

Minds 診療ガイドライン作成マニュアル 2017

第4章
システマティックレビュー

2017年12月27日版

4.0 概要

診療ガイドラインは推奨を作成するために、存在するすべ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のエビデンスの強さを決定することが望ましい。ひとつの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。Minds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。なお、この分類は、GRADEアプローチで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 2017; Turner 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、必要に応じてハンドサーチ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は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を作成することが望ましい。

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(Agency for Healthcare Research and Quality 2012; Berkman 2014; Higgins 2011a; Higgins 2011b; Deeks 2013; Schünemann 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の定義はさまざまに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うるので、完全な同義語ではない。

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）。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」と定義している（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無い、2. 採択された研究に偏りが無い、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）。

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 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 と呼ぶ。

システマティックレビュー

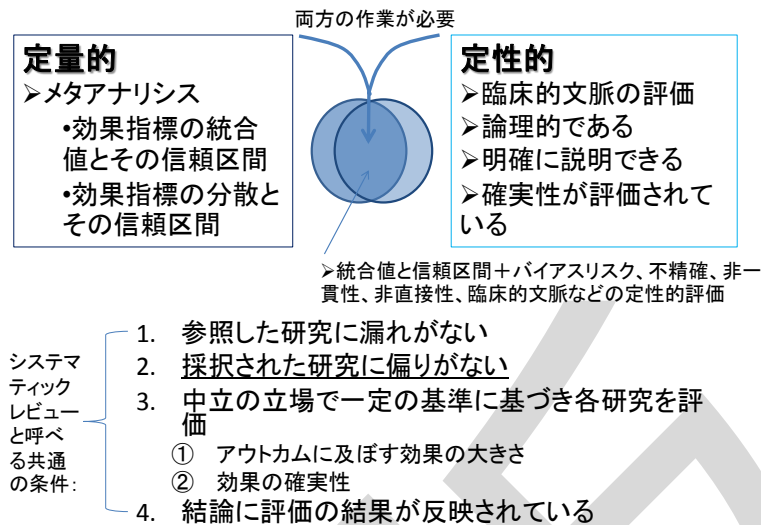


図 4-1 定量的システマティックレビューと定性的システマティックレビュー

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 (systematic review) とは、「クリニカルクエスチョンに対して、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。

コクランレビューをはじめとし、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」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の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。

・PROSPERO(International prospective register of systematic reviews)

<http://www.crd.york.ac.uk/PROSPERO>

システマティックレビューの事前登録をするときには、表 4-1 のような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。

表 4-1 システマティックレビュープロトコール

項目	記載事項	注
対象文献データベース	<input type="checkbox"/> PubMed/Medline <input type="checkbox"/> 医中誌 Web <input type="checkbox"/> The Cochrane Library <input type="checkbox"/> その他 ()	
ハンドサーチ	<input type="checkbox"/> 実施せず <input type="checkbox"/> 実施 対象医学誌： 方法：	
Grey Literature	<input type="checkbox"/> 採用せず <input type="checkbox"/> 採用 対象研究： <input type="checkbox"/> 学会抄録 <input type="checkbox"/> プロシーディングス <input type="checkbox"/> 厚労省班会議資料 <input type="checkbox"/> 行政資料 <input type="checkbox"/> その他 ()	
スクリーニング方法	一次スクリーニング： 二次スクリーニング： 不一致時の対処：	
データの抽出法		
研究デザインの分類	RCT、非ランダム化比較試験、コホート研究、症例対照研究、 横断研究、症例集積、症例報告、その他 ()	
個別研究で評価したバイアスリスクおよびその他の評価項目	選択バイアス (ランダム化、コンシールメント) 実行バイアス (盲検化) 測定バイアス (盲検化) 症例減少バイアス (不完全アウトカム報告) その他 (選択的アウトカム報告) 非直接性 (PICO)	
それぞれの項目の評価法と結果の分類	各ドメインは高、中/疑い、低の3段階 まとめは高、中、低の3段階	
エビデンス総体の評価項目	バイアスリスク、非一貫性、不精確、出版 (報告) バイアス	
メタアナリシスの方法	ランダム効果モデル、固定効果モデル、 その他 (具体的名称)	
用いられた効果指標	リスク比、オッズ比、リスク差 (率差)、NNT、率、感度、特異度、正診率など	
付随した解析	<input type="checkbox"/> 感度分析 <input type="checkbox"/> メタリグレーション <input type="checkbox"/> その他 ()	
メタアナリシスの結果の提示法	<input type="checkbox"/> 効果指標値と95%信頼区間 <input type="checkbox"/> Forest plot <input type="checkbox"/> Funnel plot <input type="checkbox"/> その他 ()	

4.0.2 定性的および定量的システマティックレビュー

・定性的システマティックレビュー

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われわれ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4-2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、比較された介入、バイアスリスクの評価などの記述にとどまらず、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づけている²。

表4-2 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	臨床課題あるいはクリニカルクエストの臨床的文脈における位置づけ。
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、非一貫性、非直接性などを評価する。
4. なぜ結果が異なるかを説明する個々の研究の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亜群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	感受性分析の結果、報告バイアスの評価をする。
9. システマティックレビューの結果が医療界や社会の通念と異なるかを記述する	専門医、患者代表の評価を取り込む。

² 定量的統合あるいは定性的統合という場合は、かならずしもシステマティックという概念を含んでいない。得られるすべ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。

・ 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は分からない場合に、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。

・ 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みなすと、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に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(Juni 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)、うまく機能することが示されているが、高度のスキルが要求されるため、一般化するにいたっていない。

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。たとえば、生存をアウトカムとしてRCTのメタアナリシスによりハザード比0.5、95%信頼区間0.41~0.61という結果が得られていて、効果が高いとみなされる場合でも、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ではなくB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、確実性が低いとみなすことになる。

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は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うる。これらすべ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、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のずれがあるため、バイアスリスク (Guyatt 2011c)、不精確 (Guyatt 2011e)、非一貫性 (Guyatt 2011f)、非直接性 (Guyatt 2011g)、出版（報告）バイアス (Guyatt 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。

たとえば、RCTが1件、症例対照研究が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

コクランレビューをはじめとし、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 4-3 にそれをまとめた。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 4-3 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

項目	診療ガイドライン作成のための SR	通常の SR
目的	益と害の判定のためにエビデンスの強さを明らかにする。	効果の大きさと確実性を明らかにする（効果指標の統合値と信頼区間）。
研究の対象者	属性が限定的でより特異的。	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。	ひとつの益の主要アウトカムを中心に効果が評価されることが多い。
害のアウトカム	・ 益のアウトカムと同じように重要性が評価される。 ・ 観察研究も対象とされ包括的に解析される。	・ 副次的に扱われる。 ・ RCT の統合では頻度の低い害は解析されない。
統合対象研究のデザイン	異なる研究デザインを含む。	ひとつの研究デザイン。
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	ひと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。	アウトカムがひとつであれば比較的単純。
研究の集合	ひとつの介入でアウトカムごとに複数の集合が必要。	ひとつの介入でひとつの集合がある。

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、通常のシステマティックレビューとは異なり、中立的立場からエビデンス総体の強さを評価し推奨の決定を目的としている。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 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。

既に同じ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 2009)。既に同じ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The Appraisal of Guidelines for Research and Evaluation (AGREE II)、A Measurement Tool to Assess systematic Reviews (AMSTAR)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。

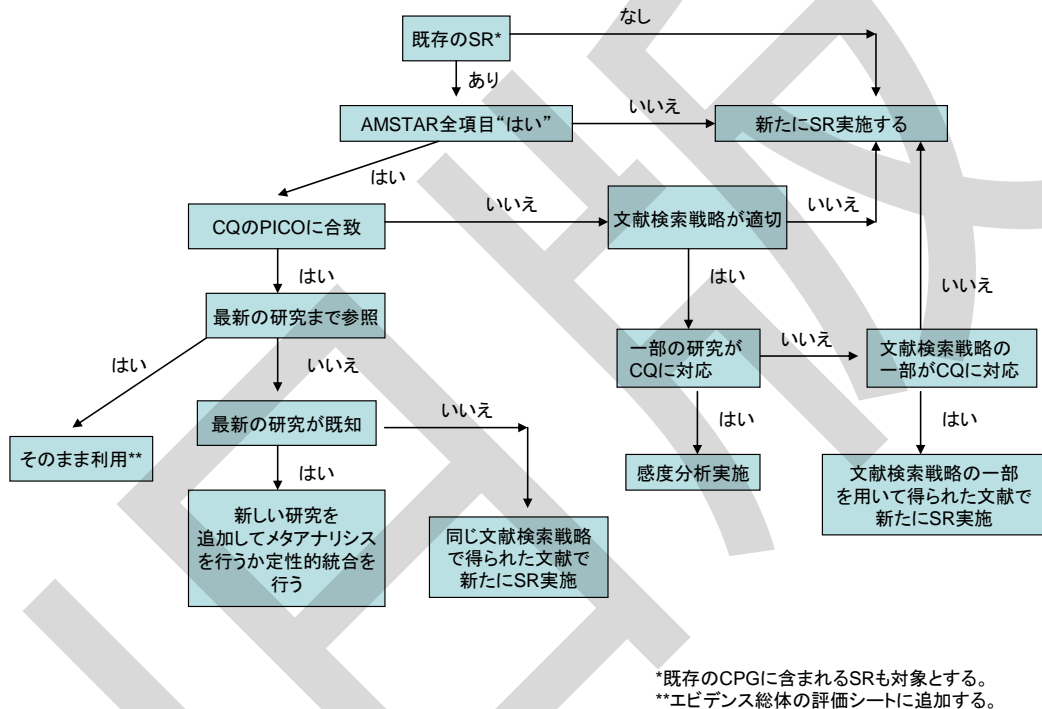


図4-2 既存のシステマティックレビューを利用する場合の方針

- 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もとの研究のデータを再評価し、もとの研究から臨床クエスチョン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。

