    u.covs12[i] ~ dbeta(alpha.covs12[i],beta.covs12[i]) ;
    u.covc12[i] ~ dbeta(alpha.covc12[i],beta.covc12[i]) ;
    covs12[i] <- u.covs12[i]*us[i] ;
    covc12[i] <- u.covc12[i]*uc[i] ;
  }

  for(j in 1:2) {
    prec[j] <- pow(sigma[j],-2)
    sigma[j] ~ dunif(0,2)
  }

  THETA ~ dunif(-1.5,1.5)
  LAMBDA ~ dunif(-3,3)
  beta ~ dunif(-0.75,0.75)
  S_overall <- phi(-(THETA-LAMBDA/2)/exp(beta/2))
  C_overall <- phi( (THETA+LAMBDA/2)*exp(beta/2))
  theta_new ~ dnorm(THETA,prec[1])
}
```


4.8 補足資料3：ネットワークメタアナリシス

(1) OpenBUGS を用いるベイジアンアプローチによるコントラストベースモデルによるネットワークメタアナリシス

通常のメタアナリシスで行われる治療1と治療2の直接比較だけでなく、さらに治療3と間接比較する場合、治療1と治療3の間接比較と、治療2と治療3の間接比較ではそれぞれの治療効果 δ に対して次の分布を想定できる。 i は研究番号を表す。

$$\begin{aligned}\delta_{i,13} &\sim N(d_{13}, \sigma_{13}^2) \\ \delta_{i,23} &\sim N(d_{23}, \sigma_{23}^2)\end{aligned}$$

これらから、 $\delta_{i,23} = \delta_{i,13} - \delta_{i,12}$ が成立する場合、1つの研究内での共通の基準との効果の比較は互いに加算・減算できるということになり、移行性が成立するという。ネットワークメタアナリシスではこの移行性が成立することを前提としている。すなわち、治療1と治療3の差から治療1と治療2の差を除くと、治療2と治療3の差が得られることが正しいことを前提としている。パラメータである d についても同様の式が成立し、 $d_{23} = d_{13} - d_{12}$ となる。

分散については、1つの研究で得られた効果指標の間には相関があることを想定し、相関係数 ρ で調整が行われる。すなわち $\sigma_{23}^2 = \sigma_{12}^2 + \sigma_{13}^2 - \rho_{23}^{(1)} \sigma_{12}^2 \sigma_{13}^2$ となる。ここでは治療1を基準として設定している。

分散の評価における $\rho_{23}^{(1)}$ は、1つの研究における治療1と比較した治療3の効果と治療1と比較した治療2の効果の相関である。すなわち、1つの研究で3つの治療がRCTで比較され、そのうち1つ（治療1）がプラセボのような場合、対象者が全体として軽症者であれば、治療2、治療3のいずれの治療も効果が高めに出る可能性がある。逆に重症者が多ければ、効果が低めに出る可能性がある。したがって、治療2と治療3に相関があることを前提とする必要がある。

以下に紹介するBUGSのコードで用いるDias Sら（2011）のモデルでは、分散に関しては同じ値、すなわち $\sigma_{12}^2 = \sigma_{13}^2 = \sigma_{23}^2 = \sigma^2$ と見なし、全ての比較されるペアの間で $\sigma^2 = 0.5$ としている。

このモデルを適用することによって、治療1と治療2の比較あるいは治療1と治療3の比較が異なる研究で行われていた場合、間接的に治療2と治療3を比較することが可能になる。その場合にそれぞれの研究の交換可能性が前提となる。

また、間接的な比較で得られる効果については次の式が成立する必要がある、これらが成立する状態を一致性（consistency）またはコヒレンス（coherence）と呼ぶ。 s は治療番号を示す。

$$\begin{aligned}d_{23} &= d_{13} - d_{12} \\ d_{24} &= d_{14} - d_{12} \\ &\vdots \\ d_{(s-1),s} &= d_{1s} - d_{1,(s-1)}\end{aligned}$$

言い換えると、直接比較の結果と間接比較の結果の間に乖離がないということが必要な条件となる。実際には乖離することもあり得るので、何らかの方法、例えば後述するネットヒートプロット（net heat plot）やノード分割モデルプロットなどで点検が行われる。

k 種類の治療、すなわち治療 $1 \sim k$ を比較する場合についても適用可能な表記を用いてランダム効果モデルを記述すると、以下ようになる。 t_k は研究 i の治療 k という意味である。 σ^2 は研究間のばらつき、すなわち研究間の異質性を表す項である。

$$\delta_{i,1k} \sim N(d_{i,t_k}, \sigma^2)$$

さらに、4種類以上の治療を比較した臨床試験の結果を取り込んでネットワークメタアナリシスを行う際に Dias S らの述べているモデルは以下のとおりである。

$$\delta_i = \begin{bmatrix} \delta_{i,12} \\ \delta_{i,13} \\ \vdots \\ \delta_{i,1a_i} \end{bmatrix} \sim N_{a_i-1} \left[\begin{bmatrix} d_{i,t_2} \\ d_{i,t_3} \\ \vdots \\ d_{i,t_{a_i}} \end{bmatrix}, \begin{bmatrix} \sigma^2 & \frac{\sigma^2}{2} & \cdots & \frac{\sigma^2}{2} \\ \frac{\sigma^2}{2} & \sigma^2 & \cdots & \frac{\sigma^2}{2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma^2}{2} & \frac{\sigma^2}{2} & \cdots & \sigma^2 \end{bmatrix} \right]$$

研究 i における効果指標 δ_i は治療の種類分の値からなるベクトルで表され [a_i は治療の個数 (2, 3, ..., k)], それぞれ正規分布に従う。研究間の分散は同じ値であるとしている。

また、治療 2 から治療 $k - 1$ の効果指標が与えられた際の $\delta_{i,1k}$ の分布は次の式で示される。

$$\delta_{i,1k} \begin{bmatrix} \delta_{i,12} \\ \delta_{i,13} \\ \vdots \\ \delta_{i,1(k-1)} \end{bmatrix} \sim N \left[(d_{1,t_k} - d_{1,t_1}) + \frac{1}{k-1} \sum_{j=1}^{k-1} [\delta_{i,1j} - (d_{1,t_j} - d_{1,t_1})], \frac{1}{2(k-1)} \sigma^2 \right]$$

これら2つの式のいずれかを用いることによって、治療間の相関を取り込んだランダム効果モデルによるネットワークメタアナリシスが可能になる。

実際の BUGS 用のコードは、以下のとおりである (Dias S 2011: Appendix)。

```

####Network meta-analysis with logit link regression, random-effects model###
model
{
for(i in 1:ns)          #ns は研究数
{
w[i,1]<-0              # 対照のアームの調整は 0 に設定。
delta[i,1]<-0          # 対照のアームの治療効果は 0 に設定。
mu[i]~dnorm(0,0.0001)  # ベースライン (対照の治療効果) のあいまい事前分布の設定。
for(k in 1:na[i])      # 研究 i の治療アーム数だけ以下を繰り返す。
{
r[i,k]~dbin(p[i,k],n[i,k]) # イベント数 (アウトカム+例数) は二項分布に従う。二項分布尤度。
logit(p[i,k])<-mu[i]+delta[i,k] # 直線回帰のロジットモデル。k=1 つまり対照の場合 delta[i,k]=0 となる。
rhat[i,k]<-p[i,k]*n[i,k]      # イベント数の期待値 (Deviance 算出に必要)。
dev[i,k]<-2*(r[i,k]*(log(r[i,k])-log(rhat[i,k]))+(n[i,k]-r[i,k])*(log(n[i,k]-r[i,k])-log(n[i,k]-rhat[i,k]))) #Deviance 要素算出
}
}
resdev[i]<-sum(dev[i,1:na[i]]) # 研究 i の Deviance 合計。
for(k in 2:na[i])          # 治療 1 を対照に、治療アーム分繰り返す。
{
delta[i,k]~dnorm(md[i,k],taud[i,k]) # 研究 i の  $\delta$  は  $\ln(\text{OR})$  の平均値 md, 分散 taud の正規分布よりサンプリング。
md[i,k]<-d[t[i,k]]-d[t[i,1]]+sw[i,k] # アーム数で調整した  $\ln(\text{OR})$  の平均値
taud[i,k]<-tau*2*(k-1)/k # 分散  $\sigma^2 * 1/[2(k-1)]$  の逆数 precision を計算。
w[i,k]<-(delta[i,k]-d[t[i,k]]+d[t[i,1]]) # アーム数の調整分の計算
sw[i,k]<-sum(w[i,1:k-1])/(k-1) # 調整分の合計の計算
}
}
totresdev<-sum(resdev[]) #Total residual deviance
d[1]<-0                    # 参照アームの効果を 0 に設定。
for(k in 2:nt)
{
d[k]~dnorm(0,0.0001)      # 治療効果のあいまい事前分布の設定。
}
sd~dunif(0,5)             # 研究間分布の標準偏差にあいまい事前分布の設定。
tau<-pow(sd,-2)           # 標準偏差から分散の逆数を求め tau に代入する。
####sampling effect measures OR, RR, RD, NNT
for (c in 1:(nt-1)) {
for (k in (c+1):nt) {
OR[c,k] <- exp(d[k] - d[c]) # 統合値のオッズ比をサンプリング。
lor[c,k] <- d[k]-d[c]      # オッズ比の自然対数をサンプリング。
RR[c,k]<-exp(d[k])*(1+exp(d[c]))/exp(d[c])/(1+exp(d[k])) # リスク比のサンプリング。
RD[c,k] <-exp(d[k])/(1+exp(d[k]))-exp(d[c])/(1+exp(d[c]))# リスク差のサンプリング。
NNT[c,k]<-1/abs(RD[c,k])   #NNT のサンプリング。
}
}
}
###
# ranking on relative scale
for (k in 1:nt)
{
rk[k] <- nt+1-rank(d[],k) # アウトカムがよい事象の場合。
#rk[k] <- rank(d[],k)    # アウトカムが悪い事象の場合。
best[k] <- equals(rk[k],1) # 治療 k が最善である確率。
for (h in 1:nt){
prob[h,k] <- equals(rk[k],h) # 治療 k が h 番目によい治療である確率。
}
}

```

```

}
###
}

#DATA
list(ns=24,nt=4)
#smokingcessation data for BUGS
r[,1]  n[,1]  r[,2]  n[,2]  r[,3]  n[,3]  t[,1]  t[,2]  t[,3]  na[]  #  study
9      140   23    140   10    138   1      3     4     3     #   1
11     78    12    85    29    170   2      3     4     3     #   2
75     731   363   714   NA    NA    1      3     NA    2     #   3
2      106   9      205   NA    NA    1      3     NA    2     #   4
58     549   237   1561  NA    NA    1      3     NA    2     #   5
0      33     9      48    NA    NA    1      3     NA    2     #   6
3      100   31     98    NA    NA    1      3     NA    2     #   7
1      31     26     95    NA    NA    1      3     NA    2     #   8
6      39     17     77    NA    NA    1      3     NA    2     #   9
79     702   77     694   NA    NA    1      2     NA    2     #  10
18     671   21     535   NA    NA    1      2     NA    2     #  11
64     642   107    761   NA    NA    1      3     NA    2     #  12
5      62     8      90    NA    NA    1      3     NA    2     #  13
20     234   34     237   NA    NA    1      3     NA    2     #  14
0      20     9      20    NA    NA    1      4     NA    2     #  15
8      116   19     149   NA    NA    1      2     NA    2     #  16
95     1107  143    1031  NA    NA    1      3     NA    2     #  17
15     187   36     504   NA    NA    1      3     NA    2     #  18
78     584   73     675   NA    NA    1      3     NA    2     #  19
69     1177  54     888   NA    NA    1      3     NA    2     #  20
20     49    16     43    NA    NA    2      3     NA    2     #  21
7      66     32     127   NA    NA    2      4     NA    2     #  22
12     76    20     74    NA    NA    3      4     NA    2     #  23
9      55     3      26    NA    NA    3      4     NA    2     #  24

END

#INITS
list(d=c(NA,0,0,0),sd=1,mu=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0))
list(d=c(NA,-1,-1,-1),sd=4,mu=c(-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3))
list(d=c(NA,2,2,2),sd=2,mu=c(-3,5,-1,-3,7,-3,-4,-3,-3,0,-3,-3,0,3,5,-3,-3,-1,-3,-7,-3,-3,3))

```

出典：Dias, S., Welton, N.J., Sutton, A.J. & Ades, A.E. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. 2011; last updated September 2016; available from <http://www.nicedsu.org.uk>. p.59 より作成

なお、将来同じような研究が行われた場合のオッズ比の予測分布を得たい場合には、

```
lor[c,k] <- d[k]-d[c]    # オッズ比の自然対数をサンプリング
```

の行に続いて以下の行を追加し、nodeとしてlornewとORnewを設定する。

```
ornew[c,k]~dnorm(lor[c,k],taud)
ORnew[c,k]~exp(lornew[c,k])
```

(2) BUGS 用のデータの準備について

BUGS でコードを走らせる際には、「Model」メニューから「Specification…」を選択し、「Specification Tool」で「check model」, 「load data」, 「compile」, 「load inits」, 「gen inits」の順に操作する。データは Excel で用意して必要な範囲を選択してコピーし、BUGS の「Edit」メニューから「Paste Special」を選び、貼り付ける形式で「Plain Text」を選択して [OK] ボタンをクリックする。このようにすることで、Excel から直接コピー&ペーストでデータを取り込める。

例えば、Excel で次のようなデータを用意したとする。

| | A | B | C | D | |
|---|------|------|------|------|----|
| 1 | rc[] | nc[] | rt[] | nt[] | |
| 2 | | 4 | 11 | 5 | 13 |
| 3 | | 3 | 14 | 1 | 13 |
| 4 | | 14 | 19 | 2 | 17 |
| 5 | | 1 | 12 | 0 | 13 |
| 6 | | 9 | 13 | 7 | 17 |
| 7 | | 13 | 47 | 18 | 48 |
| 8 | END | | | | |

セル A1 から D8 の範囲を選択してコピーし、BUGS のコード画面に貼り付けることができる。

```
DATA #k is the number of studies.
list(k=6, rc=c(4,3,14,1,9,13),nc=c(11,14,19,12,13,47),rt=c(5,1,2,0,7,18),nt=c(13,13,17,13,17,48))