4.0.5 エビデンス総体の評価 (4.4 参照)

診療ガイドラインの作成においてシステマティックレビューを実施し、診療ガイドラインには、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 (Body of Evidence) の強さと限界を明確に記載することが望ましい。エビデンス総体とは、ある臨床上の問題 (クリニカルクエスション: CQ) に対して収集しえ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、一つの 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 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。ガイドライン作成過程において、CQ に対する推奨診療を提示する場合 (第5章参照) に、エビデンス総体の総括としての強さを一つだけ決定する (Guyatt 2011a)。

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) 臨床の疑問に対し、アウトカムを列挙し、重大、重要なアウトカムを採用

2) 採用された一つ一つのアウトカムに対し以下の作業を行う

3) 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)

4) 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

- (1) バイアスリスク (risk of bias)
- (2) 非直接性 (indirectness)
- (3) 非一貫性 (inconsistency)
- (4) 不精確 (imprecision)
- (5) 出版 (報告) バイアス (publication bias)

各論文に対し
て評価:

STEP1

全論文に対して

5項目評価:

STEP2

5) 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

- (1) 効果が大きい (large effect)
- (2) 用量-反応勾配あり (dose-dependent gradient)
- (3) 可能性のある交絡因子が提示された効果を減弱させている

(plausible confounder)

注意: 各項目の重み付けは CQ によって異なる

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我々の確信が、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 2011)。その際に、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した上で決めることが GRADE Working Group から提言されており (Hultcrantz 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うる。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うえで、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、作成グループ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 Minds 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
- (4) エビデンス総体の評価
- (5) SR 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 先行するガイドラインの検索

国内・海外の先行する診療ガイドラインの検索は、現状把握のためのスコーピングサーチの一環としてスコープ作成時に行うことが望ましい。表 4-6 に示すように、NGC (National Guideline Clearinghouse)、NICE (National Institute for Health and Care Excellence) などを検索する。国内の質の高いガイドラインは Minds (マインズ) ガイドラインライブラリに掲載されているので、特に他の作成団体による先行するガイドラインが公表されている場合は十分に吟味し、整合性について予め検討する。

表 4-6 主な診療ガイドライン検索のためのデータベース

データベース名	組織名	国名	URL	閲覧要件
National Guideline Clearinghouse (NGC)	Agency for Health care Research and Quality (AHRQ)	米国	http://guideline.gov/	
NICE Evidence Search	National Institute for Health Care Excellence (NICE)	英国	http://www.evidence.nhs.uk/	
International Guideline Library	Guidelines International Network (GIN)	国際組織	http://www.g-i-n.net/	要契約
Minds ガイドラインライブラリ	(公財) 日本医療機能評価機構	日本	http://minds.jcqhc.or.jp/	

4.1.2 文献検索データベース

網羅的な検索を行うため、MEDLINE/PubMed³、The Cochrane Library⁴、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 EMBASE⁵や JMEDPlus、社会・心理学的研究の PsycINFO[®]、看護研究の CINAHL など、重要なデータベースも併せて検索する。必要と

³ MEDLINE には PubMed 以外に OVID 等の商用データベースもある。

⁴ The Cochrane Library には The Cochrane Database of Systematic Review (CDSR)、The York Database of Abstracts of Reviews of Effectiveness (DARE)、The Cochrane Controlled Trials Register (CENTRAL) 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。

⁵ ランダム化比較試験に関しては、The Cochrane Library の Cochrane Central Register of Controlled Trials (CENTRAL) から検索可能だがタイムラグがあるので少なくとも最新2年分は検索することが望ましい。

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。特に、事前登録された大規模な未発表のRCTが既知である時には、Grey Literature Report (New York Academy of Sciences) などの情報も積極的に拾い上げる。

表 4-7 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	http://www.ncbi.nlm.nih.gov/pubmed	
	MEDLINE	http://gateway.ovid.com/autologin.html など	要契約
	The Cochrane Library	http://www.thecochranelibrary.com/view/0/index.html	全文参照は要契約
	医中誌 Web	http://www.jamas.or.jp/	要契約
重要	EMBASE	http://www.embase.com/	要契約
	PsycINFO®	http://www.apa.org/pubs/databases/psycinfo/index.aspx	要契約
	CINAHL	http://www.ebscohost.com/	要契約
	JMEDPlus	http://jdream3.com/	要契約
その他の情報源	Grey Literature Report (New York Academy of Sciences)	http://www.greylit.org/home (New York Academy of Sciences)	
	Open Gray	http://www.opengrey.eu/	
	Clinical Trials.gov	http://clinicaltrials.gov/	
	Virtual Health Library	http://regional.bvsalud.org/php/index.php?lang=en	
	World Health Organization (WHO)	http://apps.who.int/trialsearch/	
	UMIN 臨床試験登録	http://www.umin.ac.jp/ctr/index-j.htm	
	国立保健医療科学院	http://rctportal.niph.go.jp/	臨床研究(試験)情報検索

4.1.3 文献検索戦略

4.1.3.1 文献検索概要

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) が独立して立

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、検索日を記載する。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。

適切な検索を行うためには、ガイドライン作成グループと検索担当者との間で重要臨床課題とCQに対する理解の共有が重要となるため、両者の共同作業が重要である。いつ共同作業が必要かは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（たとえば、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うる。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な工夫が必要である。

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が開発されている（McGowan 2016）。検索文献の引用文献、教科書の参照など、文献検索以外に情報収集を行った場合は記録しておく。ただし、検索式以外で得られた文献の追加については委員会での検討が必要である。

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、ハンドサーチが重要な手段となる。すべての検索、文献選択の経過は、テンプレート【4-2】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。

4.1.3.2 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）を用いて網羅的に検索されるが、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0は含めない。論文数が多い場合も言語などのフィルターを用いての絞り込みは原則として行わない。プロトコルに予め記載されている場合には作業量に応じて以下のような絞り込みを行うこともやむを得ない。益のアウトカムに関する検索例を以下に示す。

タイトル:	Stroke and t-PA
発症6時間以内の80歳未満の急性脳塞栓患者に対して (patients)、発症3時間以降のrt-PA投与は(intervention)は推奨されるか?	
データベース:	PubMed
日付:	2017/08/19
検索者:	NK/FK

PubMed 検索例 1 <益の検索>

#	検索式	文献数
1	"acute ischemic stroke" [TIAB]	10074
2	"acute ischaemic stroke" [TIAB]	1550
3	"Brain Ischemia" [Mesh]	93949
4	"Stroke" [Mesh]	106970
5	#1 OR #2 OR #3 OR #4	152123
6	"Tissue Plasminogen Activator" [Mesh]	17046
7	rt-pa[TIAB]	2554
8	alteplase[TW]	1559
9	#6 OR #7 OR #8	18150
10	#5 AND #9	4879
11	#10 AND (Meta-Analysis[PT] OR systematic[SB])	265
12	#10 AND "Randomized Controlled Trial" [PT]	278

PubMed 検索例 2 <益の検索>

#	検索式	文献数
1	"Cerebrovascular Disorders" [Mesh:NoExp]	44796
2	"Brain Ischemia" [Mesh]	93949
3	"Carotid Artery Diseases" [Mesh:NoExp] OR "Carotid Artery Thrombosis" [Mesh]	22895
4	"Stroke" [Mesh:NoExp] OR "Brain Infarction" [Mesh]	106697
5	"Hypoxia-Ischemia, Brain" [Mesh]	4637
6	"Cerebral Arterial Diseases" [Mesh:NoExp] OR "Intracranial Arterial Diseases" [Mesh:NoExp]	2792
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	347096

	attack*[TW] OR "transient ischaemic attack*[TW] OR tia[TW]	
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]	1687529
10	ischemi*[TW] OR ischaemi*[TW] OR infarct*[TW] OR thrombos*[TW] OR emboli*[TW] OR occlus*[TW] OR hypoxi*[TW]	1072537
11	#9 AND #10	214979
12	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #11	475382
13	"Tissue Plasminogen Activator"[Mesh]	17046
14	"PLAT protein, human" [Supplementary Concept]	90
15	plasminogen[TW] OR plasmin[TW] OR tPA[TW] OR t-PA[TW] OR rtPA[TW] OR rt-PA[TW]	74954
16	alteplase[TW]	1559
17	#13 OR #14 OR #15 OR #16	75260
18	#12 AND #17	9971
19	<u>#18 AND (Meta-Analysis[PT] OR systematic[SB])</u>	<u>449</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])	1020982
21	<u>#18 AND #20</u>	<u>1810</u>

4.1.3.3 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、検索式から研究デザイン (RCT) 等の絞込み条件を外し、“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。害に関する検索方法は確立しておらず今後の研究が期待される。また、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。

4.1.3.4 費用対効果の検索の進め方

cost や economics などの Mesh やテキストワードを掛け合わせて検索する。

PubMed 検索例<コストの検索>

#	検索式	文献数
1	"acute ischemic stroke"[TIAB]	10074
2	"acute ischaemic stroke" [TIAB]	1550
3	"Brain Ischemia"[Mesh]	93949
4	"Stroke"[Mesh]	106970
5	#1 OR #2 OR #3 OR #4	152123
6	"Tissue Plasminogen Activator" [Mesh]	17046
7	rt-pa[TIAB]	2554
8	alteplase[TW]	1559
9	#6 OR #7 OR #8	18150
10	#5 AND #9	4879
11	cost[TIAB] OR costs[TIAB] OR economic*[TW] OR "Costs and Cost Analysis"[Mesh]	893717
12	<u>#10 AND #11</u>	<u>177</u>

4.1.3.5 PubMedにおける研究デザインの絞り込み方法

Tagを用いるのが簡便で精密な検索方法である。慣れない場合や予備検索ではフィルターで publication type (PT) を指定する方法と、Clinical Queries を使用する簡易な方法がある。最終的には Tag を用いて検索式を記録することを推奨する。

- 1) フィルターで publication type (PT) を指定する場合は Type of Article のメニューから Clinical Trial、Meta-Analysis、Randomized Controlled Trial (RCT、Clinical Trial に含まれる)、Practice Guideline などを選択し限定することができる。
- 2) PubMed の Clinical Queries (高い質のエビデンス検索のフィルター) は、診療ガイドライン作成のための文献検索では勧められていない。
- 3) Cochrane には PubMed 検索用の RCT フィルターが用意されている (Higgins2011b)。

Box 6.4. a 感度最大化バージョン (2008 revision); PubMed format
 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]

Box 6.4. b 感度・正確度最大化バージョン (2008 revision); PubMed format
 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通りである。

1. 検索式は3つの検索語句のセットからなる。1) 対象（健康状態を表す語句）P、2) 介入を表す語句 I、3) 研究デザイン。
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、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、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こと。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 で結合して用いること。
5. 異なるコンセプトを AND で結合すること。
6. RCT のための感度最大化検索フィルターを最初使用し、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）。

4.1.3.6 PubMed における言語や期間、人間の絞り込み方法

- 1) 動物を除外するとき

NOT ("animals" [MeSH] NOT "humans" [MeSH])

- 2) 英語に絞るとき

(言語による絞り込みは原則としてしないが、やむをえないとき)

AND English [LA]

- 3) 改訂版など、検索期間を出版年で指定するとき

AND "2000" [PDAT] : "2013/01/31" [PDAT]

既存の診療ガイドライン・システマティックレビューに、臨床研究の文献を追加する場合にも、前回検索以降の論文を検索する。その場合は掲載されている検索式を参考にすることができるが、検索式に問題がある場合にはすべての期間で検索し直すことも考慮する。

また、今後患者・介護者の価値観・意向に関する研究、費用対効果分析に関する研究の増加が予測されるが、その際の検索方法、検索フィルターが実用的なレベルで開発される可能性がある (Wessels 2016; van Hoorn 2016; Zhang 2017)。

4.1.4 文献の再検索

文献検索にて、未発表の大規模な研究が把握された場合、または診療ガイドライン

出版までの間に重要な研究が追加された可能性がある場合は、文献検索をアップデートしてもよい。

○手順

- (1) 検索式の作成
- (2) 検索の実行

○テンプレート

- 【4-1 データベース検索結果】
- 【4-2 文献検索フローチャート】

○記入方法

- 【4-1 データベース検索結果 記入方法】
- 【4-2 文献検索フローチャート 記入方法】

○記入例

- 【4-1 データベース検索結果 記入例】
- 【4-2 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SRチーム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。たとえば、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。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SRチーム2名が独立してフルテキストを読み、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、2名の意見が異なる場合は第3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【4-3 二次スクリーニング後の一覧表】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。【4-3 二次スクリーニング後の一覧表】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、或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わかる。

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。

○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)方針にしたがって管理する
- (5)引用文献を記載する

○テンプレート

【4-3 二次スクリーニング後の一覧表】

【4-4 引用文献リスト】

○記入方法

【4-3 二次スクリーニング後の一覧表 記入方法】

【4-4 引用文献リスト 記入方法】

○記入例

【4-3 二次スクリーニング後の一覧表 記入例】

【4-4 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：個々の報告評価（STEP1）

エビデンスを評価し、強さと限界について、明確に提示することが望ましい。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、総体としてのエビデンス（Body of Evidence）の質あるいは限界を明確に記載することが望ましい。

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「エビデンスの収集」4.2「スクリーニング」参照）を基に、以下の論文選出・選択作業を行う。

○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）等の詳細な設定を詳しく評価し、さらにCQに関係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。一致しない領域に対しては、下記の「非直接性（Indirectness）」の評価時に参考にする。また、ランダム化とコンシールメント（割付の隠蔽化）について、再評価が必要である。
- ・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、その後に新たなRCT等が出ている場合は、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。

○システマティックレビューが複数抽出された場合

- ・上の「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。

○除外論文の取り扱い

- ・抽出された論文のうち、除外論文とされたものは、その論文名と除外理由を本文中に記載する。

○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。

○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 個々の報告に対する評価 (STEP 1)

アウトカムごとにまとめられた文献集合の個々の論文について、研究デザインごとに、【4-5 評価シート 介入研究】、【4-6 評価シート 観察研究】を用いて、バイアスリスク、非直接性(indirectness)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。ランダム化比較試験でもバイアスの影響を受けている可能性があるので (Chess 2013; Hamm 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。

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

4.3.2.1.1 原則

以下の Cochrane risk of bias tool による原則 (Higgins 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 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. バイアスリスクの内、重要視すべきドメインあるいは項目は一樣ではない。
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ないか注意する。
7. アウトカムごとに分けて評価を行う。

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 2013; Hartling 2011; Vale 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。

4.3.2.1.2 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一つまたは二つの評価項目が含まれている。

表 4-8 ランダム化比較試験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	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	参加者と医療提供者の盲検化
検出バイアス	アウトカム測定者の盲検化
症例減少バイアス	不完全アウトカムデータ
	ITT 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

各ドメインの概念はランダム化比較試験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 2014; Dechartres 2011; Viswanathan 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 2009）、疾患、介入、アウトカムによって異なる可能性があり（Fenwick 2008）、個別の判断が重要である。

○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が影響を受けるバイアス（Armijo-Olivo 2015; Britton 1999; Savovic 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、平均年齢が高い方が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。

◇項目：ランダム化

- ・ランダム配列の生成

ランダム系列生成：患者の割付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 2001）。

Chalmers ら（1983）の 145 件の急性心筋炎の RCT を検討した報告では、割付を盲検

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。

・ベースラインの不均衡

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、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の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。

◇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。介入現場での割付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（Kuntz 2007; Herbison 2011; Nuesch 2009a; Wood 2008; Pildal 2007; Schulz 1995）。

Herbison ら（2011）は、389 件の RCT を解析した結果、「double blind」の記載があっても、割付の隠蔽化が不明瞭な場合はバイアスリスクが高くなり、Pooled ratios of odds ratios (RORs) は 0.86 (95%CI: 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う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。

○実行バイアス

◇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（Armi jo-Olivo 2017; Krogh 2015; Gurusamy 2009）。

例：ランダム化比較試験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、患者からわからなくする（単盲検）ことに加えて、医療提供者にもわ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。

○検出バイアス

◇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

(Hrobjartsson 2012; Liu 2011)。

例：ランダム化比較試験で測定者が割り付けを知ってしまい、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。

○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 (Nuesch 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。

◇項目：不完全アウトカムデータ

症例の減少した分の症例はアウトカムが不明であると考えられるので、不完全アウトカムデータとしてとらえられる (Riley 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。

◇項目：ITT 解析非実施

ITT 解析は治療企図分析のことで、RCT の統計解析において、脱落例やプロトコー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通りに解析することである。評価統合する報告が RCT の場合は 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。

○その他のバイアスリスク

◇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内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報告バイアスを生じる可能性がある (Bialy 2014; Norris 2012; Ioannidis 2011; Dwan 2010; Kirkham 2010; Kyzas 2005; Chan 2004b; Chan 2004a)。

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

◇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にもかかわらず、あるいは適切に計画された Adaptive study design 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 2013; Schou 2013; Hughes 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。

◇項目：その他のバイアス

上記のバイアス以外のバイアスの可能性。

- ・COI (利益相反) 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・管理等の記載から、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うるか評価する。たとえば、単施設での臨床試験は多施設の臨床試験と異なり (Inaba 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 2013; Bafeta 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 2003)。

*バイアスリスク判定方法

コクランバイアスリスクツール (Higgins 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、今までのコクランレビューでは必須実行になっていない。現在コクランバイアスリスクツール 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ションに No, Probably No, No information, Probably Yes, Yes のいずれかに答えながら、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が予定されている (Higgins 2016)。

コクランバイアスリスクツールでは一つの研究でバイアスリスクのひと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づけ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通りである。

1. それぞれの論文について評価する。

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うる。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

ステップ1の表 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)