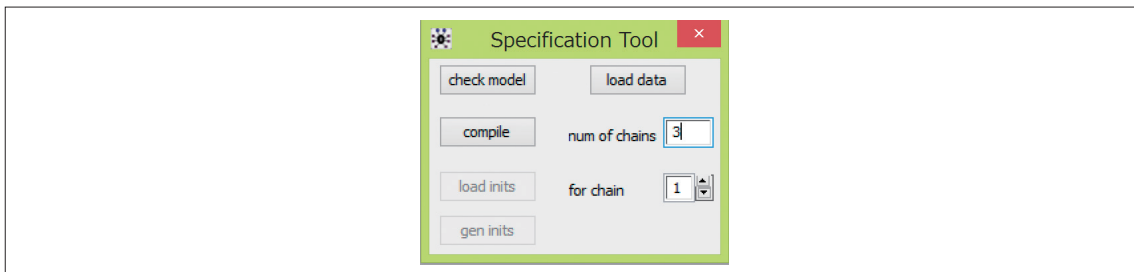
list(k=6)
rc[] nc[] rt[] nt[]
4 11 5 13
3 14 1 13
14 19 2 17
1 12 0 13
9 13 7 17
13 47 18 48
END
```

さらに、list(k=6) を書き加えたのが上記の例である。ここには2種類のデータの記述法を示している。上部の記述法の場合は、list を反転させて [load data] をクリックする。下部の記述法の場合は、list を反転させて [load data] をクリックし、次に rc[] から nt[] の行を選択して再度 [load data] をクリックする。データとして読み込まれるものはいずれの操作の場合も同じである。例えば、rc[2] で指定される値は3となりいずれの場合も同じになる。

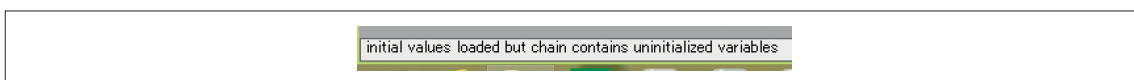
このように、配列のデータを作成する場合に、列名を変数名として用いることによって、Excel で用意したデータを利用することができる。

なお、上記のネットワークメタアナリシスの解析データは後述する netmeta パッケージのデータセット「smokingcessation」と同じものである。R でいったん読み込んで、write.csv() 関数でファイルとして書き出し、それを Excel で開いて、形式を整えてから、同様な手順で用意した。

さて、それではこのデータに対してネットワークメタアナリシスを実行してみよう。この例では、初期値を3セット用意してあり、3チェーン実行させる。「check model」, 「load data」の実行後、「num of chains」を3に変更し、「compile」をクリックする。



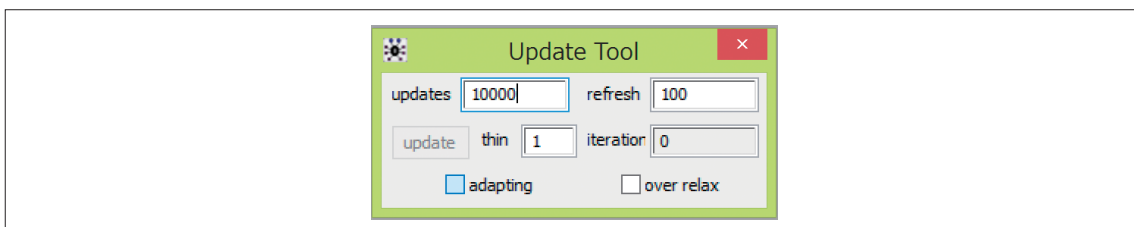
次に、#INITS の下の初期値を順に設定する。最初の list を選択して、「load inits」ボタンをクリックする。すると、「num of chains」が1から2に変わる。次の list を選択して同じことを繰り返す。3種類の初期値の「load inits」が終了すると「num of chains」が1に戻る。その際に、一部の変数の初期値が設定されていないというメッセージがウィンドウの下枠に表示されるので、「gen inits」ボタンをクリックする。



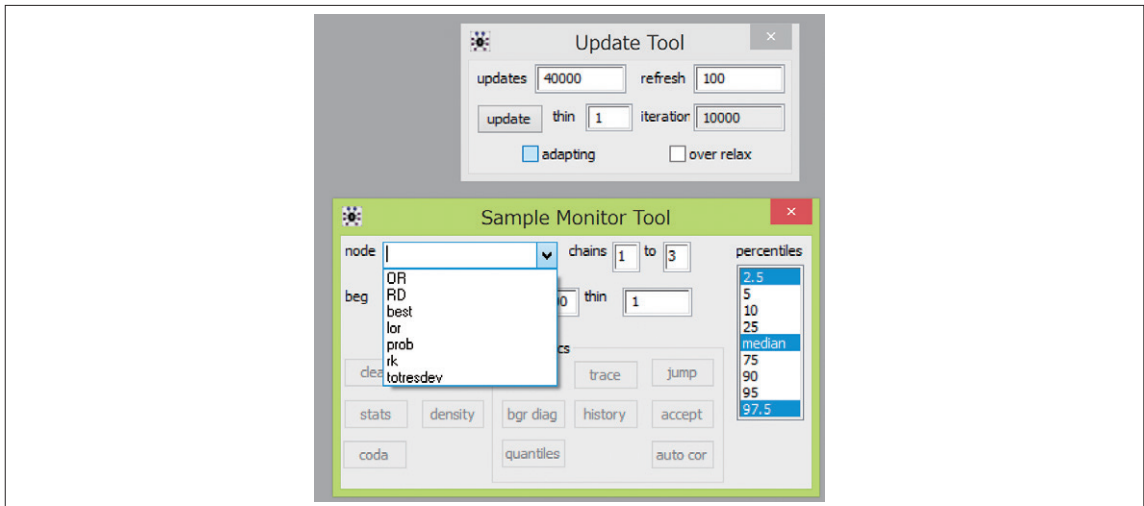
すると、メッセージが以下のように変わる。



これでMCMC (Markov chain Monte Carlo: マルコフ連鎖モンテカルロ) を走らせる準備ができた。「Model」メニューから「Update…」を選択し「Update Tool」を表示させ、「updates」を「10000」に設定して、burn-in を1万回実行する。



次に、「Inferences」メニューから「Samples…」を選択し「Sample Monitor Tool」を表示させ、「node」にモニター（記録）したい変数名を入力して「set」ボタンをクリックする。必要なものについてこれを繰り返す。例えば、「OR」, 「lor」, 「RD」, 「totresdev」, 「rk」, 「best」, 「prob」などを必要に応じて設定する。「node」の設定が終わったら、「Update Tool」で「updates」を「40000」に設定して「update」ボタンをクリックする。この時点では「Sample Monitor Tool」で「chains」が「1 to 3」に設定されているのが分かる。



「Update Tool」の「iteration」の値が「50000」に達するとMCMCが停止する。

「Sample Monitor Tool」で「node」に「*」を入力して「stats」ボタンをクリックすると次の結果が表示される。

| | mean | sd | MC_error | val2.5pc | median | val97.5pc | start | sample |
|-----------|----------|----------|----------|----------|---------|-----------|-------|--------|
| OR[1,2] | 1.772 | 0.7805 | 0.003913 | 0.7476 | 1.623 | 3.66 | 10001 | 120000 |
| OR[1,3] | 2.391 | 0.6026 | 0.003552 | 1.474 | 2.304 | 3.812 | 10001 | 120000 |
| OR[1,4] | 3.311 | 1.642 | 0.009828 | 1.298 | 2.963 | 7.36 | 10001 | 120000 |
| OR[2,3] | 1.55 | 0.6875 | 0.002731 | 0.6319 | 1.421 | 3.248 | 10001 | 120000 |
| OR[2,4] | 2.073 | 1.129 | 0.005164 | 0.7133 | 1.827 | 4.893 | 10001 | 120000 |
| OR[3,4] | 1.409 | 0.6359 | 0.003022 | 0.5726 | 1.285 | 2.968 | 10001 | 120000 |
| RD[1,2] | 0.1157 | 0.09116 | 4.523E-4 | -0.07221 | 0.1188 | 0.2854 | 10001 | 120000 |
| RD[1,3] | 0.1966 | 0.04965 | 2.803E-4 | 0.09583 | 0.1974 | 0.2922 | 10001 | 120000 |
| RD[1,4] | 0.2413 | 0.08086 | 4.774E-4 | 0.06482 | 0.2477 | 0.3804 | 10001 | 120000 |
| RD[2,3] | 0.08088 | 0.09221 | 3.794E-4 | -0.09219 | 0.0779 | 0.2724 | 10001 | 120000 |
| RD[2,4] | 0.1256 | 0.1002 | 4.387E-4 | -0.07149 | 0.1253 | 0.3247 | 10001 | 120000 |
| RD[3,4] | 0.04471 | 0.07826 | 3.579E-4 | -0.1231 | 0.04947 | 0.1861 | 10001 | 120000 |
| best[1] | 3.333E-5 | 0.005773 | 1.662E-5 | 0.0 | 0.0 | 0.0 | 10001 | 120000 |
| best[2] | 0.05717 | 0.2322 | 7.834E-4 | 0.0 | 0.0 | 1.0 | 10001 | 120000 |
| best[3] | 0.2381 | 0.4259 | 0.00165 | 0.0 | 0.0 | 1.0 | 10001 | 120000 |
| best[4] | 0.7047 | 0.4562 | 0.0018 | 0.0 | 1.0 | 1.0 | 10001 | 120000 |
| lor[1,2] | 0.4896 | 0.4016 | 0.002007 | -0.2909 | 0.4843 | 1.297 | 10001 | 120000 |
| lor[1,3] | 0.8424 | 0.2397 | 0.001375 | 0.3881 | 0.8348 | 1.338 | 10001 | 120000 |
| lor[1,4] | 1.098 | 0.4393 | 0.002655 | 0.2608 | 1.086 | 1.996 | 10001 | 120000 |
| lor[2,3] | 0.3528 | 0.4117 | 0.001678 | -0.459 | 0.3512 | 1.178 | 10001 | 120000 |
| lor[2,4] | 0.6082 | 0.4866 | 0.002235 | -0.3378 | 0.6027 | 1.588 | 10001 | 120000 |
| lor[3,4] | 0.2554 | 0.4158 | 0.001989 | -0.5576 | 0.251 | 1.088 | 10001 | 120000 |
| prob[1,1] | 3.333E-5 | 0.005773 | 1.662E-5 | 0.0 | 0.0 | 0.0 | 10001 | 120000 |
| prob[1,2] | 0.05717 | 0.2322 | 7.834E-4 | 0.0 | 0.0 | 1.0 | 10001 | 120000 |
| prob[1,3] | 0.2381 | 0.4259 | 0.00165 | 0.0 | 0.0 | 1.0 | 10001 | 120000 |
| prob[1,4] | 0.7047 | 0.4562 | 0.0018 | 0.0 | 1.0 | 1.0 | 10001 | 120000 |
| prob[2,1] | 0.002383 | 0.04876 | 1.533E-4 | 0.0 | 0.0 | 0.0 | 10001 | 120000 |
| prob[2,2] | 0.1723 | 0.3777 | 0.001169 | 0.0 | 0.0 | 1.0 | 10001 | 120000 |
| prob[2,3] | 0.6001 | 0.4899 | 0.001685 | 0.0 | 1.0 | 1.0 | 10001 | 120000 |
| prob[2,4] | 0.2251 | 0.4177 | 0.001472 | 0.0 | 0.0 | 1.0 | 10001 | 120000 |
| prob[3,1] | 0.1055 | 0.3072 | 0.001133 | 0.0 | 0.0 | 1.0 | 10001 | 120000 |
| prob[3,2] | 0.6671 | 0.4712 | 0.001501 | 0.0 | 1.0 | 1.0 | 10001 | 120000 |
| prob[3,3] | 0.1616 | 0.3681 | 0.001288 | 0.0 | 0.0 | 1.0 | 10001 | 120000 |
| prob[3,4] | 0.06575 | 0.2478 | 8.077E-4 | 0.0 | 0.0 | 1.0 | 10001 | 120000 |
| prob[4,1] | 0.8921 | 0.3103 | 0.001159 | 0.0 | 1.0 | 1.0 | 10001 | 120000 |
| prob[4,2] | 0.1033 | 0.3044 | 0.001126 | 0.0 | 0.0 | 1.0 | 10001 | 120000 |
| prob[4,3] | 1.667E-4 | 0.01291 | 3.667E-5 | 0.0 | 0.0 | 0.0 | 10001 | 120000 |
| prob[4,4] | 0.004408 | 0.06625 | 2.006E-4 | 0.0 | 0.0 | 0.0 | 10001 | 120000 |
| rk[1] | 3.89 | 0.3212 | 0.001207 | 3.0 | 4.0 | 4.0 | 10001 | 120000 |
| rk[2] | 2.817 | 0.6861 | 0.002631 | 1.0 | 3.0 | 4.0 | 10001 | 120000 |
| rk[3] | 1.924 | 0.6281 | 0.002438 | 1.0 | 2.0 | 3.0 | 10001 | 120000 |
| rk[4] | 1.37 | 0.6253 | 0.002455 | 1.0 | 1.0 | 3.0 | 10001 | 120000 |
| toresdev | 54.06 | 10.15 | 0.05225 | 35.95 | 53.47 | 75.61 | 10001 | 120000 |

「Node statistics」を見ると sample が 120000 になっており、4 万回の MCMC が初期値を変えて 3 チェーン分実行されたことが分かる。

ここでは実際の結果は示さないが、「Sample Monitor Tool」で「density」、「history」および「auto cor」も見ることができる。

「best」の結果を見ると、best[4] の mean が 0.7047 となっている。これは、治療 4 が最も効果が高い可能性が高く、その確率が 0.7047 であることを示している。また、rk の結果を見ると median が rk[4] が 1、rk[3] が 2、rk[2] が 3、rk[1] が 4 で、治療効果がこの順で高いことが分かる。また prob の項はそれぞれの治療が特定の順位になる確率を示しており、例えば prob[1,4] は治療 4 が 1 番である確率であり、mean を見ると 0.7047 で best[4] と同じ値になっていることが分かる。

(3) gemtc パッケージによるネットワークメタアナリシス

gemtc パッケージを使用する際は、R とは別に JAGS をインストールしておく必要がある。さらに、R の rjags パッケージと gemtc パッケージをインストールする。gemtc パッケージはコントラストベースモデルを用いたベイジアンの手法によるネットワークメタアナリシスを実行するパッケージである。

次に、Excel で次のような形式でデータを用意する (smokingcessation のデータ)。以下の例では、4 種類の治療のうち 1 つの RCT で直接比較されたのは 3 種類であり、「1」が対照として用いられている。

| r[,1] | n[,1] | r[,2] | n[,2] | r[,3] | n[,3] | t[,1] | t[,2] | t[,3] | na[] |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| 9 | 140 | 23 | 140 | 10 | 138 | 1 | 3 | 4 | 3 |
| 11 | 78 | 12 | 85 | 29 | 170 | 2 | 3 | 4 | 3 |
| (以下省略) | | | | | | | | | |

R のスクリプトは次のとおりである。データを Excel で用意し、データの範囲をラベルの行も含めてコピーし、R でスクリプトの全てを選択して実行する。


```

####Network meta-analysis with gemtc of Bayesian model####
#Using JAGS, rjags, Rcpp#####
library("rjags")
library("gemtc")

#Read and convert BUGS type (eg. smoking cessation) data.
if(colnames(exdat)[2]=="y..1."||colnames(exdat)[1]=="y..1."){dtype="continuous"}
if(colnames(exdat)[2]=="r..1."||colnames(exdat)[1]=="r..1."){dtype="binary"}
dtype

#Binary data:
if(dtype=="binary")
{
  gdat=mtc.data.studyrow(exdat,armVars=c('treatment'='t', 'sampleSize'='n', 'responders'='r'))
  lh="binom";lk="logit"          #responders, sampleSize -->> Odds Ratio (default)
}
#Continuous data:
if(dtype=="continuous")
{
  gdat=mtc.data.studyrow(exdat,armVars=c('sampleSize'='n', 'treatment'='t', 'mean'='y', 'std.dev'='sd'))
  lh="normal";lk="identity"     #mean, std.dev, sampleSize OR mean, std.err -->> Mean difference
}
#----->>>>>>>>>> To gemtc.
burnin=2000
cycl=20000
thn=1

netw=mtc.network(data.ab=gdat, treatments=NULL,description="Network",data.re=NULL,studies=NULL,data=NULL)
summary(netw)
plot(netw)

#lh="normal";lk="identity"     #mean, std.dev, sampleSize OR mean, std.err -->> Mean difference
#lh="binom";lk="log"          #responders, sampleSize -->> Risk Ratio
#lh="binom";lk="logit"       #responders, sampleSize -->> Odds Ratio (default)
#lh="binom";lk="cloglog"     #responders (for the same peiord), sampleSize -->> Hazard Ratio
#lh="poisson";lk="log"       #responders, exposure -->> Hazard Ratio

model=mtc.model(netw,type="consistency",likelihood=lh, link=lk, linearModel="random")#Random effects model

#Run MCMC for the model with JAGS via rjags. n.adapt=burn-in
res=mtc.run(model, sampler=NA, n.adapt=burnin,n.iter=cycl,thin=thn)
summary(res)

#plot(res)
dev.new();forest(res)

rank.p=rank.probability(res,preferredDirection=1,covariate=NA)
#rank.p=rank.probability(res,preferredDirection=-1,covariate=NA)
print(rank.p)
dev.new();plot(rank.p)

relative.e=relative.effect.table(res,covariate=NA)
print(relative.e)

gelman.diag(res)
mtc.deviance(res)

```

```

res$dic #DIC

#####
#Node-splitting analysis:
res.ns=mtc.nodesplit(netw,likelihood=lh, link=lk, linearModel="random",n.adapt=burnin,n.iter=cycl,thin=thn)
#dev.new();plot(res.ns)
summary.ns=summary(res.ns)
summary.ns
dev.new();plot(summary.ns)

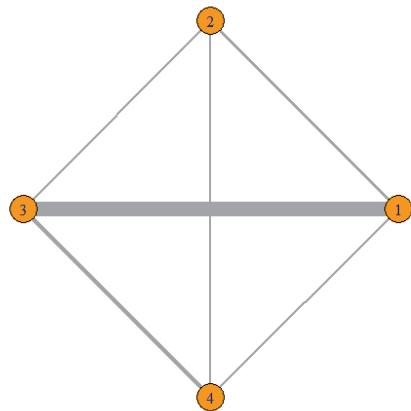
###SUCRA plot function modified from the one of Salanati's script.
sucra = function(effectiveness)
{
  # effectiveness: rank.p from GeMTC output.
  #Creates cumulative ranking curves extrapolated at the middle of each bar

  names <- names(effectiveness)
  nr.of.treat <- dim(effectiveness)[2]
  pn=round(sqrt(nr.of.treat))
  cumeffectiveness <- apply(effectiveness, 2, cumsum)

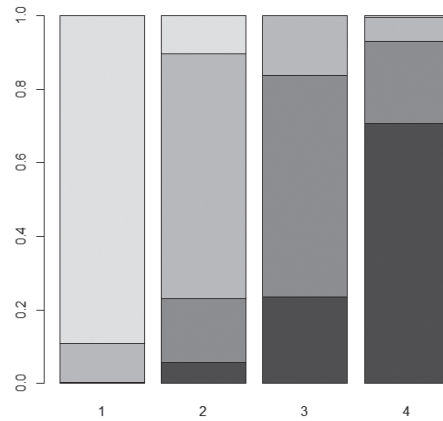
  dev.new()
  par(mfrow = c(pn,pn))
  for(i in 1:nr.of.treat) {
    plot(1:nr.of.treat, type = "n", ylim = c(0, 1), xlab =
      paste("Rank of", as.character(names[i])), ylab =
      "Cumulative Probability",xaxp=c(1,nr.of.treat,nr.of.treat-1))
    y=cumeffectiveness[c(1, 1:c(nr.of.treat - 1), c(nr.of.treat - 1)), i]
    x=c(1, c(1:c(nr.of.treat - 1) + 0.5, nr.of.treat)
    lines(lwd = 2, x, y,type="l")
  }
  dev.new()
  par(mfrow = c(pn,pn))
  for(i in 1:nr.of.treat) {
    plot(1:nr.of.treat, type = "n", ylim = c(0, 1), xlab =
      paste("Rank of", as.character(names[i])), ylab =
      "Probability",xaxp=c(1,nr.of.treat,nr.of.treat-1))
    x=c(1, c(1:c(nr.of.treat - 1) + 0.5, nr.of.treat)
    y=effectiveness[c(1, 1:c(nr.of.treat - 1), c(nr.of.treat - 1)), i]
    lines(lwd = 2, x, y,type="l")
  }
}
names(rank.p)=row.names(rank.p)
sucra(rank.p)
#####End to the script#####

```

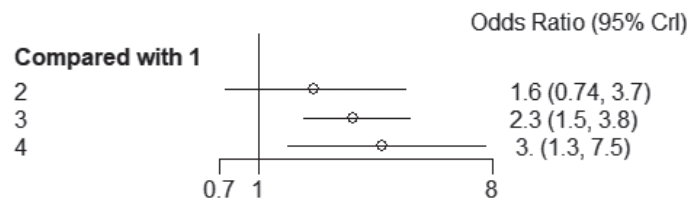
結果として、ネットワークジオメトリプロット、フォレストプロット、順位グラフ、ノード分割モデルプロット、SUCRA プロットなどが得られる。



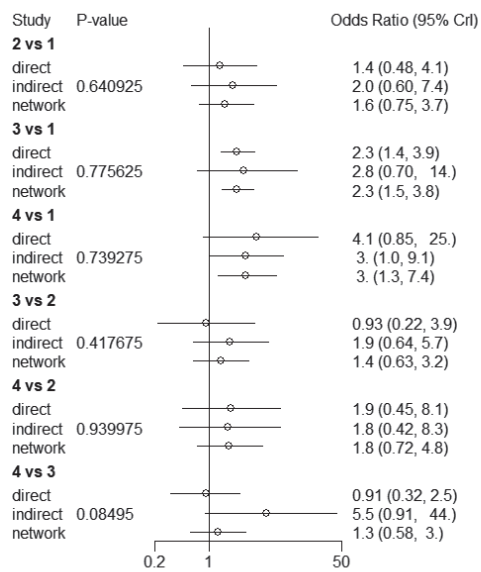
ネットワークジオメトリプロット



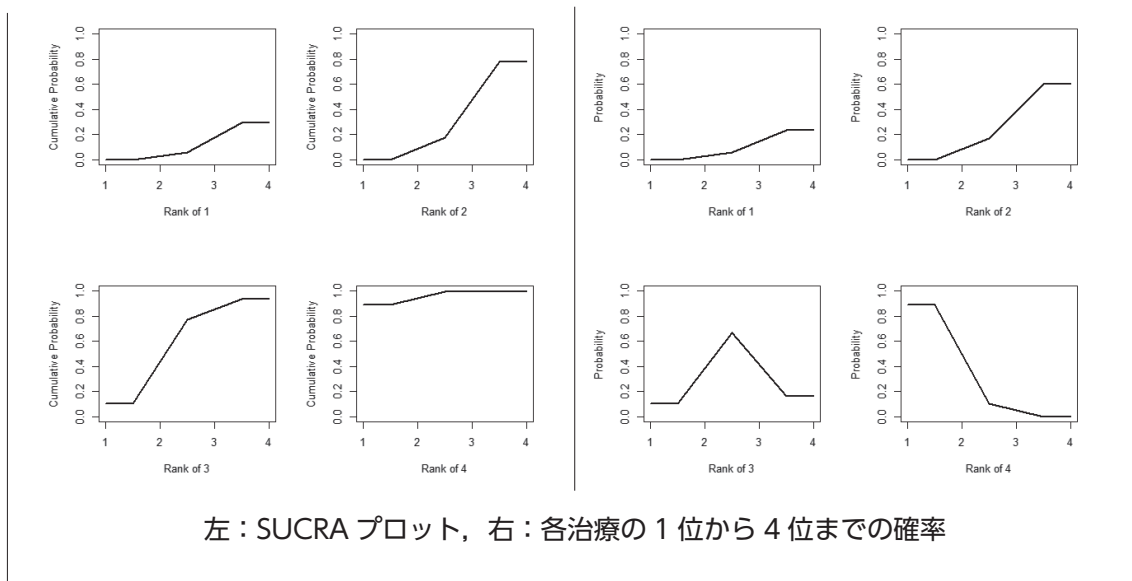
順位グラフ



フォレストプロット



ノード分割モデルプロット



この結果からは、治療4と3のペア比較でインコヒレンスが疑われるが、間接比較の信頼区間が大きく、ネットワーク推定値への貢献度は低いと考えられる。順位からは治療4が最善である確率が0.7程度で最も高いことが分かる。

上記の例は効果指標としてオッズ比を用いているが、gemtcパッケージはmtc.model関数の引数likelihoodの設定を変えることで、連続変数、リスク比、ハザード比などにも対応できるようになっている。スクリプト中の変数lhの設定を変更する。

連続変数の場合のデータの形式は以下ようになる。この例では、治療が3種類あり、1～3の番号で示されている。nは症例数、yは平均値、sdは標準偏差である。10番目の研究では3種類の介入が直接比較されているが、それ以外では治療3の群がないため、n[,3], y[,3], sd[,3]がNAとなっている。このような形式のデータをExcelなどで用意すれば、その後の操作は前述と同様である。

| n[,1] | y[,1] | sd[,1] | t[,1] | n[,2] | y[,2] | sd[,2] | t[,2] | n[,3] | y[,3] | sd[,3] | t[,3] | na[] |
|-------|--------|----------|-------|-------|--------|----------|-------|-------|-------|--------|-------|------|
| 70 | -10.26 | 13.43 | 2 | 66 | -4.88 | 11.64 | 1 | 1 | NA | 0 | NA | 2 |
| 54 | -14.78 | 12.49 | 2 | 56 | -8.13 | 12.72 | 1 | 1 | NA | 0 | NA | 2 |
| 125 | -13.36 | 8.78 | 2 | 123 | -10.39 | 10.45 | 3 | 1 | NA | 0 | NA | 2 |
| 220 | -13.11 | 8.53 | 2 | 114 | -9.1 | 9.36 | 1 | 1 | NA | 0 | NA | 2 |
| 55 | -16.6 | 12.49 | 2 | 60 | -14.9 | 11.39 | 3 | 1 | NA | 0 | NA | 2 |
| 58 | -15.49 | 8.148877 | 2 | 59 | -15.25 | 8.372449 | 1 | 1 | NA | 0 | NA | 2 |
| 187 | -11.5 | 10.9 | 3 | 177 | -9 | 10.9 | 1 | 1 | NA | 0 | NA | 2 |
| 144 | -10.1 | 10.9 | 3 | 78 | -8.7 | 10.9 | 1 | 1 | NA | 0 | NA | 2 |
| 100 | -11.71 | 7 | 2 | 101 | -8.97 | 6.934414 | 1 | 1 | NA | 0 | NA | 2 |
| 201 | -9.4 | 8.506468 | 2 | 186 | -8.2 | 8.455673 | 3 | 99 | -7.4 | 7.9599 | 1 | 3 |

(4) OpenBUGS を用いるベイジアンアプローチによるアームベースモデルによるネットワークメタアナリシス

ここでは、二値変数アウトカムの場合の BUGS コードとデータの例を示す。データのフォーマットは1行が1治療アームとなる。サンプルデータは (3) 項の smokingcessation と同じものでアウトカムは1年後の禁煙であり、治療数は4種類ある。治療1が参照治療である。変数 `probt` が各治療群の絶対リスク、各ペア比較のオッズ比は `OR`、リスク比は `RR`、リスク差は `RD` で得られる。`best` は各治療が最善である確率、`rk` はそれぞれの順位になる治療番号である。