4.3.2.2 非直接性(indirectness) : 個々の研究

非直接性にはさまざまな名称が用いられてきたが (Turner 2009; Atkins 2011)⁶、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・条件へ適応しうる程度を示す。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。

非直接性の評価は疾患専門家でないとは困難である。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のアウトカムへの影響を推定する必要がある。個々の研究報告に対する評価を行ったうえで、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 (applicability)

(例) 年齢が異なる : CQ では、85 歳未満を対称にしているが、研究論文には年齢制限を設けていない。

- ・ 介入の違い (applicability)

(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、コントロールか別の介入かが、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。

<非直接性判定方法> 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、もし個別研究については非直接性の評価は実施せず、エビデンス総体の評価の際に

⁶ 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、head-to-head 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。

実施するのであれば、テンプレート 4-3 二次スクリーニング後の一覧表を作成する際に、PICO の各項目に非直接性の判断に必要な情報を入力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。

4.3.2.3 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される。

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様な場合は、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 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) 論文の選択・選出
- (2) 個々の研究報告に対する評価
 - (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

【4-5 評価シート 介入研究】（RCT 評価用）

【4-6 評価シート 観察研究】（観察研究評価用）

○記入方法

【4-5 評価シート 介入研究 記入方法】（RCT 評価用）

【4-6 評価シート 観察研究 記入方法】（観察研究評価用）

○記入例

【4-5 評価シート 介入研究 記入例】(RCT 評価用)

【4-6 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（STEP2）

4.4.1 エビデンス総体とエビデンス総体の総括

（1）エビデンス総体（body of evidence）

ある臨床上の問題（クリニカルクエスチョン：CQ）に対して収集しえ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（body of evidence）と呼ぶ。

このアウトカムは、スコープ作成時に、一つの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。ガイドライン作成過程において、CQに対する推奨診療を提示する場合（第5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）を一つだけ決定する（Guyatt 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。たとえば、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、行なわなくてもよい。

4.4.2. エビデンス総体評価

一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、改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（inconsistency）、不精確（imprecision）、出版バイアス（publication bias）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ひとつのアウトカムが、論文の評価項目（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。
- ・上記ステップ3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され、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（indirectness）については、各個別の論文に対して行われ、かつ、一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（inconsistency）、不精確（imprecision）、出版（報告）バイアス（publication bias）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。
- ・上記5項目を用いて、一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一段階で、表4-6で示す分類で評価が付けられる。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。Minds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4-9の通りであり、GRADEアプローチと同じ分類である。

表 4-9 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の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5章推奨参照）。この過程はいわばSTEP3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、QO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、推奨をささ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。

4.4.3. 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

前述のステップ3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(indirectness)

前作業として行われているステップ3で、各個別研究に対して評価された非直接性をまとめた上で、プールされた論文全体が、現在考えているCQや臨床状況・集団・条件へ適応しう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、コントロールか別の介入かが異なる。

・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。

<非直接性判定方法>

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 (inconsistency)

アウトカムに関連して抽出された全て(複数)の研究をみると、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(heterogeneity)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。

<非一貫性判定方法>

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、結果にバラつきが生じる。ここで言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 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。(バラ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一つとして使用すべきである。コクランハンドブックあるいは GRADE ハンドブック (Schünemann 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る。

<40%	may be low
30 - 60%	may be moderate
50 - 90%	may be substantial
75 - 100%	may be considerable

臨床的な異質性と研究方法の異質性の判断を行ったうえで、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%信頼区間がすべ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 (imprecision)

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。プロットコールに示された予定症例数が達成されていることが必要。

<不精確判定方法>	とても深刻な不精確さあり (-2)
	深刻な不精確さあり (-1)
	不精確さなし (0)

不精確の判断方法については、議論の多いところである (Schünemann 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の定義を用いている (Schünemann 2013; Guyatt 2011e)。

前者の場合、メタアナリシスの結果をひとつの臨床試験の結果とみ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 が満たされない場合は、2000 から 4000 例という症例数の場合以外はレートダウンする。OIS が満たされ、95%信頼区間に含まれない場合は、ダウングレードしない。もし OIS の基準が満たされ、95%信頼区間が無効果を含む (例えば、リスク比の信頼区間が 1.0 を含む) 場合は、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? 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みる、あるいは、イ

イベント率が非常に低くサンプルサイズが非常に大きい(少なくとも2000例、おそらく4000例)を見る。いずれも満たされない場合は、不精確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のパッケージTrialSize (Zhang 2017)、ウェブツールとしてフリーで提供されているCenter for Clinical Trials (CCT) (2017)、Cancer Research And Biostatistics (CRAB): Statistical Tools (CRAB 2017)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。そのためは、一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いことで、不精確の評価を決め、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な段階である。

⑤ 出版(報告)バイアス(publication bias): 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 2011d; Hedin 2016; Driessen 2015; Roest 2015; Niemeyer 2012; Poorolajal 2010; Mcguire 2010; Lee 2006; McLeod 2004; Song 2000; Song 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 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(Aberregg 2010)。亜群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 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。

- <出版バイアス判定方法>
- とても深刻な出版バイアスあり (-2)
 - 深刻な出版バイアスあり (-1)
 - 出版バイアスなし (0)

出版バイアスの評価は、メタアナリシスの際に作成するFunnel plotの目視、およびEgger、Beg検定結果で $P < 0.1$ を基準として総合的に判断する。Funnel plot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に均等に分布していない場合、出版バイアスを疑う。

出版バイアスへの対処法としてTrim and fill method(van Aert 2016; Watzlawick 2014; Schwarzer 2010; Sterne 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 2007)。

4.4.4. 観察研究のエビデンス評価

観察研究の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の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、稀ながら、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ことは、RCTと同様である。

観察研究のエビデンスの強さ評価する場合には、前述4.4.3で挙げた項目で強さを下げることとを考慮するほかに、次の項目で強さを「中」に上げることも検討する（Guyatt 2011h）。

①介入による大きな効果（large effect）

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、使わないとほとんど死亡することが多いが、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良い。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>

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(dose-dependent gradient)

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、有意差が出ていただろうということが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良い。

<判定>

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（plausible confounder）

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良い。たとえば、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か改善しなかったが、もし、背景が均一化され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>

- 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。

4.4.5 エビデンスの統合

4.4.5.1 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、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。4.4.1～4.4.4までをまとめる作業のことである。

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。

4.4.5.2 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、Forest plot、Funnel plotが得られる。これらのプロットは非一貫性、不精確、出版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かならず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ひとつの項目となる。たとえば、ある介入が統計学的有意に、そして顕著に良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良い。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良い。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が考慮しても良い。

・メタアナリシスが省略できる場合

メタアナリシスは必ずしも必須ではない。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良い。たとえば、対象とした論文がすべ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

たとえばコクランレビュー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、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。

・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良い。

コクラン共同計画はシステマティックレビューを行うための Review Manager (RevMan) (現バージョンは 5.3) とよ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、データを入力する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、Forest plot を Word などに貼り付け可能な形で出力できる (4.6 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 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、Forest plot, Funnel plot などの結果を得ることができる (4.7 参照)。

・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、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、標準偏差と呼ばれる場合もあるが標準誤差 Standard error と呼ばれる場合もあるが同じ意味で用いられている。

以下に、これら効果指標の算出について解説する。

○四分表

Two-by-two table

	Outcome(+)	Outcome(-)	Sum
Treated: Intervention(+)	r_i^T *	$n_i^T - r_i^T$	n_i^T
Control: Intervention(-)	r_i^C	$n_i^C - r_i^C$	n_i^C

*The number of events in the treatment group in the ith study.

効果指標算出の元になる四分表

2群の率(割合)からRR、ORなどが算出される。rはアウトカム陽性(イベント生起)例の人数、nは各群の総症例数、iは研究番号を表す。

①リスク比

リスク比の算出。リスク比 = $(r_i^T/n_i^T)/(r_i^C/n_i^C)$ である。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 Standard error)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 してもと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 standard error）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 standard error）は $\sqrt{[r_i^T(n_i^T-r_i^T)/(n_i^T)^3]+[r_i^C(n_i^C-r_i^C)/(n_i^C)^3]}$ で計算される。

④標準化平均値差

標準化平均値差 SMD 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。

Effect sizes in the d family for continuous variables

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

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

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

- Hedges' $g = (M_1 - M_2)/SD_{pooled}^*$

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

連続変数の場合の効果指標の例

⑤ハザード比

HR は RR 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 HR と信頼区間を算出することが可能である。

⑥その他

治療必要数（Number Needed to Treat, NNT）は RR とベースラインリスク、あるいは RD から算出できるが、理解が容易な効果指標であり、今後可能なかぎり提示すべきで

ある。RR の信頼区間が 1.0 を挟んでいる場合には、Number Needed to Treat for Benefit (NNTB) と 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 NNT の計算も可能である。

アウトカムが害 Harm の場合には、Number needed to harm, NNH が、介入がスクリーニング検査の場合には Number needed to screen, NNS が用いられる。

RR、OR 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。

・ 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しないで、それらの研究の効果指標の平均値を求めることになるといえる。ランダム効果モデルでは、実際の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から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まったく正しいといえる。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ことが多いので、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 (Fu R 2014)。

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。

$$M = \frac{\sum W_i M_i}{\sum W_i} \quad W_i = 1/V_i = N_i/s_i^2$$
$$V_M = \frac{1}{\sum W_i} \quad M \pm 1.96 * \sqrt{V_M} \quad Z = |\ln(M)| / \sqrt{V_M}$$

重みとして分散の逆数を用いる Variance-based method

Inverse variance-based method とも呼ばれる。M が統合値、W は重みであり、分散の逆数である。V は分散である。Z は有意差検定に用いられる (1.96 以上で有意)。

研究間の異質性の検出には Q 統計値や I² 統計値が用いられている。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。

$$Q = \sum W_i M_i^2 - \frac{(\sum W_i M_i)^2}{\sum W_i} \quad Q > k - 1 = df \rightarrow \text{heterogeneity (+)}$$

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

$$W_i^* = 1/V_i^* = 1/(V_i + T^2)$$

ランダム効果モデルと Q 統計値、I² 統計値

T²は 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 は研究番号、M は効果指標の値、V は分散、W は重みである。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 を挟んでいなければ、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、このような結果はすべての研究の効果指標の平均値のとりう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。

・感度分析 Sensitivity analysis

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たり、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。

データから実行が可能である限り、感度分析を行うことが推奨される。

・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通暁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同領域の専門家の協力の下に実行することが望まれる。

4.4.6 エビデンスの強さに関する判定および表記方法

・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 (Juni 1999; Linde 1999)。たとえば、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。

・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、ケースコントロール研究)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。

・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良い。

①RCTが、一つしかない場合

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コクランレビューで内容が評価されている論文を用いる場合

コクランレビューのエビデンス評価 (バイアスリスク、非一貫性、不精確、出版バイアス等) を利用することもできる (=コクランレビューの多くは、GRADEアプローチで評価されている)。

③海外のガイドラインを利用する場合

海外のガイドラインが、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な方法によって作成されている(たとえば、GRADE アプローチ、などで作成されている)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、出版バイアス等)を利用しても良い。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。

④一つの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する。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、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

また、過去に出版された診療ガイドラインを取り上げる場合、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られない。一方、その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、出版バイアス等)については、必要に応じて利用しても良い。

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。

○手順

(1)評価を下げる5項目を評価する

(1)-1 観察研究の場合は、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)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)エビデンス総体の強さを評価する

○テンプレート

- 【4-7 評価シート エビデンス総体】
- 【4-8 定性的システマティックレビュー】
- 【4-9 メタアナリシス】

○記入方法

- 【4-7 評価シート エビデンス総体 記入方法】
- 【4-8 定性的システマティックレビュー 記入方法】

○記入例

- 【4-7 評価シート エビデンス総体 記入例】
- 【4-8 定性的システマティックレビュー 記入例】
- 【4-9 メタアナリシス 記入例】

4.5 ステップ5：SRレポートの作成

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SRレポートにまとめ、推奨作成を行う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、または検索されたすべ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。

ガイドライン作成グループは、提出されたSRレポートをもと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・十分な情報を提示できることに留意する。

SR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 Summary of Findings (SoF) table (Guyatt 2011a; Guyatt 2013b; Guyatt 2013c; Carrasco-Labra 2015; Carrasco-Labra 2016) を作成してもよい。SoFは以下の7つの要素を含む。

1. すべての重要・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

○テンプレート

【4-10 SRレポートのまとめ】

【4-11 Future research question】

○記入方法

【4-10 SRレポートのまとめ 記入方法】

【4-11 Future research question 記入方法】

○記入例

【4-10 SRレポートのまとめ 記入例】

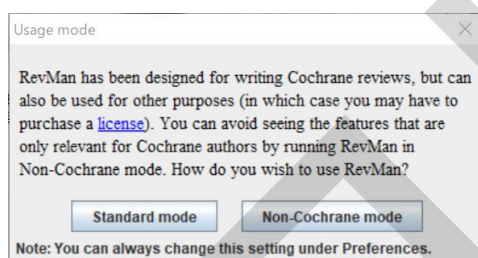
【4-11 Future research question 記入例】

4.6 (参考) Review Manager (RevMan) 5.3によるメタアナリシス

コクラン共同計画によりフリーで提供されている Review Manager (RevMan) は、システマティックレビューを効率よく、共通の基盤で行うために開発されており、文献の管理やレビューの管理、論文から抽出したデータの管理などの機能が強力である。その中にメタアナリシスのツールも付属しており、広く使用され、RevMan を用いた論文も多数発表されている。

2017年8月時点の RevMan の最新はバージョン 5.3.5 である。Cochrane のウェブサイト (<http://ims.cochrane.org/revman/download>) よりダウンロードできる。Windows、Mac、Linux それぞれの OS に対してプログラムが用意されている。ここでは、Windows 版を用いて RevMan5 を用いたメタアナリシスの手順を主に解説する。最後にバイアスリスクの評価結果の入力と解説する。

RevMan を起動する。最初は使用モードを設定する画面が出る。

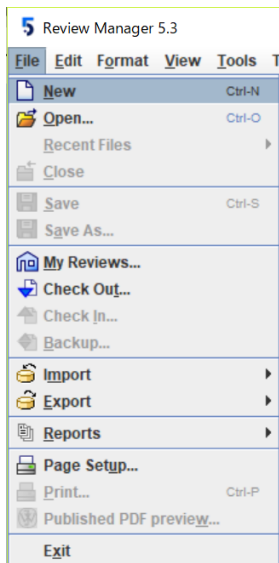


ここでは Standard mode を選択する。

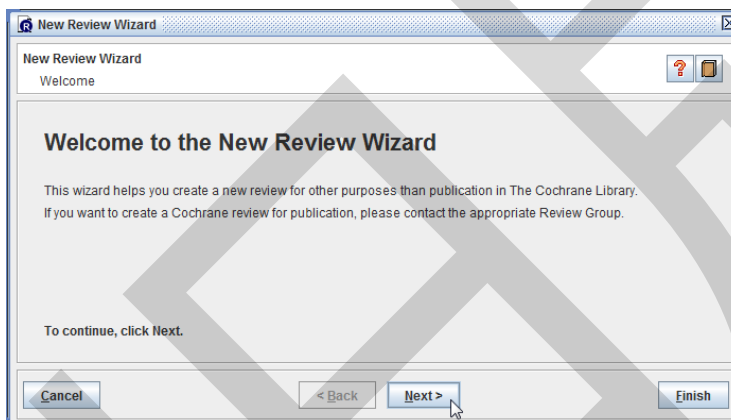
RevMan が開かれるので、右下の Close をクリックしていったんウィンドウを閉じる。



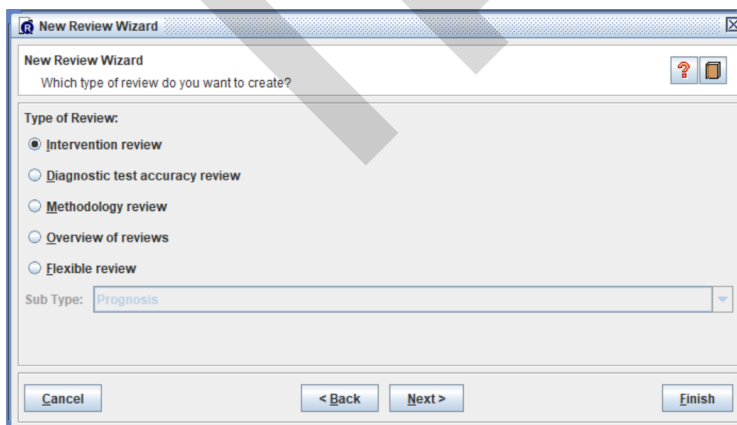
FileメニューからNewを選択する。



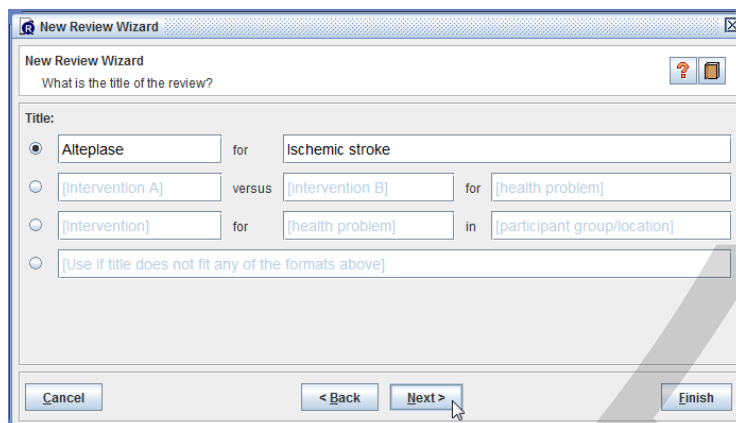
New Review Wizardの画面が開かれるのでNextをクリックする。



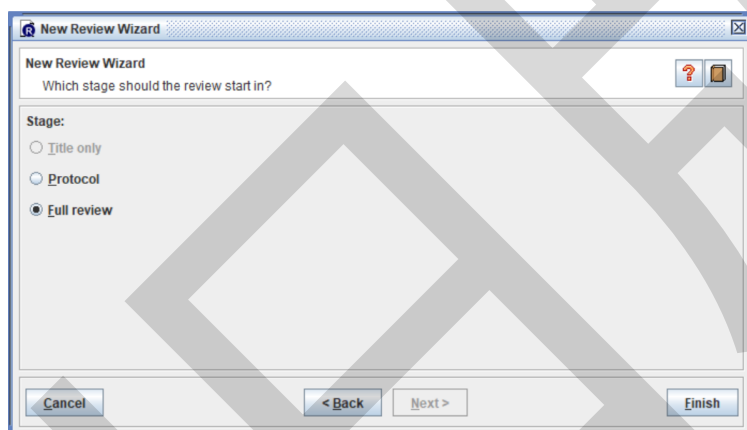
Type of Reviewを選択する。ここでは、Intervention Reviewを選択し、Nextをクリックする。