```
#Zhang J, et al: Network meta-analysis of randomized clinical trials: reporting the proper summaries. Clin Trials 2014;11:246-62.
PMID: 24096635
#Arm-based model.
model {
  for(i in 1:tS) {
    p[i]~phi(mu[t[i]]+ vi[s[i], t[i]]) # model
    r[i]~dbin(p[i], totaln[i]) # binomial likelihood
  }
  for(j in 1:sN){
    vi[j, 1:tN]~dmnorm(mn[1:tN], T[1:tN,1:tN]) # multivariate normal distribution
  }
  invT[1:tN, 1:tN]<-inverse(T[,,])
  for (j in 1:tN){
    mu[j]~dnorm(0, 0.001)
    sigma[j]~sqrt(invT[j, j])
    probt[j]~phi(mu[j]/sqrt(1+invT[j, j]))
    # population-averaged treatment specific event rate
  }
  T[1:tN,1:tN]~dwish(R[1:tN, 1:tN], tN) # Wishart prior
  for (k in 1:tN) {
    rk[k]~tN + 1 - rank(probt[,k]) # ranking
    best[k]~equals(rk[k],1) # prob {treatment k is best}
  }
  for (j in 1:tN){ # calculation of RR, RD and OR
    for (k in (j+1):tN){
      RR[j, k] <- probt[k]/probt[j]
      RD[j, k] <- probt[k]-probt[j]
      OR[j, k] <- probt[k]/(1-probt[k])/probt[j]*(1-probt[j])
    }
  }
}
#DATA
list(tN=4,sN=24,tS=50,mn=c(0, 0, 0, 0),R=structure(.Data=c(1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1),.Dim=c(4,4)))
s[]      t[]      r[]      totaln[]
1         1         9         140
1         3         23        140
1         4         10        138
2         2         11         78
2         3         12         85
2         4         29        170
3         1         75        731
3         3         363       714
4         1         2         106
4         3         9         205
```

| | | | |
|-----|---|-----|------|
| 5 | 1 | 58 | 549 |
| 5 | 3 | 237 | 1561 |
| 6 | 1 | 0 | 33 |
| 6 | 3 | 9 | 48 |
| 7 | 1 | 3 | 100 |
| 7 | 3 | 31 | 98 |
| 8 | 1 | 1 | 31 |
| 8 | 3 | 26 | 95 |
| 9 | 1 | 6 | 39 |
| 9 | 3 | 17 | 77 |
| 10 | 1 | 79 | 702 |
| 10 | 2 | 77 | 694 |
| 11 | 1 | 18 | 671 |
| 11 | 2 | 21 | 535 |
| 12 | 1 | 64 | 642 |
| 12 | 3 | 107 | 761 |
| 13 | 1 | 5 | 62 |
| 13 | 3 | 8 | 90 |
| 14 | 1 | 20 | 234 |
| 14 | 3 | 34 | 237 |
| 15 | 1 | 0 | 20 |
| 15 | 4 | 9 | 20 |
| 16 | 1 | 8 | 116 |
| 16 | 2 | 19 | 146 |
| 17 | 1 | 95 | 1107 |
| 17 | 3 | 134 | 1031 |
| 18 | 1 | 15 | 187 |
| 18 | 3 | 35 | 504 |
| 19 | 1 | 78 | 584 |
| 19 | 3 | 73 | 675 |
| 20 | 1 | 69 | 1177 |
| 20 | 3 | 54 | 888 |
| 21 | 2 | 20 | 49 |
| 21 | 3 | 16 | 43 |
| 22 | 2 | 7 | 66 |
| 22 | 4 | 32 | 127 |
| 23 | 3 | 12 | 76 |
| 23 | 4 | 20 | 74 |
| 24 | 3 | 9 | 55 |
| 24 | 4 | 3 | 26 |
| END | | | |

OpenBUGSで「burn in」を10000回、「MCMC」を50000回実行後「stats」で得られた結果は以下のとおりである。

| | mean | sd | MC_error | val2.5pc | median | val97.5pc | start | sample |
|----------|---------|---------|----------|----------|---------|-----------|-------|--------|
| OR[1,2] | 2.517 | 1.284 | 0.07639 | 0.8763 | 2.229 | 5.924 | 10001 | 50000 |
| OR[1,3] | 2.629 | 0.7801 | 0.04756 | 1.404 | 2.517 | 4.508 | 10001 | 50000 |
| OR[1,4] | 3.782 | 1.753 | 0.1029 | 1.591 | 3.388 | 8.563 | 10001 | 50000 |
| OR[2,3] | 1.244 | 0.6155 | 0.03546 | 0.4579 | 1.129 | 2.845 | 10001 | 50000 |
| OR[2,4] | 1.819 | 1.132 | 0.06503 | 0.4546 | 1.529 | 4.995 | 10001 | 50000 |
| OR[3,4] | 1.483 | 0.6152 | 0.03562 | 0.6914 | 1.342 | 3.027 | 10001 | 50000 |
| RD[1,2] | 0.09202 | 0.06456 | 0.003849 | -0.01078 | 0.08138 | 0.2531 | 10001 | 50000 |
| RD[1,3] | 0.1012 | 0.03234 | 0.001993 | 0.03745 | 0.1008 | 0.169 | 10001 | 50000 |
| RD[1,4] | 0.1576 | 0.06774 | 0.003864 | 0.04821 | 0.1484 | 0.317 | 10001 | 50000 |
| RD[2,3] | 0.00922 | 0.06626 | 0.003951 | -0.1459 | 0.01715 | 0.1222 | 10001 | 50000 |
| RD[2,4] | 0.06561 | 0.09525 | 0.005655 | -0.1434 | 0.06561 | 0.2656 | 10001 | 50000 |
| RD[3,4] | 0.05639 | 0.06979 | 0.003991 | -0.05385 | 0.04746 | 0.2138 | 10001 | 50000 |
| RR[1,2] | 2.18 | 0.886 | 0.05251 | 0.8875 | 2.025 | 4.418 | 10001 | 50000 |
| RR[1,3] | 2.315 | 0.6025 | 0.03674 | 1.345 | 2.235 | 3.744 | 10001 | 50000 |
| RR[1,4] | 3.016 | 1.048 | 0.0613 | 1.5 | 2.83 | 5.734 | 10001 | 50000 |
| RR[2,3] | 1.194 | 0.491 | 0.02823 | 0.5561 | 1.106 | 2.46 | 10001 | 50000 |
| RR[2,4] | 1.572 | 0.7569 | 0.04348 | 0.5501 | 1.404 | 3.559 | 10001 | 50000 |
| RR[3,4] | 1.33 | 0.4038 | 0.0233 | 0.7376 | 1.262 | 2.258 | 10001 | 50000 |
| best[1] | 0 | 0 | 4.47E-13 | 0 | 0 | 0 | 10001 | 50000 |
| best[2] | 0.1794 | 0.3837 | 0.01995 | 0 | 0 | 1 | 10001 | 50000 |
| best[3] | 0.1539 | 0.3608 | 0.0161 | 0 | 0 | 1 | 10001 | 50000 |
| best[4] | 0.6667 | 0.4714 | 0.0228 | 0 | 1 | 1 | 10001 | 50000 |
| probt[1] | 0.08299 | 0.01644 | 0.000991 | 0.05367 | 0.08177 | 0.1177 | 10001 | 50000 |
| probt[2] | 0.175 | 0.06423 | 0.003843 | 0.07964 | 0.1626 | 0.3368 | 10001 | 50000 |
| probt[3] | 0.1842 | 0.02659 | 0.001635 | 0.1384 | 0.1818 | 0.2409 | 10001 | 50000 |
| probt[4] | 0.2406 | 0.06579 | 0.003732 | 0.1378 | 0.2308 | 0.3934 | 10001 | 50000 |
| rk[1] | 3.954 | 0.2098 | 0.009321 | 3 | 4 | 4 | 10001 | 50000 |
| rk[2] | 2.436 | 0.8324 | 0.04505 | 1 | 3 | 4 | 10001 | 50000 |
| rk[3] | 2.167 | 0.6688 | 0.0326 | 1 | 2 | 3 | 10001 | 50000 |
| rk[4] | 1.442 | 0.6841 | 0.03449 | 1 | 1 | 3 | 10001 | 50000 |

これより、gemtcパッケージのコントラストベースモデルの結果とほぼ同じ結果が得られた。medianの列を見ると、rk[4]すなわち治療4の順位が1で95%確信区間(val2.5pcとval97.5pc)が1~3となっており、また、治療4が最善である確率はbest[4]のmeanを見ると0.6667であることが分かる。参照治療1に対して治療4のオッズ比3.388(95%確信区間1.591~8.563)となっており、gemtcパッケージの結果よりも全体として少し大きめの値になっていることが分かる。

OpenBUGS を、R から BRugs パッケージを介して動かすことも可能である。以下にその一例を示す。

```

###Arm-based Network Meta-analysis for binary outcomes#####
###based on the BUGS code by Zhang: J Clin Trials 2014;11:246-62. PMID: 24096635###
#You need OpenBUGS installed.(You can get it at http://openbugs.net/w/FrontPage)
#-----
packneed=c("BRugs", "forestplot", "tcltk2");current=installed.packages();addpack=setdiff(packneed,rownames(current));url
="https://cran.ism.ac.jp/";if(length(addpack)>0){install.packages(addpack,repos=url);if(length(addpack)==0){print("Already
installed.")}
#-----
###To use BRugs, tcltk, and forestplot.
library(BRugs)
library(tcltk2)
library(forestplot)
#-----
###Set the working directory where you have data, inits text files, and Results folder.
path0=tk_choose.dir(default="",caption="Select a folder to save data files")
path0=paste(path0, "/", sep="")
setwd(path0)
#Create a folder for results in the working directory if there is not.
paths=paste(path0, "Results/", sep="")
if(!dir.exists(paths)){dir.create(paths)}

#Set a name to data and result files.
titl="NWM_smoke" #Set title as you want..
datfilename=paste(titl, "_dat.txt", sep="") #File name for data.
numfilename=paste(titl, "_num.txt", sep="") #File name for the number of studies and others.

#Open and read a model text and save as a text file in the working directory you set.
filnam=tclvalue(tkgetOpenFile(initialfile="",filetypes="{{Text Files} {.txt}} {{All files} *}"));if(filnam!=""){nma_
model=readLines(filnam)}
#nma_model==readLines("https://stat.zanet.biz/bugs/nma/ab_model_bi.txt")
sink(modelfilename) #The model file name to be used.
writeLines(nma_model)
sink()

#-----
#Make data to be analyzed and save as text files.
dat=read.delim("clipboard",sep="\t",header=FALSE) #Select only number cells in Excel and copy and get data via
clipboard into dat.

colnames(dat)=c("s[]", "t[]", "r[]", "total[]") #Add column names as the names of variables.
tS=length(dat[,1]) #The number of all comparisons.
#The number of studies.
#Add "END" to the data. dat holds the case numbers data to be analyzed.
tN=length(unique(dat[,2])) #The number of treatment options.
sN=length(unique(dat[,1])) #The number of studies.
dat=rbind(dat,c("END", "", "", "", ""))

write.table(dat,datfilename,quote=FALSE,sep="\t",row.names=FALSE,col.names=TRUE) #Save a text file of the data.

numex=paste("list(tN=",tN,",sN=",sN,",tS=",tS,sep="") #tN the number of
treatment options; sN the number of studies; tS the number of all comparisons.
zeros=rep(0,tN)

```

```

numex=paste(numex,"mn=c(",toString(zeros),")",",",R=structure(.Data=c(",sep=""")
rmat=rep(0,tN*tN)
for(i in 1:tN){
  rmat[tN*(i-1)+i]=1
}
numex=paste(numex,toString(rmat),"),.Dim=c(",tN,",",tN,"))",sep=""")

write.table(numex,numfilename,quote=FALSE,row.names=FALSE,col.names=FALSE) #Save a text file of the data in the
working directory.

###Check the model, load the data from the data files, compile,
#read the inits file and set inits values and generate inits for not-given inits.
modelCheck(modelfilename)
modelData(datfilename)
modelData(numfilename)

modelCompile(numChains=1)
#modelInits("inits.txt")
modelGenInits()

#Burn-in MCMC, set variables for sampling, and do MCM
modelUpdate(10000) #<---You can change iteration number.

#For imperfect reference standards with covariance-adjustments (model #3).
samplesSet(c("OR","RR","RD","probt","rk","best"))

dicSet() #To get deviance data.
modelUpdate(50000) #<---MCMC iteration number you can change.

#Show stats data in the console window.
res=samplesStats("*")
print(titl)
#print(modelname)
print(res)

#-----
#-----
###Save files of the results.
setwd(paths)

#Make file names for stats, density, DIC, History.
resstats=paste(modelname,titl,"Stats.csv",sep="_")
resstatstxt=paste(modelname,titl,"Stats.txt",sep="_")
resdensity=paste(modelname,titl,"Density.pdf",sep="_") #This file would be large in size, several hundred KB.
resdic=paste(modelname,titl,"DIC.txt",sep="_")
reshistory=paste(modelname,titl,"History.pdf",sep="_")
resmodelname=paste(modelname,titl,"model.txt",sep="_")

#Save model name, path, and model script as a text file in the "Results" directory.
sink(resmodelname)
cat("# ")
writeLines(modelname)
cat("# ")
writeLines(modelfilename)
setwd(path0)
modlines=readLines(modelfilename)
setwd(paths)

```

```

writeLines(nma_model)
sink()

#Save stats as a csv file.
write.csv(modelname,resstats)
write.csv(samplesStats(""),resstats,append=TRUE)

#Save density plots as a PDF file.
pdf(resdensity)
samplesDensity("")
dev.off()

#Save deviance data as a text file.
sink(resdic)
print(dicStats())
sink()

#Save history data file.
pdf(reshistory)
samplesHistory("")
dev.off()

#Set working directory back to path0.
setwd(path0)
#-----
#-----

```

#Open and read a model text の部分で、BUGS コードの部分のみを記述したテキストファイルを開くと、その結果は設定したフォルダ内の「Results」フォルダに出力される。

(5) pcnetmeta パッケージによるアームベースモデルによるネットワークメタアナリシス

pcnetmeta パッケージは、R から rjags を介して JAGS を動かしてベイジアンの手法でネットワークメタアナリシスを実行するパッケージである。

```

###-----###
###pcnetmeta for network meta-analysis###
library(pcnetmeta)

###Setting the working directory.
library("tcltk2")
#Set a folder to save files.
paths=tk_choose.dir(default="",caption="#1 Select a folder to save data files")
paths=paste(paths,"/",sep="")
setwd(paths)

###Use smoke cessation data.
data("smoke")
exdat=smoke
#or ###Read data via clipboard. Remove the head # to use thie line.
#exdat=read.delim("clipboard",sep="\t",header=TRUE)

```

```

#Set the names of treatments and the reference treatment name.
tx.name=c("NC", "SH", "IC", "GC")
cont="NC"

#Set title and study name, and make a network geometry plot.
title="Smoking Cessation"
stnam="smoke_pcnm"
#nma.networkplot(s.id=s.id, t.id=t.id, data=exdat,trtname = tx.name, title = title)
#Save the geometry plot in a pdf file.
filpdf=paste("net_plot_",stnam, ".pdf",sep="")
pdf(filpdf) #File name to save plots in a PDF file.
nma.networkplot(s.id=s.id, t.id=t.id, data=exdat,trtname =tx.name, title = title)
dev.off()

#--Run MCMC. This process takes time.
modelnam="het_cor" #The default model for the variance-covariance matrix.
#modelnam="hom_eqcor"

#Run MCMC: burn-in 10000, MCMC 50000.
set.seed(12345)
exdat.out <- nma.ab.bin(s.id, t.id, r, n, data = exdat, trtname = tx.name,
param = c("AR", "OR", "RR", "LOR", "LRR", "RD", "rank.prob"),
model = modelnam, higher.better = TRUE, digits = 3,
n.adapt = 10000, n.iter = 50000, n.thin = 1, conv.diag = TRUE,
dic = TRUE, trace = "OR", postdens = TRUE)