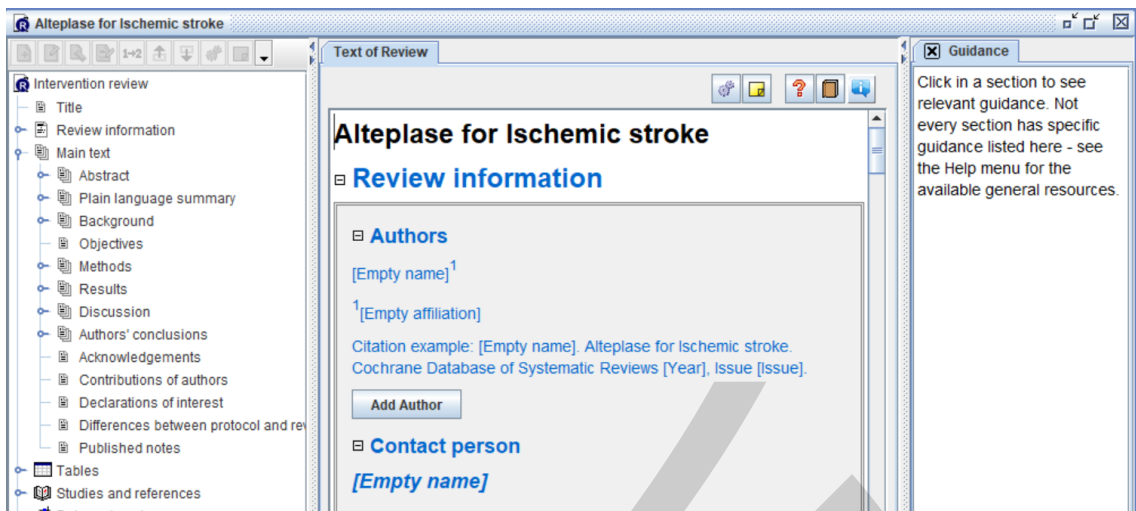
Review に付けるタイトルを入力する。いずれかのラジオボタンを選択し書き込み、Next をクリックする。[Alteplase] for [ischemic stroke]と入力（カギカッコ内を入力）。



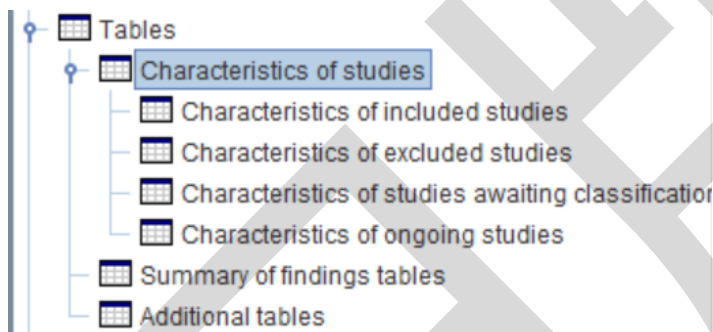
Review のステージを選択する。ここでは、Full review を選択し、Finish をクリックする。




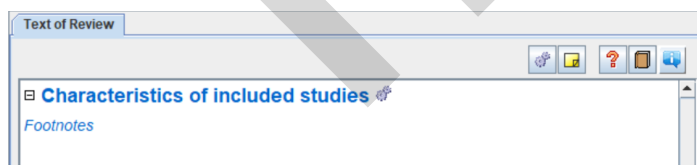
右側の Text of Review の画面にタイトルそのほかの情報が表示される。左側のサイドバーには、Review を構成するさまざまな要素が表示される。サイドバーではプレットの向きによってメニューが開かれたりたたまれたりする。この時点で File メニューからファイルとして保存することもできる。拡張子は .rm5 でファイル名はデフォルトではタイトルと同じであるが、変更することもできる。



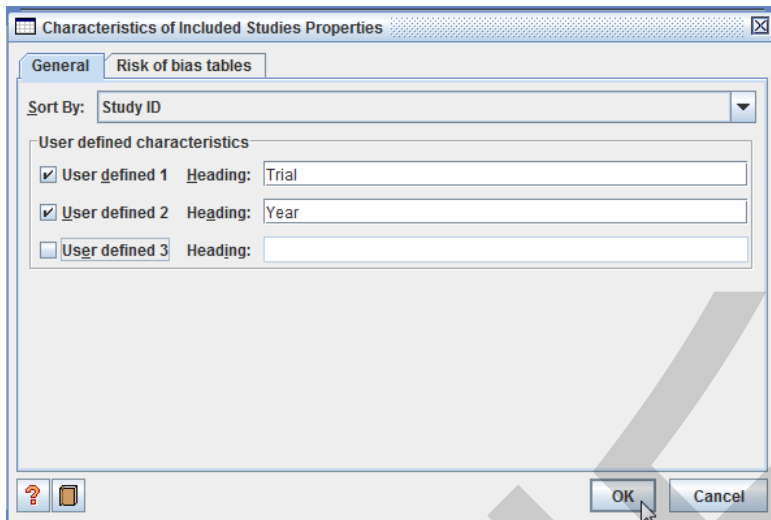
ここで、Tables のプレットをクリックし、次いで Characteristics of included studies のプレットをクリックする。



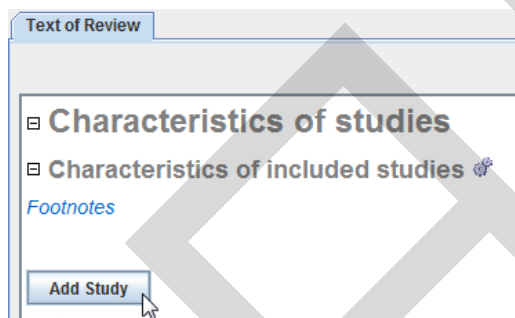
Characteristics of studies の行をクリックすると、右の Text of Review の画面は自動的にスクロールして一番上の行が Characteristics of studies となる。左側の Characteristics of included studies をダブルクリックするか、右の画面の Characteristics of included studies の右側のプロパティを表す  ボタンをクリックする。



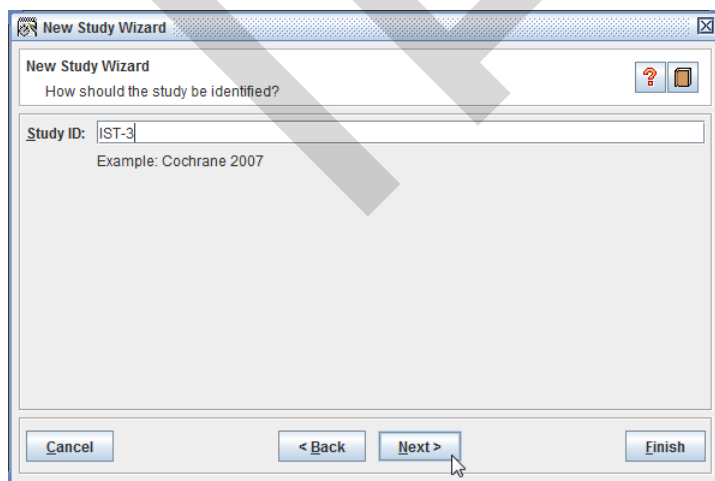
臨床試験の名称と発表年を入力するために Heading を Trial と Year の2つ設定し、OK をクリックする。



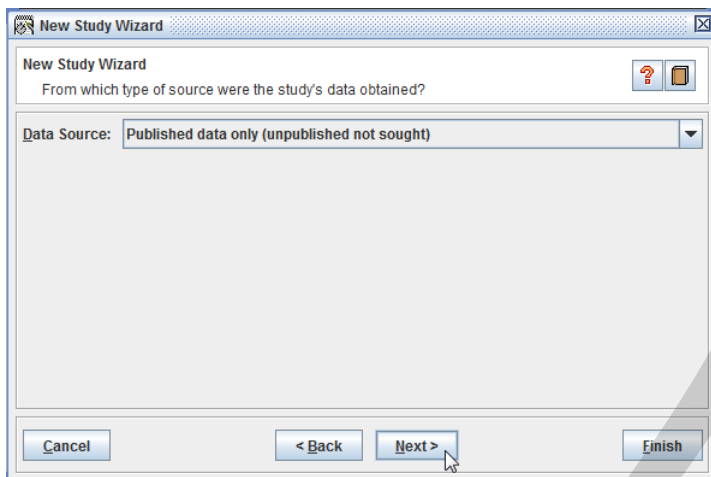
右の Text of Review の画面で、Characteristics of included studies の左の+のブレットをクリックし、Add Study ボタンをクリックする。