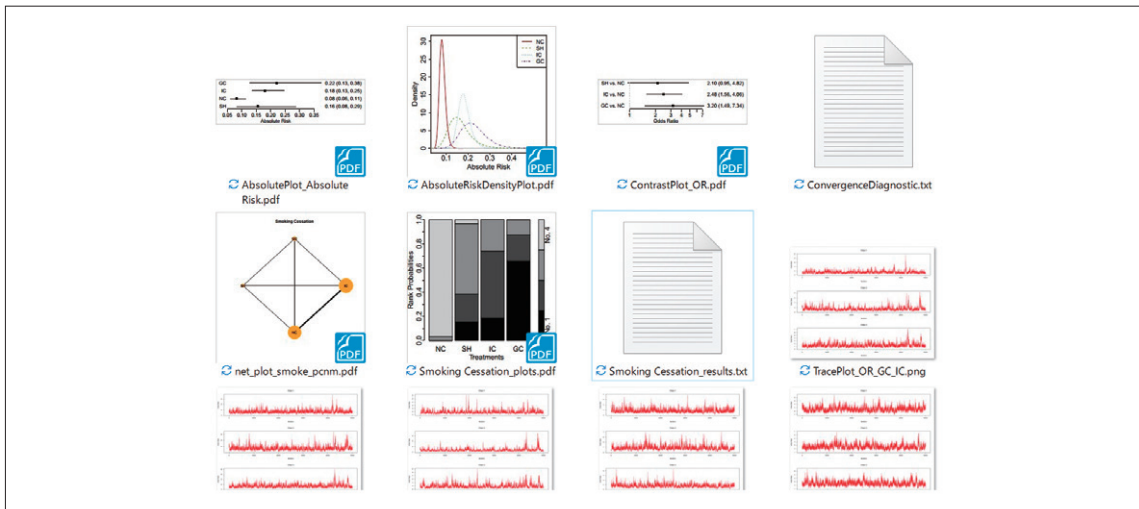
#--
#Save the results as a text file.
resfilnam=paste(title, "_results.txt",sep="") #Set the file name (xxx.txt)
sink(resfilnam)
print(exdat.out$model)
print(exdat.out$AbsoluteRisk)
print("Odds Ratio")
print(exdat.out$OddsRatio)
print("Risk Ratio")
print(exdat.out$RelativeRisk)
print("Risk Difference")
print(exdat.out$RiskDifference)
print(exdat.out$TrtRankProb)
print(exdat.out$DIC)
sink()

#Save plots as a PDF file.
filpdf=paste(title, "_plots.pdf",sep="") #File name to save plots in a PDF file.
pdf(filpdf)
absolute.plot(exdat.out, width = 5, height = 1.5)
contrast.plot(exdat.out, reference = cont, width = 5, height = 1.5)
rank.prob(exdat.out, cex.axis = 2, cex.lab = 2)
dev.off()

#####

```

結果として以下のファイルが出力される。



(6) netmeta パッケージによる頻度論派の手法によるネットワークメタアナリシス

netmeta パッケージは、アウトカムが二値変数の場合はオッズ比，リスク比，率差を，連続変数の場合は平均値差，ハザード比などを取り扱えるとともに，それらのサンプルデータが提供されている。今回は率差を効果指標とした解析例を示す。

はじめに，R を起動して以下のスクリプトを実行する。

```
library(netmeta)      #netmeta パッケージを読み込む。
data(smokingcessation) # データセット smokingcessation を読み込む。
dat=smokingcessation # データセットの内容を変数 dat に格納する。
```

必要なデータ形式は次のようになる。

| | A | B | C | D | E | F | G | H | I | J |
|----|-------|--------|-----|--------|------|--------|-----|--------|--------|--------|
| 1 | study | event1 | n1 | event2 | n2 | event3 | n3 | treat1 | treat2 | treat3 |
| 2 | 1 | 9 | 140 | 23 | 140 | 10 | 138 | A | C | D |
| 3 | 2 | 11 | 78 | 12 | 85 | 29 | 170 | B | C | D |
| 4 | 3 | 75 | 731 | 363 | 714 | NA | NA | A | C | |
| 5 | 4 | 2 | 106 | 9 | 205 | NA | NA | A | C | |
| 6 | 5 | 58 | 549 | 237 | 1561 | NA | NA | A | C | |
| 7 | 6 | 0 | 33 | 9 | 48 | NA | NA | A | C | |
| 8 | 7 | 3 | 100 | 31 | 98 | NA | NA | A | C | |
| 9 | 8 | 1 | 31 | 26 | 95 | NA | NA | A | C | |
| 10 | 9 | 6 | 39 | 17 | 77 | NA | NA | A | C | |
| 11 | 10 | 79 | 702 | 77 | 694 | NA | NA | A | B | |
| 12 | 11 | 18 | 671 | 21 | 535 | NA | NA | A | B | |

これは netmeta パッケージのデータセットに含まれる smokingcessation のデータの一部であるが、4つの治療法の名称は A, B, C, D である。これらのうち3つを直接比較した研究は 1, 2 の2つであり、それ以外の研究では2つの治療法が比較されている。したがって、1つの研究で比較された群は最大で3つあり、それぞれの群の症例数は n1, n2, n3 で表されている。それぞれの群のアウトカム生起症例数は event1, event2, event3 で表されている。2群の比較しか行われていない研究の場合、データなしとしてセルには NA が入力されている。どの治療法が用いられたかは、treat1, treat2, treat3 のカラムに入力されている。こちらは、該当する治療なしの場合は空欄となっている。treat1 の内容は研究によって同じではなく、treat1, treat2, treat3 はそれぞれの研究の群 1, 群 2, 群 3 の受けた治療に対応している。

Excel で用意したデータをクリップボード経由で読み込んで解析する場合には、セル B1 から右下のデータの部分をコピーする（カラム A は不要）。そして、R の側で以下のスクリプトを実行する。R のエディターにこれを書き込み、Excel から必要なデータ範囲を選択してコピーし、R に戻ってこのスクリプトの行にカーソルを置き [実行ボタン] をクリックする。Mac の場合は [Command] + [Enter] キーを押す。Mac の場合は Numbers からクリップボード経由でデータを読み込むこともできる。

```
dat=read.delim("clipboard",sep="\t",header=TRUE)      #Windows の場合
exdat=read.delim(pipe("pbpaste"),sep="\t",header=F);reiv()  #Mac の場合
```

このデータはアウトカムが二値変数で表され、効果指標としてオッズ比、リスク比、リスク差が用いられる場合の例となる。pairwise() 関数の sm 引数に「OR」, 「RR」, 「RD」 と設定することで、それぞれ対応した解析が行われる。

まず、pairwise() 関数を実行し、結果を変数 p3rd に格納する（変数名は任意）。以下の例では、リスク差を指標としている。次に、p3rd の最初の 6 行を head() 関数で表示させる。

```
p3rd=pairwise(list(treat1,treat2,treat3),list(event1,event2,event3),list(n1,n2,n3),data=dat,sm="RD")
head(p3rd)

> p3rd=pairwise(list(treat1,treat2,treat3),list(event1,event2,event3),list(n1,n2,n3),data=dat,sm="RD")
> head(p3rd)
      TE      seTE studlab treat1 treat2 event1  n1 event2  n2
1 -0.1000000000 0.03755462      1      A      C      9 140      23 140
2 -0.0081780538 0.03027728      1      A      D      9 140      10 138
3  0.0918219462 0.03831102      1      C      D     23 140      10 138
4 -0.0001508296 0.05458446      2      B      C     11  78      12  85
5 -0.0295625943 0.04883979      2      B      D     11  78      29 170
6 -0.0294117647 0.04752580      2      C      D     12  85      29 170
```

引き続き、pairwise() 関数の実行結果に対して、netmeta() 関数を実行する。comb.random 引数を TRUE に設定するとランダム効果モデル、FALSE に設定すると固定効果モデルが用いられる。続いて net3r を実行して結果を表示させる。

```
net3r=netmeta(TE,seTE,treat1,treat2,studlab,comb.random=TRUE,data=p3rd)
net3r
```

```
Random effects model

Treatment estimate (sm='RD'):
      A      B      C      D
A      . -0.0412 -0.0944 -0.1291
B  0.0412      . -0.0531 -0.0879
C  0.0944  0.0531      . -0.0347
D  0.1291  0.0879  0.0347      .

Lower 95%-confidence limit:
      A      B      C      D
A      . -0.1294 -0.1438 -0.2237
B -0.0469      . -0.1456 -0.1956
C  0.0449 -0.0394      . -0.1268
D  0.0345 -0.0199 -0.0573      .

Upper 95%-confidence limit:
      A      B      C      D
A      .  0.0469 -0.0449 -0.0345
B  0.1294      .  0.0394  0.0199
C  0.1438  0.1456      .  0.0573
D  0.2237  0.1956  0.1268      .

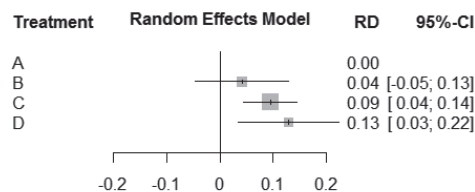
Quantifying heterogeneity/inconsistency:
tau^2 = 0.0092; I^2 = 94%

Test of heterogeneity/inconsistency:
      Q d.f. p.value
380.61  23 < 0.0001
```

comb.random=TRUE の設定では、固定効果モデルの計算も行われるので、両方のモデルの結果がコンソールに出力される。

続いて、forest() 関数でフォレストプロットを作成する。この際には、ref 引数に基準とする治療の名称を設定する。

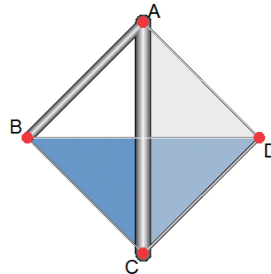
```
dev.new();forest(net3r,ref="A")
```



A を基準に設定したので、A の RD は 0、B、C、D に対しては A を基準とした RD の統合値と 95% 信頼区間が表示される。この結果から治療 A と C および D では差があるが、A と B では差がなく、B、C、D の間でも差がないと判定される。また、A と C の比較が最も信頼区間が狭く、症例数が多く、かつ研究数も多いことが反映されていると考えられる。次に示すネットワークグラフにおいて A と C を結ぶ線が一番太いことと対応している。

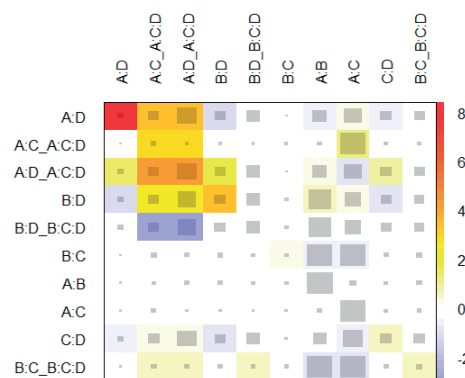
続いて、`netgraph()` 関数でネットワークグラフを作成する。点の大きさは `cex` で指定でき、点を結合する線の太さは標準誤差の逆数に比例させるように設定してある。

```
dev.new();netgraph(net3r,points=TRUE,cex.points=3,cex=1.5,thickness="se.random")
```



次に、ネットヒートプロットを作成する。

```
dev.new();netheat(net3r)
```



ネットヒートプロットは Krahn U ら (2013) が提唱した方法で、直接比較の結果と間接比較の結果の乖離の程度すなわち不一致の程度を表したものである。左側の列に研究で実際に直接比較された治療のペアが示されている。「_」の右側は研究で直接比較された治療を示している。上の行には間接比較も含めたネットワークメタアナリシスで比較された治療のペアを示している。グレーの四角はネットワークメタアナリシスの結果への直接比較の影響 (contribution) を表している。直接比較の結果を除いた場合に不一致の程度が減少するか、増加するかを色で示しており、暖色は減少、寒色は増加することを示し、その程度は色の濃さで表されている。色が濃いほど不一致の程度の変動が大きいことを示す。この例では、治療 A と D の比較は、直接比較を除いてしまうと、不一致の程度が大きく減少することを示しているため、間接比較を含めたネットワークメタアナリシスの結果は異質性が高いことを示している。

最後に、`netmeasures()` 関数で、各指標の値をコンソールに出力する。