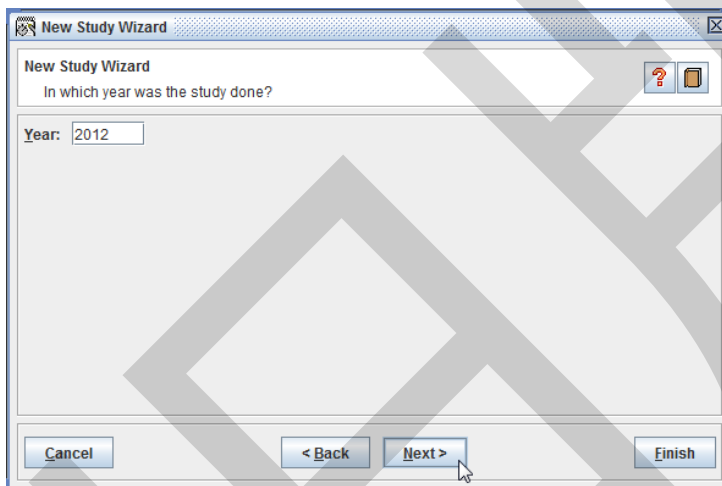
Trial の名称を入力して Next をクリックする。Trial の名称ではなく、著者名や番号などでも構わない。



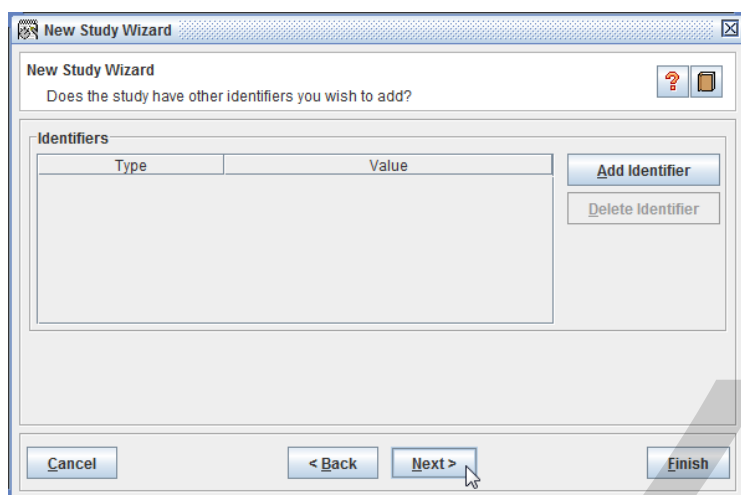
出版されたデータを用いるので、そのまま Next をクリックする。



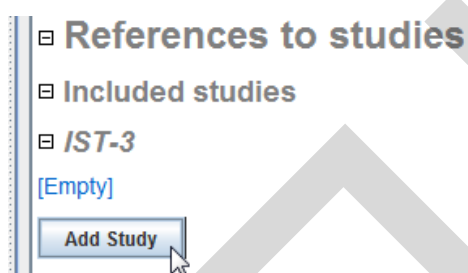
年度を入力して Next をクリックする。



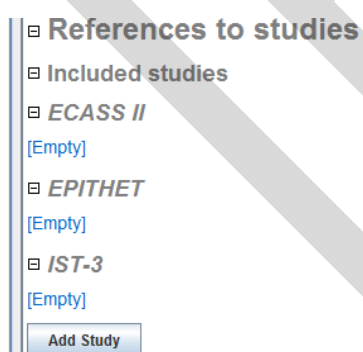
次に、研究の特徴などを入力する画面となる。ここでは、不要なので、そのまま Finish をクリックする。



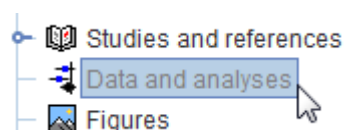
以上で1つの研究の属性を入力した。入力した研究の名称が右の画面に表示されるようになる。



さらに Add Study をクリックして必要な研究をすべて入力する。ここでは、3つの研究を入力した。以下の画面の様に入力した3つの研究のタイトルが表示される。データはまだ入力していないので、各名称の下に [Empty] と表示されている。

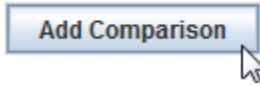


次に、左側のサイドバーの Data and analyses をクリックする。

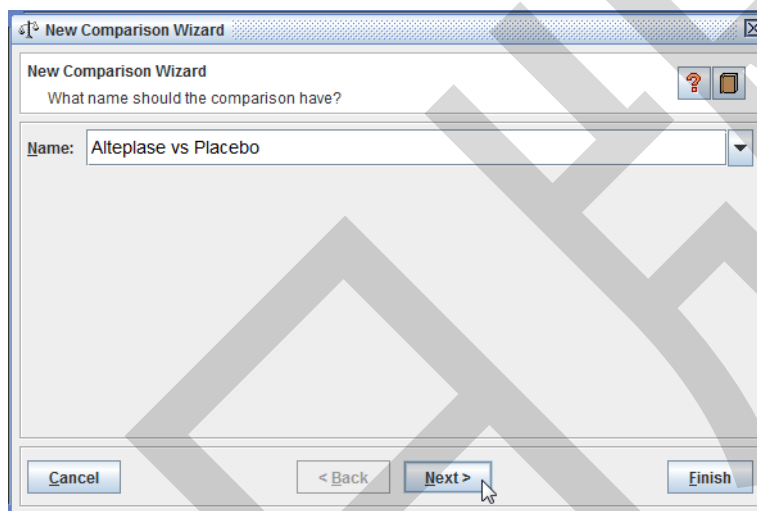


右側の画面に Add comparison のボタンが表示されているので、それをクリックする。

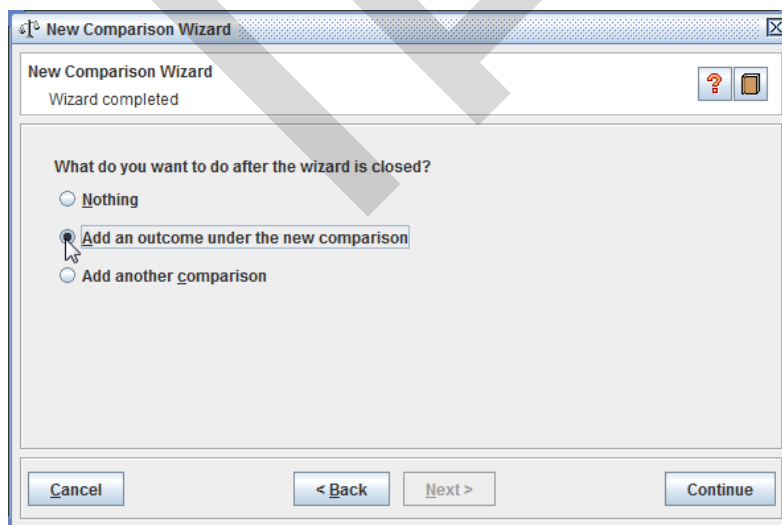
☐ Data and analyses



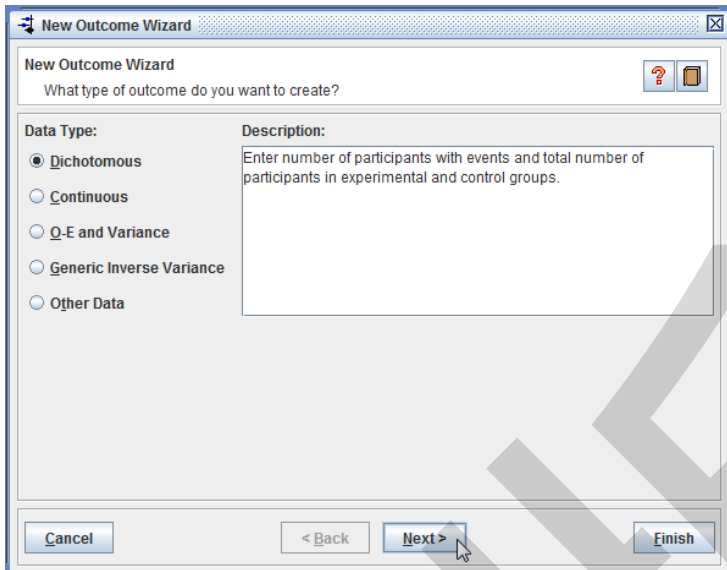
New Comparison Wizard の画面が表示される。まず、比較の名称を入力し、Next をクリックする。ここでは Alteplase vs Placebo と入力。(ここで、いったん Finish で終了してから、あとで右側の Text of Review の画面で表示される Add Outcome ボタンをクリックして以下のデータの入力を行うこともできる。)



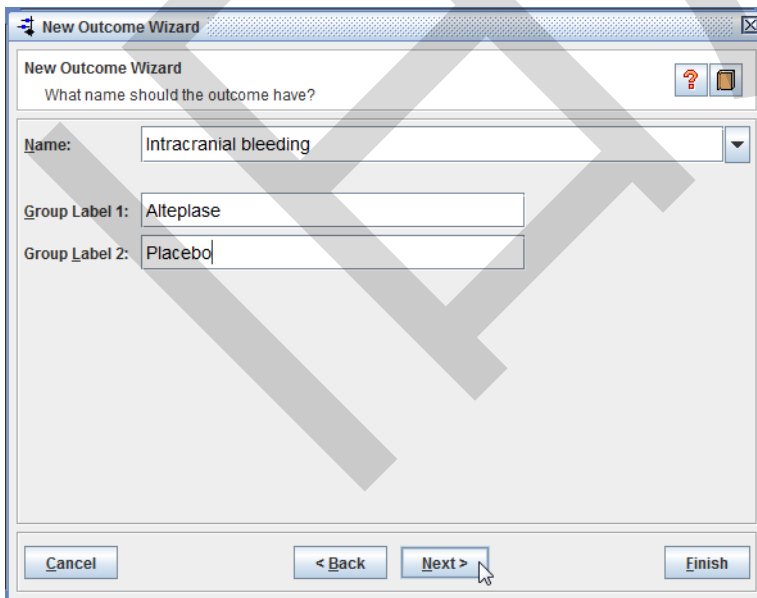
続けて比較の対象となるアウトカムを入力するため、Add an outcome under comparison を選択し Continue をクリックする。