```
dnetmeasures(net3r)
```

4.9 補足資料4：システマティックレビューツール

• GRADEpro GDT

診療ガイドライン作成作業全般を支援するウェブツールとして、McMaster University and Evidence Prime Inc. が提供する GRADEpro GDT がある (GRADEpro GDT 2021)。これは GRADE ワーキンググループの Schünemann H らが開発に参加しており、GRADE アプローチに沿った作業をクラウドの機能を利用して行えるように、スコープ作成から推奨作成まで、チーム作業を支援する形でさまざまな機能が提供されている。診療ガイドライン作成グループが利用する場合は有料となっている。個人の使用は無料、2名のチームは\$29 (12CQ まで)、15名までのチームは\$2,000 (25CQ まで)、それ以上の場合は個別に利用料が設定される (2021年2月24日現在)。

• Rayyan

文献検索の結果から選定する作業は時間と労力を要するが、無料のクラウドサービスとして Qatar Computing Research Institute (Data Analytics) が提供している Rayyan (レイヤン) がある (Ouzzani M 2016)。文献リストの作成を共同作業で進めることが可能である。また、JavaScript で作成された PubMed 検索用ウェブツール pmSearch も利用可能である (pmSearch 2019)。

• RevMan web

現在の RevMan (使用方法は『Minds 診療ガイドライン作成マニュアル 2017』第4章4.6節を参照) はダウンロードしてスタンドアロンのアプリケーションとして動かすことができるが、Cochrane は今後クラウドを利用したウェブツールに変更する計画である。そのため、今後自由に利用できなくなる可能性がある。

• 機械学習を利用したシステマティックレビュー支援ツール

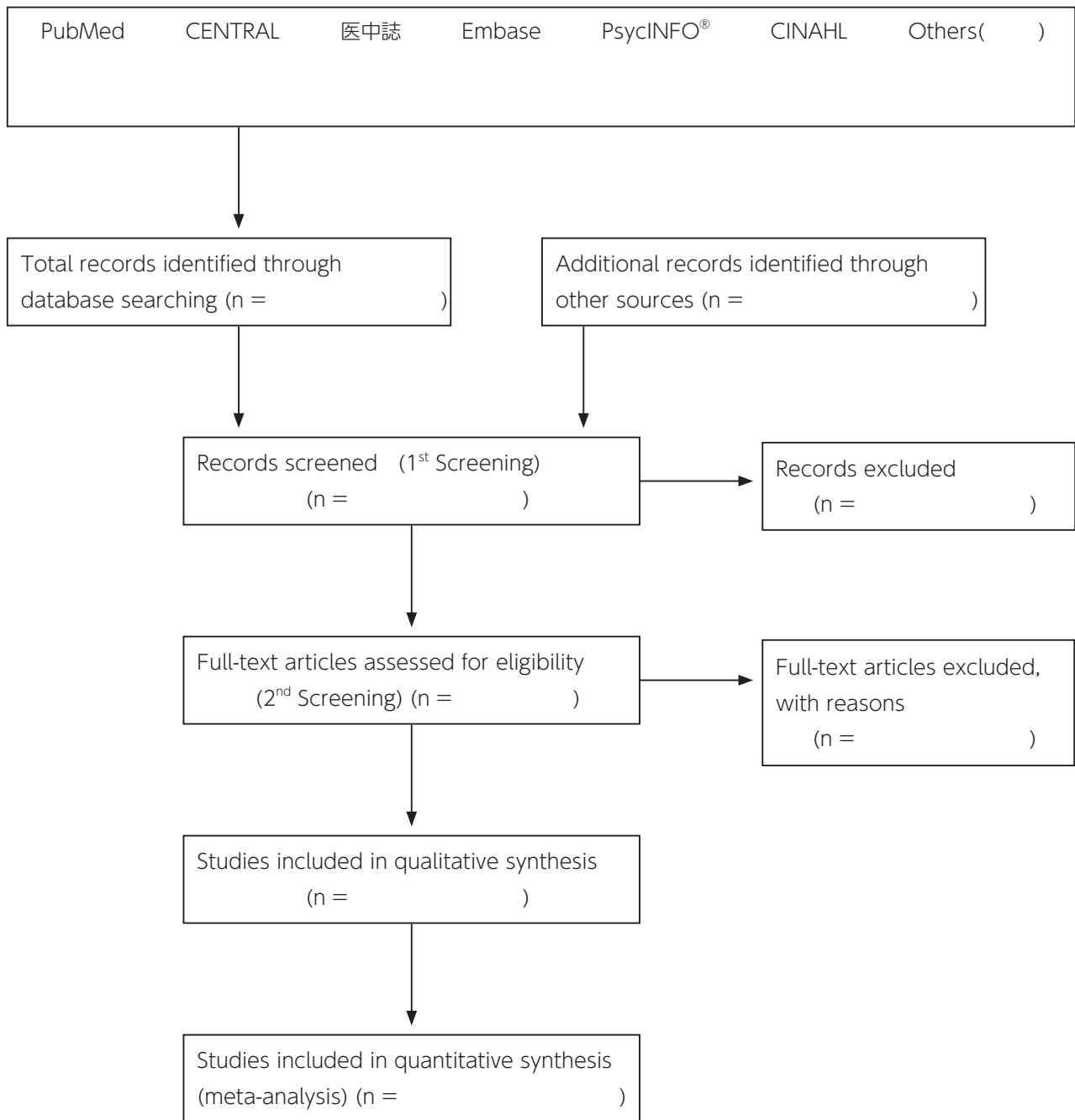
機械学習を利用し、文献選定作業を省力化するツールとして SWIFT-Review (Howard BE 2016, O'Mara-Eves A 2015) が発表されているが、感度を100%にすることは困難であることが示されている。また、バイアスリスクの評価を自動化することも試みられており、バイアスドメインの評価結果を High, Low の二値変数として扱ったときに、Cochrane の熟練したレビューの評価結果との一致率が70%程度となることが示されている [Robot Reviewer (Marshall IJ 2016), robot analysis 4.8.7 (Millard LA 2016)]。

また、PubMed では Best match という機能が提供されており、アブストラクト付き、フルテキストなどに限定するメニューの中から選択できる。Best match では各論文のタイトルとアブストラクトの単語の出現頻度のデータを参照し、検索式の語句の出現頻度が高い論文から順番に表示する。その機能は感度100%には到達していないため、PubMed 自体はシステマティックレビューの文献検索には用いないことを推奨している。

• PICORON

PDF ファイルの論文中で「randomization」, 「outcome」などのキーワードをハイライトする機能を持つ。

【SR-2 文献検索フローチャート】



【SR-4 引用文献リスト】

| | 文献 ID | 書誌情報 |
|----------|-------|------|
| 採用論文 | | |
| 不採用論文 | | |
| その他の引用論文 | | |

【SR-8 評価シート エビデンス総体 (絶対効果指標の結果を記入する場合)】

| | |
|----------|--|
| 診療ガイドライン | |
| 対象 | |
| 介入 | |
| 対照 | |

エビデンスの強さについて、RCTは「強 (A)」からスタート、観察研究は「弱 (C)」からスタート。
 *各ドメインは「高 (-2)」、「中/疑い (-1)」、「低 (0)」の3段階。
 **エビデンスの強さは「強 (A)」、「中 (B)」、「弱 (C)」、「非常に弱 (D)」の4段階。
 ***重要性はアウトカムの重要性 (1~9)。

| エビデンス総体 | リスク人数 (アウトカム率) | | | | | | 効果指標 (種類) | 効果指標統合値 | 95%信頼区間 | 絶対効果 | コメント | 重要性 | エビデンスの強さ | 95%信頼区間 | 効果指標統合値 | 95%信頼区間 | |
|---------|----------------|-----------|-------|-----------|-------|-----------|-----------|---------|---------|------|------|-----|----------|---------|---------|---------|--|
| | 対照群分母 | 対照群分子 (%) | 介入群分母 | 介入群分子 (%) | 介入群分母 | 介入群分子 (%) | | | | | | | | | | | |
| アウトカム | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

コメント (該当するセルに記入)

| | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |

【SR-9 定性的システマティックレビュー】

| | | |
|-------|--|--|
| CQ | | |
| P | | |
| I | | |
| C | | |
| 臨床的文脈 | | |

| | | |
|-------------|--|--|
| 01 | | |
| 非直接性のまとめ | | |
| バイアスリスクのまとめ | | |
| 非一貫性その他のまとめ | | |
| コメント | | |

| | | |
|----|--|--|
| 02 | | |
|----|--|--|

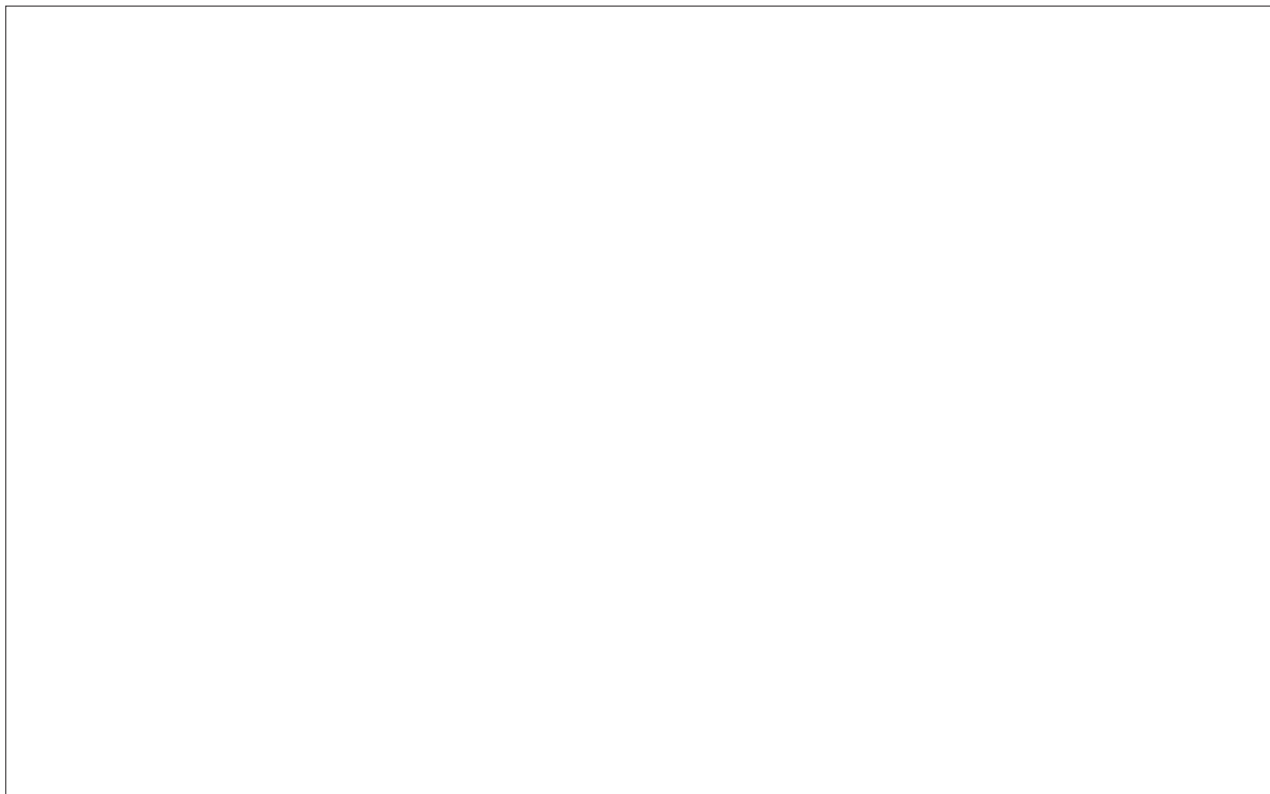
| | | |
|----|--|--|
| 03 | | |
|----|--|--|

⋮

【SR-10 メタアナリシス】

| | | | |
|---------------|--|-------|----------|
| CQ | | | |
| P | | I | |
| C | | O | |
| 研究デザイン | | 文献数 | |
| コード | | | |
| モデル | | 方法 | |
| 効果指標 | | 統合値 | (-) P= |
| フォレスト プロット | | | |
| | | コメント： | |
| ファンネル プロット | | | |
| | | コメント： | |
| その他の解析 | | コメント： | |
| メタリグ レッション | | | |
| 感度分析 | | | |

【SR-11 システマティックレビューレポートのまとめ】



【SR-12 結果のまとめ (SoF 表) (ペア比較のメタアナリシス)】

| 重要臨床課題： | | | | | | |
|--------------------------------------|-------------------|----------------------|-------|------------------|------------------------|-------------|
| 疾患／対象者： | | | | | | |
| セッティング： | | | | | | |
| 介入： | | | | | | |
| 対照： | | | | | | |
| アウトカム 対象者数 (研究数) | 相対効果 (95%信頼区間) | 期待される絶対効果* (95%信頼区間) | | | エビデンス 確実性 | 何が 起きるか？ |
| | | 対照 | 介入 | 差 | | |
| アウトカム 1： 介入 経過観察期間： | | | | | | |
| 対象者 1 対象者数 (研究数) | | % | % (～) | % 少ない (～ 少ない) | ⊕⊕⊕⊕ 高 Due to | |
| 対象者 1 対象者数 (研究数) | | % | % (～) | % 少ない (～ 少ない) | ⊕⊕⊕○ 中 Due to | |
| アウトカム 2： 経過観察期間： 対象者数 (研究数) | | % | % (～) | % 少ない (～ 少ない) | ⊕⊕○○ 低 Due to | |
| アウトカム 3： 経過観察期間： 対象者数 (研究数) | | % | % (～) | % 少ない (～ 少ない) | ⊕○○○ 非常に低 Due to | |
| アウトカム 4： 経過観察期間： 対象者数 (研究数) | | % | % (～) | % 少ない (～ 少ない) | ⊕○○○ 非常に低 Due to | |
| * | | | | | | |
| ** | | | | | | |
| 解説 | | | | | | |
| 1 | | | | | | |
| 2 | | | | | | |
| 3 | | | | | | |
| 4 | | | | | | |
| 5 | | | | | | |
| 6 | | | | | | |
| 7 | | | | | | |
| 8 | | | | | | |
| 9 | | | | | | |
| 10 | | | | | | |
| 11 | | | | | | |
| 12 | | | | | | |

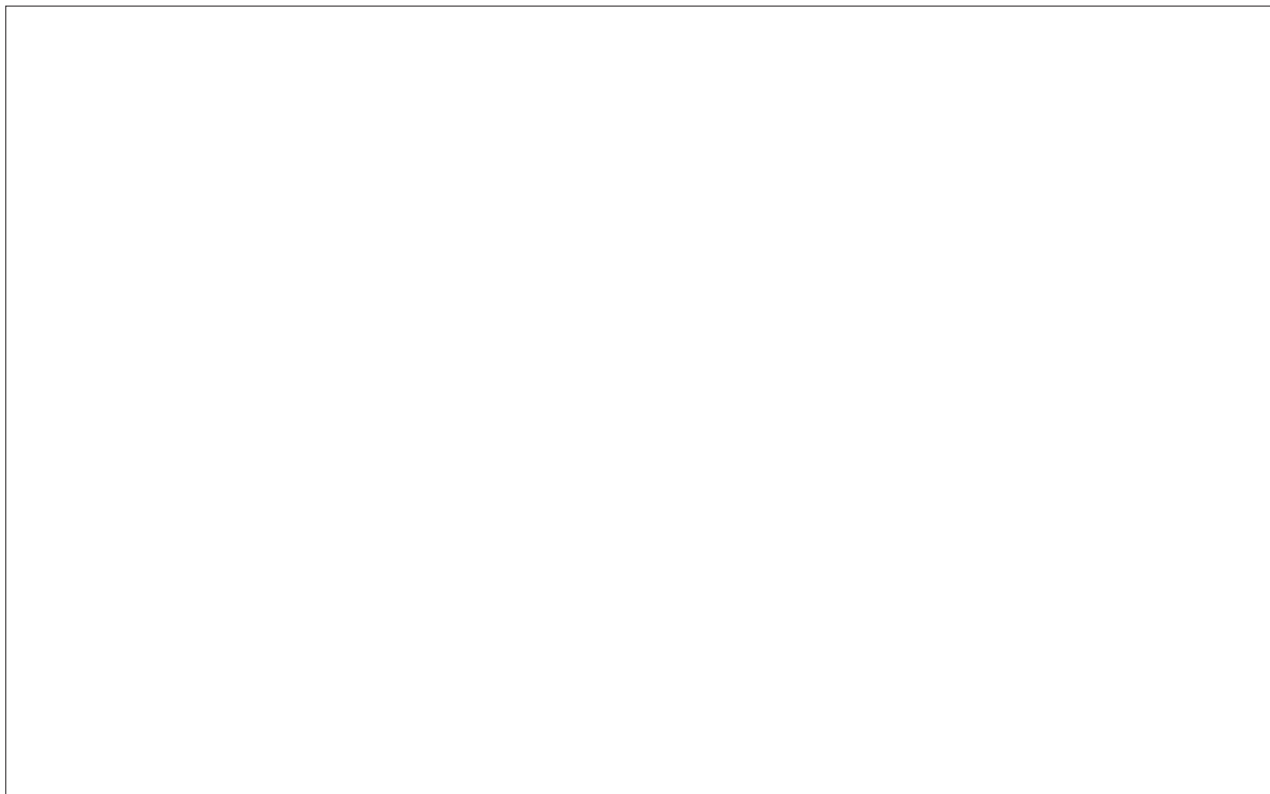
出典：Guyatt GH, et al. (2012) GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. J Clin Epidemiol 66:158-172.
By courtesy of Dr. Guyatt より作成

【SR-14 結果のまとめ (SoF 表) (ネットワークメタアナリシス：様式 2)】

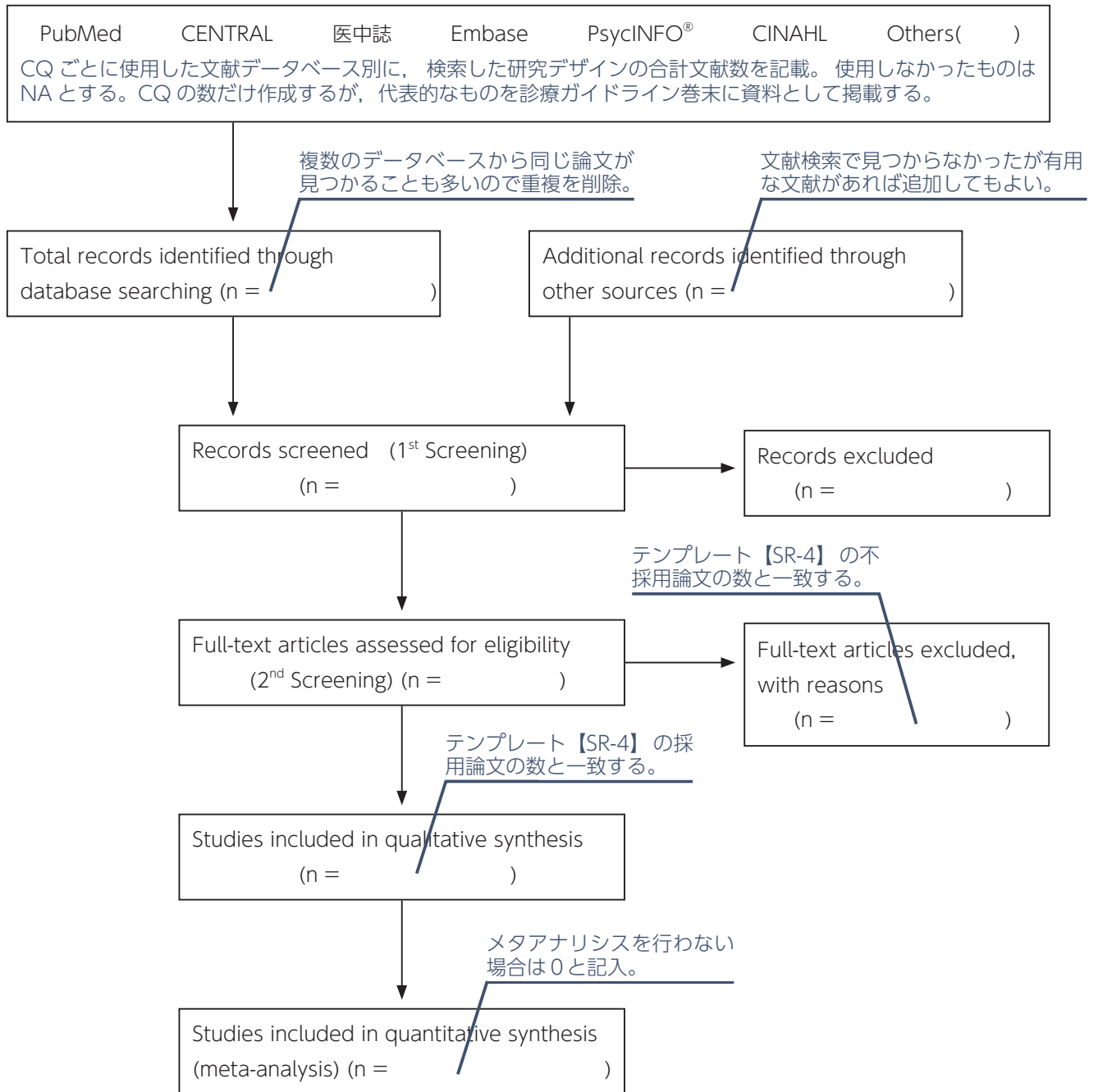
| 益／害 | | | | | | | |
|--|---|------------------|---|------------------|---|------------------|------|
| 患者／参加者 (P) : 介入 (I) : 対照 (参照) (C) : ネットワークメタアナリシス Geomety plot* アウトカム (O) : セッティング (S) : | | | | | | | |
| アウトカム | 効果推定値と確信区間 | | | | | | コメント |
| | 治療 1 | | 治療 2 | | 治療 3 | | |
| アウトカム 1 | | | | | | | |
| 参照治療 人／人 (%)*** | ORなど相対 効果 (95%CrI) ネットワーク 推定値 | 絶対効果 (95%CrI) | ORなど相対 効果 (95%CrI) ネットワーク 推定値 | 絶対効果 (95%CrI) | ORなど相対 効果 (95%CrI) ネットワーク 推定値 | 絶対効果 (95%CrI) | |
| | 効果推定値に対する確信度 | | 高 ⊕⊕⊕⊕ | | 中 ⊕⊕⊕⊖ | | |
| 順位 (95%CrI) | 順位 (95%CrI) | | 3(2~9) | | 3(1~10) | | |
| | 参加者総数 (研究数) | | 823(3RCT) | | 427(1RCT) | | |
| アウトカム 1 | | | | | | | |
| 参照治療 人／人 (%) | ORなど相対 効果 (95%CrI) ネットワーク 推定値 | 絶対効果 (95%CrI) | ORなど相対 効果 (95%CrI) ネットワーク 推定値 | 絶対効果 (95%CrI) | ORなど相対 効果 (95%CrI) ネットワーク 推定値 | 絶対効果 (95%CrI) | |
| | 効果推定値に対する確信度 | | 効果推定値に対する確信度 | | 効果推定値に対する確信度 | | |
| 順位 (95%CrI) | 順位 (95%CrI) | | 順位 (95%CrI) | | 順位 (95%CrI) | | |
| | 参加者総数 (研究数) | | 参加者総数 (研究数) | | 参加者総数 (研究数) | | |
| NMA-SoF 表の定義 *実線は直接比較 **推定値はオッズ比。CrIは確信区間 (ベイジアンアプローチによる)。 ***想定されるリスク (例えば、研究横断的な対照群の中央値) は脚注に記載。 CrI : Credible Interval, OR : Odds Ratio | | | | | | | |
| GRADE によるエビデンスのグレーディング 高い質：将来の研究が効果推定値に対する我々の確信度を変える可能性は非常に低い。 中等度の質：将来の研究が効果推定値に対する我々の確信度へ重要な影響を持つ可能性があり、推定値を変えるかもしれない。 低い質：将来の研究が我々の効果推定値に対する確信度へ重要な影響を持つ可能性がかなりあり、推定値を変える可能性がある。 非常に低い質：推定値に対して我々は非常に不確かである。 | | | | | | | |
| 解説脚注： 1 ベースラインリスク (想定される対照群のリスク) は National Cancer Institute の集約プログラムから得た。 2 95%確信区間が同等の値と交差するので非常に重篤な不精確性があり、広い確信区間から害の可能性が高いことを示す。 3 概念的にはあり得る修飾因子分布は異なる化学療法薬を用いる研究間で差がないので、明らかな非移行性はない。 4 順位は中央値 (順位 1 ~ 10) と 95%確信区間を示す。 | | | | | | | |

一番左の列には Geomety plot への参照を含める (ノードの色など)。

【SR-15 Future Research Question】



【SR-2 文献検索フローチャート 記入方法】



【SR-4 引用文献リスト 記入方法】

| | 文献 ID | 書誌情報 |
|----------|-------------------------------|---|
| 採用論文 | ハーバード方式による文献整理 ID で、本文と対応させる。 | テンプレート【SR-3】に残った、システマティックレビューに使用した論文の書誌情報を記載する。 |
| 不採用論文 | ハーバード方式による文献整理 ID で、本文と対応させる。 | テンプレート【SR-3】から、その後のステップで削除された論文の書誌情報を記載する。 |
| その他の引用論文 | ハーバード方式による文献整理 ID で、本文と対応させる。 | 総論や疫学など、システマティックレビューを行わない章において引用した論文の書誌情報を記載する。 |

【SR-9 定性的システマティックレビュー 記入方法】

| | | |
|-------|------|--|
| CQ | 管理番号 | CQの文章を記入する。 |
| P | | CQの文章を補足する十分に詳細な記述をする。タイミングやセッティングも必要に応じて記述する。 |
| I | | 介入の詳細を記述する。 |
| C | | 1:1の比較が望ましいが、必要に応じて複数の比較を記述する。 |
| 臨床的文脈 | | 診療のプロセスのどこに位置付けられるか、診断、治療、予防、予後予測、その他のいずれに分類されるかなどを記述する。 |

| | | |
|-------------|--|---|
| 01 | | 害のようにCQの文章に表れない場合もあるが、エビデンス総体の対象が分かるように記述する。 |
| 非直接性のまとめ | | CQと得られたエビデンス総体間の乖離について記述する。介入の比較が非直接的である場合や日本人患者への適用に問題がある場合には必ず記述する。 |
| バイアスリスクのまとめ | | 特に問題となるバイアスリスク、推奨の決定に影響を及ぼすバイアスリスクを記述する。 |
| 非一貫性その他のまとめ | | 複数の研究間の相違の大きさ、信頼区間やバイアスリスクから判定した不確実性、効果の大きさなどについてまとめを記述する。研究デザインが異なるものをまとめた場合はそれについて記述する。 |
| コメント | | このアウトカムに関するエビデンス総体の評価において特に注意すべき点があればそれを記述する。 |

| | | |
|----|--|--|
| 02 | | |
|----|--|--|

| | | |
|----|--|--|
| 03 | | |
|----|--|--|

⋮

【SR-11 システマティックレビューレポートのまとめ 記入方法】

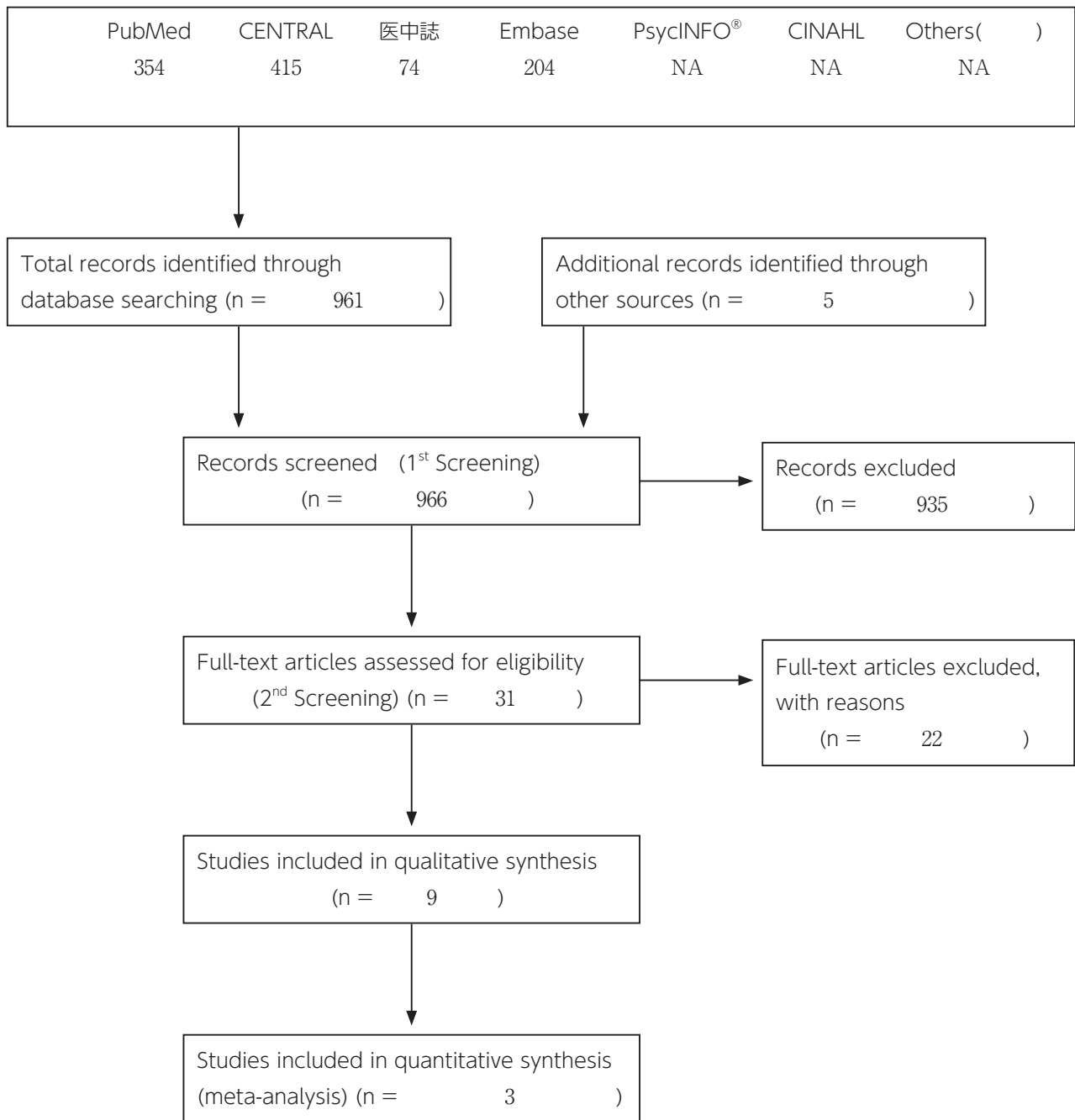
テンプレート【SR-9】、【SR-10】の結果をまとめる。

【SR-15 Future Research Question 記入方法】

現在の研究では解明されていない課題から、将来の研究で回答が得られるであろう CQ を作成する。以下の項目を記述する。

- ・リサーチエスチョン
- ・背景
- ・可能な研究計画の概略

【SR-2 文献検索フローチャート 記入例】



【SR-4 引用文献リスト 記入例】

| | 文献 ID | 書誌情報 |
|----------|-------------------------------------|---|
| 採用論文 | The IST-3 collaborative group, 2012 | The benefits and harms of intravenous thrombolysis with recombinant tissue plasminogen activator within 6 h of acute ischaemic stroke (the third international stroke trial [IST-3]): a randomised controlled trial. <i>Lancet</i> . 2012. 379(9834):2352-2363. |
| 不採用論文 | | |
| その他の引用論文 | | |

【SR-5 評価シート 介入研究 記入例】

| | |
|----------|------------------------------|
| 診療ガイドライン | 虚血性脳卒中に対するアルテプラーゼ |
| 対象 | 虚血性脳卒中 (79歳までの成人) |
| 介入 | 6時間以内のアルテプラーゼ 0.6mg/kg 静注 |
| 対照 | プラセボ |

*各項目の評価は「高 (-2)」、「中/疑い (-1)」、「低 (0)」の3段階。
 まとめては「高 (-2)」、「中 (-1)」、「低 (0)」の3段階でエビデンス総体に反映させる。
 アウトカムごとに別紙にまとめる。

| アウトカム | 発症6ヵ月後の総死亡 | | | | | | | | | | | | | | | | | | | | | |
|-------|------------|--------|--------|--------|--------|----------|----------------|------------|----|----|------|-------|-----|-------|-------|-------|-------|----------|---------|-------------|--------|-----|
| | バイアスリスク* | | | 非直接性* | | | リスク人数 (アウトカム率) | | | | 信頼区間 | | | | | | | | | | | |
| 個別研究 | 研究コード | 研究デザイン | 選択バイアス | 実行バイアス | 検出バイアス | バイアスリスク* | | | 対象 | 介入 | 対照 | アウトカム | まとめ | 対照群分母 | 対照群分子 | 介入群分母 | 介入群分子 | 効果指標 (値) | 効果指標 種類 | | | |
| | | | | | | ランダム化 | コンシールメント | 選択的アウトカム報告 | | | | | | | | | | | | 早期試験中止 | その他のアス | (%) |
| | EST-3 | RCT | 0 | 0 | 0 | 0 | 0 | 0 | -1 | -1 | 0 | 0 | -1 | 1520 | 407 | 1515 | 408 | 1.006 | RR | 0.89 - 1.13 | | |
| | EPHETH | RCT | 0 | 0 | 0 | 0 | 0 | 0 | -1 | -1 | 0 | -1 | -1 | 49 | 7 | 52 | 13 | 1.75 | RR | 0.76 - 4.02 | | |
| | ECASS II | RCT | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | -1 | -1 | 391 | 42 | 409 | 43 | 0.979 | RR | 0.65 - 1.46 | | |
| | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | |

コメント (該当するセルに記入)

| | | | | | | | | | | | | | | | | | | | | | | | |
|----------|--|--|--------------|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| EST-3 | | | | | | | | | | | | | | | | | | | | | | | |
| EPHETH | | | 4名のプロックランダム化 | | | | | | | | | | | | | | | | | | | | |
| ECASS II | | | | | | | | | | | | | | | | | | | | | | | |

【SR-7 評価シート エビデンス総体 記入例 (介入研究)】

| | |
|----------|----------------------------|
| 診療ガイドライン | 虚血性脳卒中に対するアルテプラナーゼ |
| 対象 | 虚血性脳卒中 (79歳までの成人) |
| 介入 | 6時間以内のアルテプラナーゼ 0.6mg/kg 静注 |
| 対照 | プラセボ |

*バイアスリスク、非直接性
各ドメインの評価は「高 (+2)」、「中/疑い (-1)」、「低 (0)」の3段階。
まともは「高 (+2)」、「中 (-1)」、「低 (0)」の3段階でエビデンス総体に反映させる。
** 上昇要因
各項目の評価は「高 (+2)」、「中 (+1)」、「低 (0)」の3段階。
まともは「高 (+2)」、「中 (+1)」、「低 (0)」の3段階でエビデンス総体に反映させる。
アウトカムごとに別紙にまとめる。

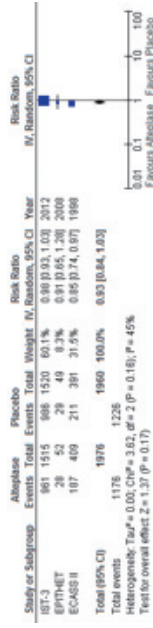
エビデンス総体

| アウトカム | リスク人数 (アウトカム率) | | | | | | | | | | コメント | | | | |
|-------------------|----------------|---------|------|------|------|----------------|-------------|-------|--------------|--------------|--------------|-------------|---|---|---------------|
| | 研究デザイン / 研究数 | バイアスリスク | 非一貫性 | 不精確性 | 非直接性 | その他 (出版バイアスなど) | 上昇要因 (観察研究) | 対照群分母 | 対照群分子 (%) | 介入群分母 (%) | | 介入群分子 (%) | | | |
| 6ヵ月後の総死亡 | RCT / 3 | 0 | 0 | 0 | 0 | -1 | 0 | 1960 | 456 (0.223) | 1976 (0.23) | 464 (0.23) | 0.905-1.139 | A | 9 | 差がない。 |
| 6ヵ月後の依存性 (mRS0-2) | RCT / 3 | 0 | 0 | 0 | 0 | -1 | 0 | 1960 | 1226 (0.626) | 1976 (0.6) | 1176 (0.6) | 0.827-1.041 | B | 7 | mRS0-2 がアウトカム |
| 1週間以内の脳出血 | RCT / 3 | 0 | 0 | 0 | 0 | -1 | 0 | 1961 | 29.5 (0.015) | 1977 (0.015) | 144.5 (0.07) | 2.013-9.777 | A | 9 | 脳出血例のほとんどが死亡 |

総死亡



依存性



頭蓋内出血



【SR-7 評価シート エビデンス総体 記入例 (観察研究)】

| | |
|----------|--------------------|
| 診療ガイドライン | 自己免疫性肝炎 |
| 対象 | 自己免疫性肝炎 |
| 介入 | PSL+AZA または PSL 単独 |
| 対照 | 無治療 |

* バイアスリスク、非直接性
各ドメインの評価は「高 (-2)」、「中/疑い (-1)」、「低 (0)」の3段階。
まとめは「高 (-2)」、「中 (-1)」、「低 (0)」の3段階でエビデンス総体に反映させる。
** 上昇要因
各項目の評価は「高 (+2)」、「中 (+1)」、「低 (0)」の3段階。
まとめは「高 (+2)」、「中 (+1)」、「低 (0)」の3段階でエビデンス総体に反映させる。
アウトカムごとに別紙にまとめる。

| エビデンス総体 | リスク人数 (アウトカム率) | | | | | | | | | | コメント | | | | | | | |
|----------------------|----------------|---------|------|------|------|----------------|-------------|-------|-------|-----|------|-------|-------|-----|-----------|---------|----------|----------|
| | 研究デザイン / 研究数 | バイアスリスク | 非一貫性 | 不精確性 | 非直接性 | その他 (出版バイアスなど) | 上昇要因 (観察研究) | 対照群分母 | 対照群分子 | (%) | | 介入群分母 | 介入群分子 | (%) | 効果指標 (種類) | 効果指標統合値 | 95% 信頼区間 | エビデンスの強さ |
| アウトカム 生化学的・組織学的寛解 | コホート研究 / 3 | -2 | 0 | -1 | -2 | | | 8 | 7 | 0.9 | 540 | 101 | 0.187 | NA | NA | NA | 弱 (C) | 8 |
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |

コメント (該当するセルに記入)

| | | | | | | | | | | | | | | | | | | | |
|--|-----------------------------|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| | Czaja 2008 の研究の介入は1つにまとめられる | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |

[SR-10 メタアナリシス 記入例]

| | | | |
|--|--|--|----------------------------------|
| CQ | | 成人の急性虫垂炎において抗菌薬投与による保存的治療は外科的虫垂切除術と比べ1年以内の再発を伴わない治療が改善するか？ | |
| P | 成人の急性虫垂炎 | I | 抗菌薬投与による保存的治療 |
| C | 外科手術 | O | 1ヵ月以降1年以内の再発 |
| 研究デザイン | RCT | 文献数 | 4 |
| コード | Salminen P 2015 Vons C 2011 Styrud J 2006 Eriksson S 1995 | | |
| モデル | ランダム効果モデル | 方法 | 分散逆数法 |
| 効果指標 | リスク比 (RR) | 統合値 | 41.399 (10.260-167.05) P=<0.0001 |
| 効果指標 | リスク差 (RD) | 統合値 | 0.221 (0.14-0.303) P=<0.0001 |
| フォレスト プロット | | | |
| | | | |
| コメント： | | | |
| ファンネル プロット | | | |
| | | | |
| コメント： ファンネルプロットからは出版バイアスは見られない。 | | | |
| その他の解析 | | コメント： | |
| メタリグ レッション | | | |
| 感度分析 | | | |

【SR-12 結果のまとめ (SoF 表) (ペア比較のメタアナリシス) 記入例】

重要臨床課題：成人の急性虫垂炎において抗菌薬投与による保存的治療は外科的虫垂切除術と比べ推奨できるか？

疾患／対象者：18歳以上の成人で、急性虫垂炎で穿孔、膿瘍形成などの合併症をとまなわないもの

セッティング：医療体制の確立した地域

介入：アモキシシリン＋クラブラン酸の抗菌薬投与による保存的治療

対照：外科的虫垂切除（開腹術および腹腔鏡下手術は問わない）

| アウトカム 対象者数 (研究数) | 相対効果 (95%信頼区間) | 期待される絶対効果* (95%信頼区間) | | | エビデンス 確実性 | 何が 起きるか？ |
|--------------------------------------|-----------------------------------|----------------------|---------------|---------------------------------------|--|---|
| | | 対照 | 介入 | 差 | | |
| 虫垂炎の再発 1年 949名 (4件) | リスク比 41.39 (10.26 ~ 167.05) | 100.0% | 77.80% | 22.1% 多い (14.0 ~ 30.3% 多い) | ⊕⊕⊕⊕ 高 バイアスリスク、非直接 性が認められる1,2が絶 対リスク減少は0.14 ~ 0.303で、臨床的閾値を 0.05に設定した場合確実 性は高い | 抗菌薬投与は治療率が低 く、1ヵ月以内に手術を 受けなくて済んでも、1 年以内に再発して手術を 受けることになる例が 14 ~ 30%程度ある。 |
| 虫垂切除術 (1ヵ月以内) 1年 1061名 (4件) | リスク比 0.09 (0.061 ~ 0.143) | 99.8% | 8.60% | 90.4% 多い (86.3% ~ 94.4% 少ない) | ⊕⊕⊕⊖ 中 バイアスリスク、非直接 性があるが1,2閾値を 85%少ないに設定すると 確実性は中等度と考 える。 | 抗菌薬投与で最初の1ヵ 月間手術を受けなくて済 む者が90%程度いる。 |
| 主要な合併症 1年 969名 (4件) | リスク比 0.58 (0.355 ~ 0.955) | 8.0% | 5.00% | 2.5% 少ない (6.3% 少ない ~ 1.2% 多い) | ⊕⊕⊖⊖ 低 バイアスリスク、非直接 性があり1,2閾値2% 少ない~2%多いに設定 すると、範囲から外れる 確率がかなり高く、確実 性は低いと考える。 | 各割り付け群全体でみ ると、抗菌薬投与群のほう が手術を受けなくて済む 者が多いので、合併症は 少なくなる。 |
| その他の合併 症 1年 949名 (4件) | リスク比 0.22 (0.088 ~ 0.540) | 13.0% | 2.00% | 9% 少ない (17.2% 少 ない~2% 少 ない) | ⊕⊕⊕⊖ 中 バイアスリスク、非直 接性があるが1,2閾 値を2%少ないに設定 すると確実性は中等度 と考える。 | 各割り付け群全体でみ ると、抗菌薬投与群のほう が手術を受けなくて済 む者が多いので、合併症 は少なくなる。 |
| 入院期間 0 ~ 14日 1050名 (4件) | — | 平均値 2.83日 | 平均値 3.23日 | 0.39日長い (0.25日延長 ~ 0.59日延 長) | ⊕⊕⊕⊖ 中 バイアスリスク、非直 接性1,2が認められるが、 1日以上延長の可能性 は低い。 | 抗菌薬投与の方が入院期 間がわずかに長くなる。 |
| 病休期間 0 ~ 14日 1017名 (3件) | — | 平均値 13.8日 | 平均値 9.043日 | 3.6日短い (7.93日短縮 ~ 0.74日延 長) | ⊕⊕⊖⊖ 低 バイアスリスク、非直 接性、非一貫性1,2,3が 認められる。 | 抗菌薬投与の方が病休期 間が短くなる可能性があ る |

*略語：CI (信頼区間)

解説

- 1 ランダム割り付け、コンシールメントに問題がある研究が多く、また、手術の施行に伴う実行・検出の盲検化ができないため、バイアスリスクに問題がある。
- 2 合併症を伴う症例が多く含まれているために、古い研究では診断法に問題があり、日本の現状と比べ、より進行した例が多く対象に含まれている可能性が高いため、非直接性に問題がある研究が多い。抗菌薬の種類も研究により異なる。
- 3 Forest plot で信頼区間の重ならない研究があり、 I^2 統計量が大きく非一貫性に問題がある。

4章 文献

- Aberlegg SK, et al. (2010) Delta inflation: a bias in the design of randomized controlled trials in critical care medicine. *Crit Care* 14:R77. doi: [10.1186/cc8990](https://doi.org/10.1186/cc8990). PMID: [20429873](https://pubmed.ncbi.nlm.nih.gov/20429873/)
- Agency for Healthcare Research and Quality (AHRQ). (2012) *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. AHRQ Publication No. 10(12)-EHC063-EF. <https://effectivehealthcare.ahrq.gov/> (参照 2021/02/24)
- Als-Nielsen B, et al. (2003) Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 290:921-928. doi: [10.1001/jama.290.7.921](https://doi.org/10.1001/jama.290.7.921). PMID: [12928469](https://pubmed.ncbi.nlm.nih.gov/12928469/)
- Arber M, et al. (2017) Which information sources should be used to identify studies for systematic reviews of economic evaluations in healthcare? *Value in Health* 20:PA738.
- Armijo-Olivo S, et al. (2015) What is the influence of randomisation sequence generation and allocation concealment on treatment effects of physical therapy trials? A meta-epidemiological study. *BMJ Open* 5:e008562. doi: [10.1136/bmjopen-2015-008562](https://doi.org/10.1136/bmjopen-2015-008562). PMID: [26338841](https://pubmed.ncbi.nlm.nih.gov/26338841/)
- Armijo-Olivo S, et al. (2017) Blinding in physical therapy trials and its association with treatment effects: a meta-epidemiological study. *Am J Phys Med Rehabil* 96:34-44. doi: [10.1097/PHM.0000000000000521](https://doi.org/10.1097/PHM.0000000000000521). PMID: [27149591](https://pubmed.ncbi.nlm.nih.gov/27149591/)
- Atkins D, et al. (2011) *Assessing the Applicability of Studies When Comparing Medical Interventions*. Agency for Healthcare Research and Quality. *Methods Guide for Comparative Effectiveness Reviews*. AHRQ Publication No. 11-EHC019-EF. <https://effectivehealthcare.ahrq.gov/> (参照 2021/02/24)
- Bafeta A, et al. (2012) Impact of single centre status on estimates of intervention effects in trials with continuous outcomes: meta-epidemiological study. *BMJ* 344:e813. doi: [10.1136/bmj.e813](https://doi.org/10.1136/bmj.e813). PMID: [22334559](https://pubmed.ncbi.nlm.nih.gov/22334559/)
- Balshem H, et al. (2011) GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 64:401-406. doi: [10.1016/j.jclinepi.2010.07.015](https://doi.org/10.1016/j.jclinepi.2010.07.015). PMID: [21208779](https://pubmed.ncbi.nlm.nih.gov/21208779/)
- Bassler D, et al. (2013) Reflections on meta-analyses involving trials stopped early for benefit: is there a problem and if so, what is it? *Stat Methods Med Res* 22:159-168. doi: [10.1177/0962280211432211](https://doi.org/10.1177/0962280211432211). [Epub 2011 Dec 13](https://pubmed.ncbi.nlm.nih.gov/22170891/). PMID: [22170891](https://pubmed.ncbi.nlm.nih.gov/22170891/)
- Berkman ND, et al. (2014) *The Empirical Evidence of Bias in Trials Measuring Treatment Differences*. Agency for Healthcare Research and Quality. PMID: [25392898](https://pubmed.ncbi.nlm.nih.gov/25392898/)
- Bialy L, et al. (2014) A meta-epidemiological study to examine the association between bias and treatment effects in neonatal trials. *Evid Based Child Health* 9:1052-1059. doi: [10.1002/ebch.1985](https://doi.org/10.1002/ebch.1985). PMID: [25504975](https://pubmed.ncbi.nlm.nih.gov/25504975/)
- Brignardello-Petersen R, et al. (2019) GRADE approach to rate the certainty from a network meta-analysis: addressing incoherence. *J Clin Epidemiol* 108:77-85. doi: [10.1016/j.jclinepi.2018.11.025](https://doi.org/10.1016/j.jclinepi.2018.11.025). [Epub 2018 Dec 5](https://pubmed.ncbi.nlm.nih.gov/30529648/). PMID: [30529648](https://pubmed.ncbi.nlm.nih.gov/30529648/)
- Britton A, et al. (1999) Threats to applicability of randomised trials: exclusions and selective participation.

- J Health Serv Res Policy 4:112-121. doi: [10.1177/135581969900400210](https://doi.org/10.1177/135581969900400210). PMID: [10387403](https://pubmed.ncbi.nlm.nih.gov/10387403/)
- Cancer Research and Biostatistics (CRAB). Statistical Tools. <https://stattools.crab.org/> (参照 2021/02/24)
- Carrasco-Labra A, et al. (2015) Comparison between the standard and a new alternative format of the Summary-of-Findings tables in Cochrane review users: study protocol for a randomized controlled trial. *Trials* 16:164. doi: [10.1186/s13063-015-0649-6](https://doi.org/10.1186/s13063-015-0649-6). PMID: [25873338](https://pubmed.ncbi.nlm.nih.gov/25873338/)
- Carrasco-Labra A, et al. (2016) Improving GRADE evidence tables part 1: a randomized trial shows improved understanding of content in summary of findings tables with a new format. *J Clin Epidemiol* 74:7-18. doi: [10.1016/j.jclinepi.2015.12.007](https://doi.org/10.1016/j.jclinepi.2015.12.007). Epub 2016 Jan 11. PMID: [26791430](https://pubmed.ncbi.nlm.nih.gov/26791430/)
- Centre for Clinical Research and Biostatistics (CCRB). (2017) Sample Size Calculation. <http://www2.ccrb.cuhk.edu.hk/web/> (参照 2021/02/24)
- Chalmers TC, et al. (1983) Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 309:1358-1361. doi: [10.1056/NEJM198312013092204](https://doi.org/10.1056/NEJM198312013092204). PMID: [6633598](https://pubmed.ncbi.nlm.nih.gov/6633598/)
- Chan AW, et al. (2004a) Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 291:2457-2465. doi: [10.1001/jama.291.20.2457](https://doi.org/10.1001/jama.291.20.2457). PMID: [15161896](https://pubmed.ncbi.nlm.nih.gov/15161896/)
- Chan AW, et al. (2004b) Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 171:735-740. doi: [10.1503/cmaj.1041086](https://doi.org/10.1503/cmaj.1041086). PMID: [15451835](https://pubmed.ncbi.nlm.nih.gov/15451835/)
- Chang SM, et al. (2012) Methods Guide for Medical Test Reviews [Internet]. Agency for Healthcare Research and Quality (US) PMID: 22834019. <https://www.ncbi.nlm.nih.gov/books/NBK98241/> (参照 2021/02/24)
- Chess LE, et al. (2013) Risk of bias of randomized controlled trials published in orthopaedic journals. *BMC Med Res Methodol* 13:76. doi: [10.1186/1471-2288-13-76](https://doi.org/10.1186/1471-2288-13-76). PMID: [23758875](https://pubmed.ncbi.nlm.nih.gov/23758875/)
- Clark T, et al. (2013) Sample size determinations in original research protocols for randomised clinical trials submitted to UK research ethics committees: review. *BMJ* 346:f1135. doi: [10.1136/bmj.f1135](https://doi.org/10.1136/bmj.f1135). PMID: [23518273](https://pubmed.ncbi.nlm.nih.gov/23518273/)
- Dechartres A, et al. (2011) Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. *J Clin Epidemiol* 64:136-144. doi: [10.1016/j.jclinepi.2010.04.015](https://doi.org/10.1016/j.jclinepi.2010.04.015). PMID: [20705426](https://pubmed.ncbi.nlm.nih.gov/20705426/)
- Deeks JJ, et al. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration. <http://methods.cochrane.org/sdt/handbook-dta-reviews> (参照 2021/02/24)
- Dias S, et al. TSDs completed or currently in progress. <http://nicedsu.org.uk/technical-support-documents/technical-support-documents/> (参照 2021/02/24)
- Dias S, et al. (2010) Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* 29:932-944. doi: [10.1002/sim.3767](https://doi.org/10.1002/sim.3767). PMID: [20213715](https://pubmed.ncbi.nlm.nih.gov/20213715/)
- Dias S, et al. (2011) NICE DSU Technical Support Document 2: A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. <http://www.nicedsu.org.uk> (参照 2021/02/24)

- Dias S, et al. (2018) Network Meta-Analysis for Decision Making. John Wiley & Sons Ltd.
- Driessen E, et al. (2015) Does publication bias inflate the apparent efficacy of psychological treatment for major depressive disorder? A systematic review and meta-analysis of US national institutes of health-funded trials. PLoS One 10:e0137864. doi: [10.1371/journal.pone.0137864](https://doi.org/10.1371/journal.pone.0137864). PMID: [26422604](https://pubmed.ncbi.nlm.nih.gov/26422604/)
- Dwan K, et al. (2010) Assessing the potential for outcome reporting bias in a review: a tutorial. Trials 11:52. doi: [10.1186/1745-6215-11-52](https://doi.org/10.1186/1745-6215-11-52). PMID: [20462436](https://pubmed.ncbi.nlm.nih.gov/20462436/)
- Fenwick J, et al. (2008) The effect of bias on the magnitude of clinical outcomes in periodontology: a pilot study. J Clin Periodontol 35:775-782. doi: [10.1111/j.1600-051X.2008.01291.x](https://doi.org/10.1111/j.1600-051X.2008.01291.x). PMID: [18840153](https://pubmed.ncbi.nlm.nih.gov/18840153/)
- Gagnier JJ, et al. (2012) Investigating clinical heterogeneity in systematic reviews: a methodologic review of guidance in the literature. BMC Med Res Methodol 12:111. doi: [10.1186/1471-2288-12-111](https://doi.org/10.1186/1471-2288-12-111). PMID: [22846171](https://pubmed.ncbi.nlm.nih.gov/22846171/)
- Gordon M. et al. (2020) forestplot: Advanced Forest Plot Using 'grid' Graphics. <https://cran.r-project.org/web/packages/forestplot/index.html> (参照 2021/02/24)
- GRADEpro GDT. (2021) <https://gradepro.org/> (参照 2021/02/24)
- Gurusamy KS, et al. (2009) Assessment of risk of bias in randomized clinical trials in surgery. Br J Surg 96:342-349. doi: [10.1002/bjs.6558](https://doi.org/10.1002/bjs.6558). PMID: [19283747](https://pubmed.ncbi.nlm.nih.gov/19283747/)
- Guyatt G, et al. (2011a) GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 64:383-394. doi: [10.1016/j.jclinepi.2010.04.026](https://doi.org/10.1016/j.jclinepi.2010.04.026). PMID: [21195583](https://pubmed.ncbi.nlm.nih.gov/21195583/)
- Guyatt GH, et al. (2011b) GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 64:395-400. doi: [10.1016/j.jclinepi.2010.09.012](https://doi.org/10.1016/j.jclinepi.2010.09.012). PMID: [21194891](https://pubmed.ncbi.nlm.nih.gov/21194891/)
- Guyatt GH, et al. (2011c) GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). J Clin Epidemiol 64:407-415. doi: [10.1016/j.jclinepi.2010.07.017](https://doi.org/10.1016/j.jclinepi.2010.07.017). PMID: [21247734](https://pubmed.ncbi.nlm.nih.gov/21247734/)
- Guyatt GH, et al. (2011d) GRADE guidelines: 5. Rating the quality of evidence--publication bias. J Clin Epidemiol 64:1277-1282. doi: [10.1016/j.jclinepi.2011.01.011](https://doi.org/10.1016/j.jclinepi.2011.01.011). PMID: [21802904](https://pubmed.ncbi.nlm.nih.gov/21802904/)
- Guyatt GH, et al. (2011e) GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol 64:1283-1293. doi: [10.1016/j.jclinepi.2011.01.012](https://doi.org/10.1016/j.jclinepi.2011.01.012). PMID: [21839614](https://pubmed.ncbi.nlm.nih.gov/21839614/)
- Guyatt GH, et al. (2011f) GRADE guidelines: 7. Rating the quality of evidence--inconsistency. J Clin Epidemiol 64:1294-1302. doi: [10.1016/j.jclinepi.2011.03.017](https://doi.org/10.1016/j.jclinepi.2011.03.017). PMID: [21803546](https://pubmed.ncbi.nlm.nih.gov/21803546/)
- Guyatt GH, et al. (2011g) GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol 64:1303-1310. doi: [10.1016/j.jclinepi.2011.04.014](https://doi.org/10.1016/j.jclinepi.2011.04.014). PMID: [21802903](https://pubmed.ncbi.nlm.nih.gov/21802903/)
- Guyatt GH, et al. (2011h) GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol 64:1311-1316. doi: [10.1016/j.jclinepi.2011.06.004](https://doi.org/10.1016/j.jclinepi.2011.06.004). PMID: [21802902](https://pubmed.ncbi.nlm.nih.gov/21802902/)
- Guyatt GH, et al. (2012) GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. J Clin Epidemiol 66:158-172. doi: [10.1016/j.jclinepi.2012.01.012](https://doi.org/10.1016/j.jclinepi.2012.01.012). PMID: [22609141](https://pubmed.ncbi.nlm.nih.gov/22609141/)
- Guyatt G, et al. (2013a) GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. J Clin Epidemiol 66:151-157. doi: [10.1016/j.jclinepi.2012.01.006](https://doi.org/10.1016/j.jclinepi.2012.01.006). PMID: [22542023](https://pubmed.ncbi.nlm.nih.gov/22542023/)
- Guyatt GH, et al. (2013b) GRADE guidelines: 13. Preparing summary of findings tables and evidence

- profiles-continuous outcomes. *J Clin Epidemiol* 66:173-183. doi: [10.1016/j.jclinepi.2012.08.001](https://doi.org/10.1016/j.jclinepi.2012.08.001). PMID: [23116689](https://pubmed.ncbi.nlm.nih.gov/23116689/)
- Hahn S, et al. (2000) Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. *Stat Med* 19:3325-3336. doi: [10.1002/1097-0258\(20001230\)19:24<3325::aid-sim827>3.0.co;2-d](https://doi.org/10.1002/1097-0258(20001230)19:24<3325::aid-sim827>3.0.co;2-d). PMID: [11122498](https://pubmed.ncbi.nlm.nih.gov/11122498/)
- Hamm MP, et al. (2010) A descriptive analysis of a representative sample of pediatric randomized controlled trials published in 2007. *BMC Pediatr* 10:96. doi: [10.1186/1471-2431-10-96](https://doi.org/10.1186/1471-2431-10-96). PMID: [21176224](https://pubmed.ncbi.nlm.nih.gov/21176224/)
- Hartling L, et al. (2011) Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma. *PLoS One* 6:e17242. doi: [10.1371/journal.pone.0017242](https://doi.org/10.1371/journal.pone.0017242). PMID: [21390219](https://pubmed.ncbi.nlm.nih.gov/21390219/)
- Hartling L, et al. (2013) Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. *J Clin Epidemiol* 66:973-981. doi: [10.1016/j.jclinepi.2012.07.005](https://doi.org/10.1016/j.jclinepi.2012.07.005). PMID: [22981249](https://pubmed.ncbi.nlm.nih.gov/22981249/)
- Hawkins N, et al. (2016) ‘Arm-based’ parameterization for network meta-analysis. *Res Synth Methods* 7:306-313. doi: [10.1002/jrsm.1187](https://doi.org/10.1002/jrsm.1187). PMID: [26610409](https://pubmed.ncbi.nlm.nih.gov/26610409/)
- Haynes RB, et al. (2004) Optimal search strategies for retrieving scientifically strong studies of diagnosis from MEDLINE: analytical survey. *BMJ* 328:1040. doi: [10.1136/bmj.38068.557998.EE](https://doi.org/10.1136/bmj.38068.557998.EE). PMID: [15073027](https://pubmed.ncbi.nlm.nih.gov/15073027/).
- Hedin RJ, et al. (2016) Publication bias and nonreporting found in majority of systematic reviews and meta-analyses in anesthesiology journals. *Anesth Analg* 123:1018-1025. doi: [10.1213/ANE.00000000000001452](https://doi.org/10.1213/ANE.00000000000001452). PMID: [27537925](https://pubmed.ncbi.nlm.nih.gov/27537925/)
- Herbison P, et al. (2011) Different methods of allocation to groups in randomized trials are associated with different levels of bias. A meta-epidemiological study. *J Clin Epidemiol* 64:1070-1075. doi: [10.1016/j.jclinepi.2010.12.018](https://doi.org/10.1016/j.jclinepi.2010.12.018). PMID: [21474279](https://pubmed.ncbi.nlm.nih.gov/21474279/)
- Higgins JP, et al. (2011a) The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials. *BMJ* 343:d5928. doi: [10.1136/bmj.d5928](https://doi.org/10.1136/bmj.d5928). PMID: [22008217](https://pubmed.ncbi.nlm.nih.gov/22008217/)
- Higgins JPT, et al. (2011b) *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. <http://handbook-5-1.cochrane.org/> (参照 2021/02/24)
- Higgins JPT, et al. (2019) RoB 2 tool A revised Cochrane risk of bias tool for randomized trials. <https://www.riskofbias.info/welcome/rob-2-0-tool> (参照 2021/02/24)
- Hong H, et al. (2015) Incorporation of individual-patient data in network meta-analysis for multiple continuous endpoints, with application to diabetes treatment. *Stat Med* 34:2794–2819. doi: [10.1002/sim.6519](https://doi.org/10.1002/sim.6519). PMID: [25924975](https://pubmed.ncbi.nlm.nih.gov/25924975/)
- Hong H, et al. (2016) A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Res Synth Methods* 7:6–22. doi: [10.1002/jrsm.1153](https://doi.org/10.1002/jrsm.1153). PMID: [26536149](https://pubmed.ncbi.nlm.nih.gov/26536149/)
- Howard BE, et al. (2016) SWIFT-review: a text-mining workbench for systematic review. *Syst Rev* 5:87. doi: [10.1186/s13643-016-0263-z](https://doi.org/10.1186/s13643-016-0263-z). PMID: [27216467](https://pubmed.ncbi.nlm.nih.gov/27216467/)
- Hróbjartsson A, et al. (2012) Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ* 344:e1119. doi: [10.1136/bmj](https://doi.org/10.1136/bmj).

[e1119](#). PMID: [22371859](#)

Hughes MD, et al. (1992) The impact of stopping rules on heterogeneity of results in overviews of clinical trials. *Biometrics* 48:41-53. PMID: [1581492](#)

Hultcrantz M, et al. (2017) The GRADE working group clarifies the construct of certainty of evidence. *J Clin Epidemiol* 87:4-13. doi: [10.1016/j.jclinepi.2017.05.006](#). PMID: [28529184](#)

Hutton B, et al. (2015) The PRISMA extension statement. *Ann Intern Med* 163:566-567. doi: [10.7326/L15-5144-2](#). PMID: [26436629](#)

Inaba Y, et al. (2009) Impact of single or multicentre study design on the results of trials examining the efficacy of adjunctive devices to prevent distal embolisation during acute myocardial infarction. *EuroIntervention* 5:375-383. doi: [10.4244/v5i3a59](#). PMID: [19736164](#)

Institute of Medicine (IOM). (2011a) *Clinical Practice Guidelines We Can Trust*. National Academies Press

Institute of Medicine (IOM). (2011b) *Finding What Works in Health Care: Standards for systematic reviews*. National Academies Press

Ioannidis JP. (2011) Excess significance bias in the literature on brain volume abnormalities. *Arch Gen Psychiatry* 68:773-780. doi: [10.1001/archgenpsychiatry.2011.28](#). PMID: [21464342](#)

Jüni P, et al. (1999) The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 282:1054-1060. doi: [10.1001/jama.282.11.1054](#). PMID: [10493204](#)

Kirkham JJ, et al. (2010) The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 340:c365. doi: [10.1136/bmj.c365](#). PMID: [20156912](#)

Kjaergard LL, et al. (2001) Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 135:982-989. doi: [10.7326/0003-4819-135-11-200112040-00010](#). PMID: [11730399](#)

Knudsen AB, et al. (2016) Estimation of benefits, burden, and harms of colorectal cancer screening strategies: modeling study for the US preventive services task force. *JAMA* 315:2595-2609. doi: [10.1001/jama.2016.6828](#). PMID: [27305518](#)

Krahn U, et al. (2013) A graphical tool for locating inconsistency in network meta-analyses. *BMC Med Res Methodol* 13:35. doi: [10.1186/1471-2288-13-35](#). PMID: [23496991](#)

Krogh J, et al. (2015) DEPEROR: Risks of systematic errors in drug and non-drug randomized clinical trials assessing intervention effects in patients with unipolar depression. *J Affect Disord* 179:121-127. doi: [10.1016/j.jad.2015.03.042](#). PMID: [25863907](#)

Kunz R, et al. (2007) Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* MR000012. doi: [10.1002/14651858.MR000012.pub2](#). PMID: [17443633](#)

Kyzas PA, et al. (2005) Selective reporting biases in cancer prognostic factor studies. *J Natl Cancer Inst* 97:1043-1055. doi: [10.1093/jnci/dji184](#). PMID: [16030302](#)

Lee A, et al. (2006) Publication bias affected the estimate of postoperative nausea in an acupoint stimulation systematic review. *J Clin Epidemiol* 59:980-983. doi: [10.1016/j.jclinepi.2006.02.003](#). PMID: [16895822](#)

- Lefebvre C, et al. (2020) Chapter 4: Searching for and selecting studies. In: Higgins JPT, et al. editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.1. <https://training.cochrane.org/handbook/current> (参照 2021/02/24)
- Lin L, et al. (2017) Performing arm-based network meta-analysis in R with the pcnetmeta package. J Stat Softw 80. doi: [10.18637/jss.v080.i05](https://doi.org/10.18637/jss.v080.i05). PMID: [28883783](https://pubmed.ncbi.nlm.nih.gov/28883783/) [パッケージ : <https://cran.r-project.org/web/packages/pcnetmeta/>] (参照 2021/02/24)
- Linde K, et al. (1999) Impact of study quality on outcome in placebo-controlled trials of homeopathy. J Clin Epidemiol 52:631-636. doi: [10.1016/s0895-4356\(99\)00048-7](https://doi.org/10.1016/s0895-4356(99)00048-7). PMID: [10391656](https://pubmed.ncbi.nlm.nih.gov/10391656/)
- Liu CJ, et al. (2011) Do unblinded assessors bias muscle strength outcomes in randomized controlled trials of progressive resistance strength training in older adults? Am J Phys Med Rehabil 90:190-196. doi: [10.1097/PHM.0b013e31820174b3](https://doi.org/10.1097/PHM.0b013e31820174b3). PMID: [21173683](https://pubmed.ncbi.nlm.nih.gov/21173683/)
- Marshall IJ, et al. (2016) RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. J Am Med Inform Assoc 23:193-201. doi: [10.1093/jamia/ocv044](https://doi.org/10.1093/jamia/ocv044). PMID: [26104742](https://pubmed.ncbi.nlm.nih.gov/26104742/)
- McDonagh M, et al. (2013) Avoiding Bias in Selecting Studies. Methods Guide for Comparative Effectiveness Reviews. (Prepared by the Oregon Health & Science University, McMaster University, and Southern California Evidencebased Practice Centers under Contract No. 290-2007-10057-I.) AHRQ Publication No. 13- EHC045-EF. Agency for Healthcare Research and Quality. https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/methods-guidance-bias_methods.pdf (参照 2021/02/24)
- McGowan J, et al. (2016) PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Elaboration (PRESS E&E). Ottawa: CADTH. https://www.cadth.ca/sites/default/files/pdf/CP0015_PRESS_Update_Report_2016.pdf (参照 2021/03/08)
- McGuire W, et al. (2010) The effect of fluconazole prophylaxis in very low birthweight infants is overestimated by before-after studies. Arch Dis Child Fetal Neonatal Ed 95:F182-187. doi: [10.1136/adc.2009.160796](https://doi.org/10.1136/adc.2009.160796). PMID: [19822526](https://pubmed.ncbi.nlm.nih.gov/19822526/)
- McInnes MDF, et al. (2018) Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA 319:388-396. doi: [10.1001/jama.2017.19163](https://doi.org/10.1001/jama.2017.19163). PMID: [29362800](https://pubmed.ncbi.nlm.nih.gov/29362800/)
- McLeod BD, et al. (2004) Using dissertations to examine potential bias in child and adolescent clinical trials. J Consult Clin Psychol 72:235-251. doi: [10.1037/0022-006X.72.2.235](https://doi.org/10.1037/0022-006X.72.2.235). PMID: [15065958](https://pubmed.ncbi.nlm.nih.gov/15065958/)
- Mhaskar R, et al. (2012) Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. J Clin Epidemiol 65:602-609. doi: [10.1016/j.jclinepi.2011.10.016](https://doi.org/10.1016/j.jclinepi.2011.10.016). PMID: [22424985](https://pubmed.ncbi.nlm.nih.gov/22424985/)
- Millard LA, et al. (2016) Machine learning to assist risk-of-bias assessments in systematic reviews. Int J Epidemiol 45:266-277. doi: [10.1093/ije/dyv306](https://doi.org/10.1093/ije/dyv306). PMID: [26659355](https://pubmed.ncbi.nlm.nih.gov/26659355/)
- Mills EJ, et al. (2013) Demystifying trial networks and network meta-analysis. BMJ 346:f2914. doi: [10.1136/bmj.f2914](https://doi.org/10.1136/bmj.f2914). PMID: [23674332](https://pubmed.ncbi.nlm.nih.gov/23674332/)
- Morton SC, et al. (2018) Quantitative Synthesis-An Update. Methods Guide for Comparative Effectiveness

- Reviews. (Prepared by the Scientific Resource Center under Contract No. 290-2012-0004-C). AHRQ Publication No. 18-EHC007- EF. Agency for Healthcare Research and Quality. Posted final reports are located on the Effective Health Care Program search page. <https://doi.org/10.23970/AHRQPCMETHGUIDE3> (参照 2021/02/24)
- Niemeyer H, et al. (2012) Publication bias in meta-analyses of the efficacy of psychotherapeutic interventions for schizophrenia. *Schizophr Res* 138:103-112. doi: [10.1016/j.schres.2012.03.023](https://doi.org/10.1016/j.schres.2012.03.023). PMID: [22484024](https://pubmed.ncbi.nlm.nih.gov/22484024/)
- Norris SL, et al. (2012) Selective outcome reporting as a source of bias in reviews of comparative effectiveness. Agency for healthcare research and quality. PMID: [22993870](https://pubmed.ncbi.nlm.nih.gov/22993870/)
- Nüesch E, et al. (2009a) The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. *Arthritis Rheum* 61:1633-1641. doi: [10.1002/art.24894](https://doi.org/10.1002/art.24894). PMID: [19950329](https://pubmed.ncbi.nlm.nih.gov/19950329/)
- Nüesch E, et al. (2009b) The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. *BMJ* 339:b3244. doi: [10.1136/bmj.b3244](https://doi.org/10.1136/bmj.b3244). PMID: [19736281](https://pubmed.ncbi.nlm.nih.gov/19736281/)
- O'Mara-Eves A, et al. (2015) Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Syst Rev* 4:5. doi: [10.1186/2046-4053-4-5](https://doi.org/10.1186/2046-4053-4-5). PMID: [25588314](https://pubmed.ncbi.nlm.nih.gov/25588314/)
- Ouzzani M, et al. (2016) Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 5:210. doi: [10.1186/s13643-016-0384-4](https://doi.org/10.1186/s13643-016-0384-4). PMID: [27919275](https://pubmed.ncbi.nlm.nih.gov/27919275/). <https://rayyan.qcri.org/welcome> (解説動画 : <https://www.youtube.com/watch?v=irAOQgzFMs4>) (参照 2021/02/24)
- Peters JL, et al. (2007) Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Stat Med* 26:4544-4562. doi: [10.1002/sim.2889](https://doi.org/10.1002/sim.2889). PMID: [17476644](https://pubmed.ncbi.nlm.nih.gov/17476644/)
- Pildal J, et al. (2007) Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol* 36:847-857. doi: [10.1093/ije/dym087](https://doi.org/10.1093/ije/dym087). PMID: [17517809](https://pubmed.ncbi.nlm.nih.gov/17517809/)
- Plummer M. (2017) JAGS. <http://mcmc-jags.sourceforge.net/> (参照 2021/02/24)
- Plummer M, et al. (2019) jags: Bayesian Graphical Models using MCMC. <https://cran.r-project.org/web/packages/rjags/> (参照 2021/02/24)
- pmSearch (2019) <https://stat.zanet.biz/sr/pmsearch.htm> (参照 2021/02/24)
- Poorolajal J, et al. (2010) Capture-recapture method for assessing publication bias. *J Res Med Sci* 15:107-115. PMID: [21526067](https://pubmed.ncbi.nlm.nih.gov/21526067/)
- Puhan MA, et al. (2014) A GRADE working group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 349:g5630. doi: [10.1136/bmj.g5630](https://doi.org/10.1136/bmj.g5630). PMID: [25252733](https://pubmed.ncbi.nlm.nih.gov/25252733/)
- Riley RD, et al. (2004) Sensitivity analyses allowed more appropriate and reliable meta-analysis conclusions for multiple outcomes when missing data was present. *J Clin Epidemiol* 57:911-924. doi: [10.1016/j.jclinepi.2004.01.018](https://doi.org/10.1016/j.jclinepi.2004.01.018). PMID: [15504634](https://pubmed.ncbi.nlm.nih.gov/15504634/)
- Roest AM, et al. (2015) Reporting bias in clinical trials investigating the efficacy of second-generation antidepressants in the treatment of anxiety disorders: a report of 2 meta-analyses. *JAMA Psychiatry* 72:500-510. doi: [10.1001/jamapsychiatry.2015.15](https://doi.org/10.1001/jamapsychiatry.2015.15). PMID: [25806940](https://pubmed.ncbi.nlm.nih.gov/25806940/)
- Rücker G, et al. (2019) Network meta-analysis of multicomponent interventions. *Biometrical Journal*

- 62:808-821. doi: [10.1002/bimj.201800167](https://doi.org/10.1002/bimj.201800167). PMID: [31021449](https://pubmed.ncbi.nlm.nih.gov/31021449/) [Rのパッケージ: <https://cran.r-project.org/web/packages/netmeta/> (参照 2021/03/08)]
- Salanti G, et al. Multiple-Treatments Meta-Analysis. <http://www.mtm.uoi.gr/> (参照 2021/02/24)
- Salanti G, et al. (2011) Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* 64:163-171. doi: [10.1016/j.jclinepi.2010.03.016](https://doi.org/10.1016/j.jclinepi.2010.03.016). PMID: [20688472](https://pubmed.ncbi.nlm.nih.gov/20688472/)
- Savovi J, et al. (2012) Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health Technol Assess* 16:1-82. doi: [10.3310/hta16350](https://doi.org/10.3310/hta16350). PMID: [22989478](https://pubmed.ncbi.nlm.nih.gov/22989478/)
- Schou IM, et al. (2013) Meta-analysis of clinical trials with early stopping: an investigation of potential bias. *Stat Med* 32:4859-4874. doi: [10.1002/sim.5893](https://doi.org/10.1002/sim.5893). PMID: [23824994](https://pubmed.ncbi.nlm.nih.gov/23824994/)
- Schulz KF, et al. (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 273:408-412. doi: [10.1001/jama.273.5.408](https://doi.org/10.1001/jama.273.5.408). PMID: [7823387](https://pubmed.ncbi.nlm.nih.gov/7823387/)
- Schünemann H, et al. (2013) GRADE Handbook. <http://gdt.guidelinedevelopment.org/app/handbook/handbook.html> (参照 2021/02/24)
- Schünemann HJ (2016) Interpreting GRADE's levels of certainty or quality of the evidence: GRADE for statisticians, considering review information size or less emphasis on imprecision? *J Clin Epidemiol* 75:6-15. doi: [10.1016/j.jclinepi.2016.03.018](https://doi.org/10.1016/j.jclinepi.2016.03.018). PMID: [27063205](https://pubmed.ncbi.nlm.nih.gov/27063205/)
- Schünemann HJ, et al. (2019a) 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. doi: [10.1016/j.jclinepi.2018.01.012](https://doi.org/10.1016/j.jclinepi.2018.01.012). PMID: [29432858](https://pubmed.ncbi.nlm.nih.gov/29432858/)
- Schünemann HJ, et al. (2019b) GRADE guidelines: 22. The GRADE approach for tests and strategies—from test accuracy to patient-important outcomes and recommendations. *J Clin Epidemiol* 111:69-82. doi: [10.1016/j.jclinepi.2019.02.003](https://doi.org/10.1016/j.jclinepi.2019.02.003). PMID: [30738926](https://pubmed.ncbi.nlm.nih.gov/30738926/)
- Schwarzer G, et al. (2010) Empirical evaluation suggests Copas selection model preferable to trim-and-fill method for selection bias in meta-analysis. *J Clin Epidemiol* 63:282-288. doi: [10.1016/j.jclinepi.2009.05.008](https://doi.org/10.1016/j.jclinepi.2009.05.008). PMID: [19836925](https://pubmed.ncbi.nlm.nih.gov/19836925/)
- Selva A, et al. (2017) Development and use of a content search strategy for retrieving studies on patients' views and preferences. *Health Qual Life Outcomes* 15:126. doi: [10.1186/s12955-017-0698-5](https://doi.org/10.1186/s12955-017-0698-5). PMID: [28851437](https://pubmed.ncbi.nlm.nih.gov/28851437/)
- Shea BJ, et al. (2007) Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 7:10. doi: [10.1186/1471-2288-7-10](https://doi.org/10.1186/1471-2288-7-10). PMID: [17302989](https://pubmed.ncbi.nlm.nih.gov/17302989/)
- Shea BJ, et al. (2017) AMSTAR 2: a critical appraisal tool for systematic reviews that include randomized or non-randomised studies of healthcare interventions, or both. *BMJ* 358:j4008. doi: [10.1136/bmj.j4008](https://doi.org/10.1136/bmj.j4008). PMID: [28935701](https://pubmed.ncbi.nlm.nih.gov/28935701/)
- Shemilt I, et al. (2019) Chapter 20: Economic evidence. In: Higgins JPT, et al. editors. *Cochrane Handbook for Systematic Reviews of Interventions* version 6.0. <https://training.cochrane.org/handbook/archive/>

v6 (参照 2021/02/24)

- Song F, et al. (2000) Publication and related biases. *Health Technol Assess* 4:1-115. PMID: [10932019](#)
- Song F, et al. (2010) Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 14:iii, ix-xi, 1-193. doi: [10.3310/hta14080](#). PMID: [20181324](#)
- Sox HC, et al. (2013) *Medical decision making*, 2nd ed. Wiley-Blackwell
- Sterne JA, et al. (2016) ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 355:i4919. doi: [10.1136/bmj.i4919](#). PMID: [27733354](#). <https://www.riskofbias.info/welcome/home/current-version-of-robins-i> (参照 2021/02/24)
- Sterne JA, et al. (2001) Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 323:101-105. doi: [10.1136/bmj.323.7304.101](#). PMID: [11451790](#)
- Sterne JAC, et al. (2019) RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 366:14898. doi: [10.1136/bmj.14898](#). PMID: [31462531](#)
- Tierney JF, et al. (2007) Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 8:16. doi: [10.1186/1745-6215-8-16](#). PMID: [17555582](#)
- Turner RM, et al. (2009) Bias modelling in evidence synthesis. *J R Stat Soc Ser A Stat Soc* 172:21-47. doi: [10.1111/j.1467-985X.2008.00547.x](#). PMID: [19381328](#)
- Unverzagt S, et al. (2013) Single-center trials tend to provide larger treatment effects than multicenter trials: a systematic review. *J Clin Epidemiol* 66:1271-1280. doi: [10.1016/j.jclinepi.2013.05.016](#). PMID: [23972520](#)
- Vale CL, et al. (2013) Can trial quality be reliably assessed from published reports of cancer trials: evaluation of risk of bias assessments in systematic reviews. *BMJ* 346:f1798. doi: [10.1136/bmj.f1798](#). PMID: [23610376](#)
- van Aert RC, et al. (2016) Conducting meta-analyses based on p values: reservations and recommendations for applying p-uniform and p-curve. *Perspect Psychol Sci* 11:713-729. doi: [10.1177/1745691616650874](#). PMID: [27694466](#)
- van Hoorn R, et al. (2016) The development of PubMed search strategies for patient preferences for treatment outcomes. *BMC Med Res Methodol* 16:88. doi: [10.1186/s12874-016-0192-5](#). PMID: [27473226](#)
- van Tulder MW, et al. (2009) Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. *Spine (Phila Pa 1976)* 34:1685-1692. doi: [10.1097/BRS.0b013e3181ab6a78](#). PMID: [19770609](#)
- van Valkenhoef G, et al. gemtc. <https://cran.r-project.org/web/packages/gemtc/> (参照 2021/02/24)
- van Valkenhoef G, et al. (2016) Automated generation of node-splitting models for assessment of inconsistency in network meta-analysis. *Res Synth Methods* 7:80-93. doi: [10.1002/jrsm.1167](#). PMID: [26461181](#)
- Viechtbauer W. (2020) metafor: Meta-Analysis Package for R. <https://cran.r-project.org/web/packages/metafor/index.html> (参照 2021/02/24)
- Viswanathan M, et al. (2012) Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epidemiol* 65:163-178. doi: [10.1016/j.jclinepi.2011.05.008](#). PMID: [21959223](#)

- Viswanathan M, et al. (2017) Assessing the Risk of Bias in Systematic Reviews of Health Care Interventions. Methods Guide for Comparative Effectiveness Reviews. (Prepared by the Scientific Resource Center under Contract No. 290-2012-0004-C). AHRQ Publication No. 17(18)-EHC036-EF. Rockville, MD: Agency for Healthcare Research and Quality. Posted final reports are located on the Effective Health Care Program search page. <https://doi.org/10.23970/AHRQEPCMETHGUIDE2> (参照 2021/02/24)
- Watzlawick R, et al. (2014) Effect and reporting bias of RhoA/ROCK-blockade intervention on locomotor recovery after spinal cord injury: a systematic review and meta-analysis. JAMA Neurol 71:91-99. doi: [10.1001/jamaneurol.2013.4684](https://doi.org/10.1001/jamaneurol.2013.4684). PMID: [24297045](https://pubmed.ncbi.nlm.nih.gov/24297045/)
- Wessels M, et al. (2016) How to identify existing literature on patients' knowledge, views, and values: the development of a validated search filter. J Med Libr Assoc 104:320-324. doi: [10.3163/1536-5050.104.4.014](https://doi.org/10.3163/1536-5050.104.4.014). PMID: [27822157](https://pubmed.ncbi.nlm.nih.gov/27822157/)
- White CM, et al. (2009) Using existing systematic reviews to replace de novo processes in conducting Comparative Effectiveness Reviews. In: Agency for Healthcare Research and Quality (AHRQ). Methods Guide for Comparative Effectiveness Reviews. https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/methods-guidance-de-novo-processes_methods.pdf (参照 2021/02/24)
- White IR, et al. (2019) A comparison of arm-based and contrast-based models for network meta-analysis. Stat Med 38:5197-5213. doi: [10.1002/sim.8360](https://doi.org/10.1002/sim.8360). PMID: [31583750](https://pubmed.ncbi.nlm.nih.gov/31583750/)
- Whiting PF, et al. (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155:529-536. doi: [10.7326/0003-4819-155-8-201110180-00009](https://doi.org/10.7326/0003-4819-155-8-201110180-00009). PMID: [22007046](https://pubmed.ncbi.nlm.nih.gov/22007046/)
- World Health Organization (WHO). (2019) WHO Handbook for Guideline Development. 2nd edition. 17: Developing guideline recommendations for tests or diagnostic tools. <https://apps.who.int/iris/bitstream/handle/10665/145714/9789241548960-chap17-eng.pdf?sequence=20&isAllowed=y> (参照 2021/02/24)
- Wood L, et al. (2008) Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 336:601-605. doi: [10.1136/bmj.39465.451748.AD](https://doi.org/10.1136/bmj.39465.451748.AD). PMID: [18316340](https://pubmed.ncbi.nlm.nih.gov/18316340/)
- Yepes-Nuñez JJ, et al. (2019) Development of the summary of findings table for network meta-analysis. J Clin Epidemiol 115:1-13. doi: [10.1016/j.jclinepi.2019.04.018](https://doi.org/10.1016/j.jclinepi.2019.04.018). PMID: [31055177](https://pubmed.ncbi.nlm.nih.gov/31055177/)
- Zhang J, et al. (2014) Network meta-analysis of randomized clinical trials: reporting the proper summaries. Clin Trials 11:246-262. doi: [10.1177/1740774513498322](https://doi.org/10.1177/1740774513498322). PMID: [24096635](https://pubmed.ncbi.nlm.nih.gov/24096635/)
- Zhang J, et al. (2019) Bayesian hierarchical methods for meta-analysis combining randomized-controlled and single-arm studies. Stat Methods Med Res 28:1293-1310. doi: [10.1177/0962280218754928](https://doi.org/10.1177/0962280218754928). PMID: [29433407](https://pubmed.ncbi.nlm.nih.gov/29433407/)
- Zhang Y, et al. (2017) 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. doi: [10.1186/s12955-017-0621-0](https://doi.org/10.1186/s12955-017-0621-0). PMID: [28460638](https://pubmed.ncbi.nlm.nih.gov/28460638/)
- Zhang E, et al. (2017) TrialSize. <https://cran.r-project.org/web/packages/TrialSize/index.html> (参照

2021/02/24)

- 上岡洋晴 他. (2018) 「AMSTAR 2：ヘルスケア介入のランダム化または非ランダム化研究あるいは両方を含むシステマティック・レビューのための厳密な評価ツール」の紹介と解. 薬理と治療 46:111-113.
- 河合富士美 他. (2017) システマティックレビュー検索式査読のためのガイドライン：PRESS Peer Review of Electronic Search Strategies. 医学図書館 64:75-78.
- 森實敏夫 他. (2015) 特別寄稿 5 診断に関する診療ガイドライン (CPG) の作成. https://minds.icqhc.or.jp/s/guidance_special_articles5_1 (参照 2021/1/28)