New Outcome Wizard の画面が表示されるので、比較されるデータ型を選択し、Next をクリックする。ここでは、Dichotomous を選択。



アウトカムの名称と比較される 2 群のラベルを入力する。Group Label 1: と Group Label 2: はデフォルトでは Experimental と Control となっているので、それぞれを書き換える。ここでは、名称を Intracranial bleeding、群のラベルは Alteplase、Placebo とする。入力したら、Next をクリックする。



メタアナリシスの方法、効果指標、モデルを設定する画面が出るので、それぞれラジオボタンを選択する。ここでは、Inverse Variance、Risk Difference、Random Effects を選択する。Next をクリックする。

New Outcome Wizard
Which analysis method do you want to use?

Statistical Method

- Peto
- Mantel-Haenszel
- Inverse Variance
- Exp[(O-E) / Var]

Analysis Model

- Fixed Effect
- Random Effects

Effect Measure

- Peto Odds Ratio
- Odds Ratio
- Risk Ratio
- Risk Difference
- Mean Difference
- Std. Mean Difference
- Name of Effect Measure:

Cancel < Back Next > Finish

メタアナリシスの結果を出力する際の項目を選択する画面となる。Totals and subtotals、95%、95%を選択して、Next をクリックする。

New Outcome Wizard
Which analysis details do you want to use?

Totals

- Totals and subtotals
- Subtotals only
- No totals

Study Confidence Interval

- 90%
- 95%
- 99%

Total Confidence Interval

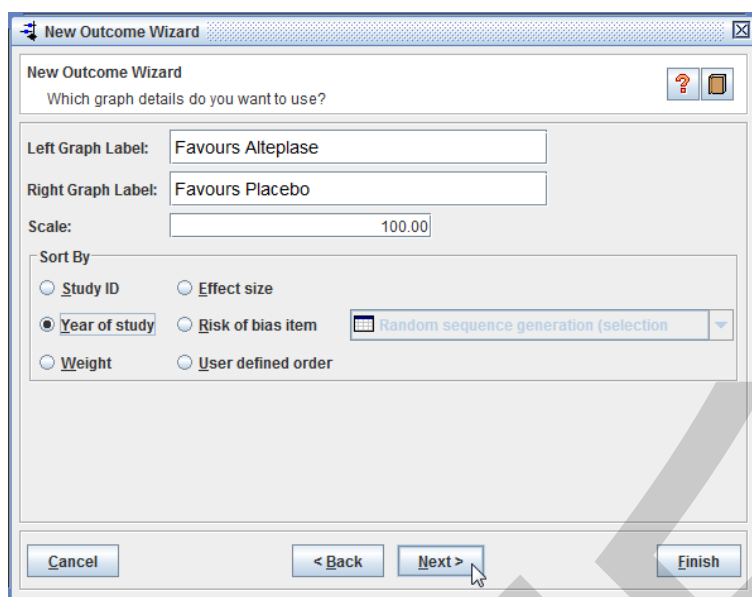
- 90%
- 95%
- 99%

Advanced Options

- Test for subgroup differences
- Swap event and non-event

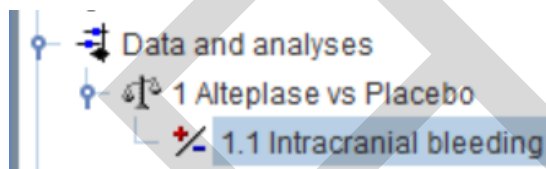
Cancel < Back Next > Finish

フォレストプロットでの表示に関するデータ設定の画面が出るので、必要に応じて入力・選択する。Left Graph Label、Right Graph Label はデフォルトで Favors [experimental]、Favors [control]となっているが、ここでは、かぎかっこの部分をそれぞれ Alteplase、Placebo と書き換え、Sort By は Year of study を選択する。Next をクリックする。



次の画面でいったん Finish する。

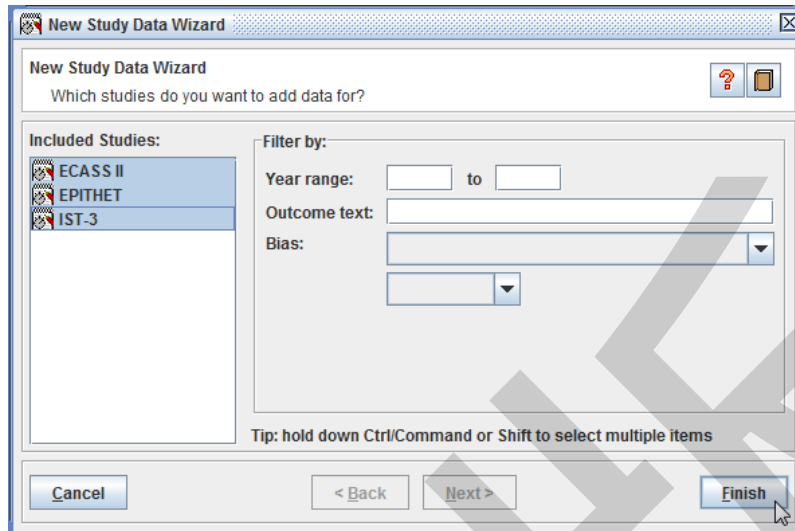
次に、左のサイドバーの Data and analyses の下に今設定した Outcome が表示されているので、その部分 1.1 Intracranial bleeding をダブルクリックする。



右の Text Review の画面にそのアウトカムに関するデータが表示される。まだデータを入力していない状態なので、症例数は0が表示されている。

Study or Subgroup	Alteplase		Placebo		Weight	Risk IV, Random, ...
	Events	Total	Events	Total		
Total (95% CI)		0		0		
Total events	0		0			
Heterogeneity: Not ...						
Test for overall effe...						

ここで、上に並んでいるアイコンの一番左の+のついたアイコンをクリックすると、データ入力する研究名称を選択する画面が出る。左側に研究名称の一覧が表示されるので、そこから選択して Finish をクリックする。今回は、3つ全部を選択する (Control キーを押しながら選択すると複数選択が可)。



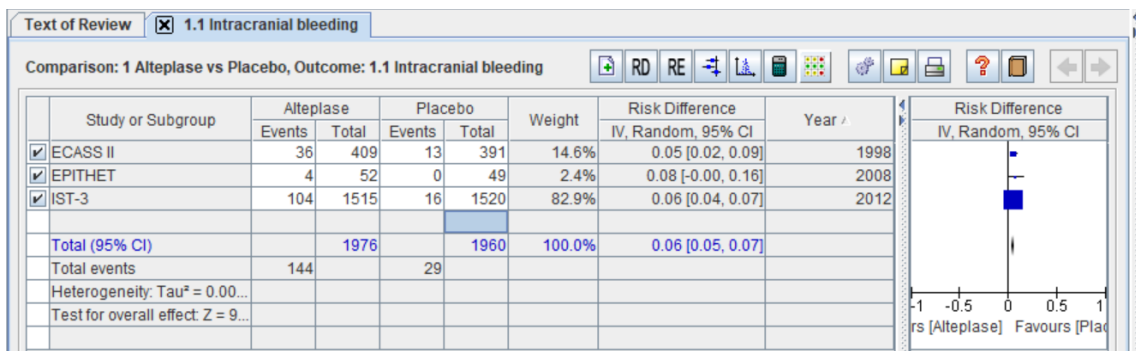
右側の Text of Review の画面に表が表示されるので、各研究の介入群と対照群の総症例数とイベント生起例数を入力する。先ほど Year でソートする設定にしたので、年度順に研究名称が並んで表示されている。

Text of Review [X] 1.1 Intracranial bleeding

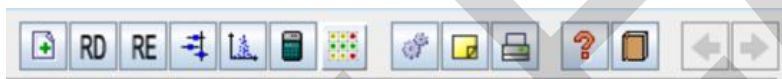
Comparison: 1 Alteplase vs placebo, Outcome: 1.1 Intracranial bleeding

Study or Subgroup	Alteplase		Placebo		Weight	Risk Diff IV, Rando
	Events	Total	Events	Total		
<input checked="" type="checkbox"/> ECASS II	0	0	0	0		No
<input checked="" type="checkbox"/> EPITHET	0	0	0	0		No
<input checked="" type="checkbox"/> IST-3	0	0	0	0		No
Total (95% CI)		0		0		No
Total events	0		0			
Heterogeneity: Not applicable						
Test for overall effect: Not applicable						

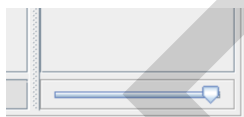
背景が白のセルにデータを順次入力し、すべての値の入力を終了すると以下の結果が得られる。



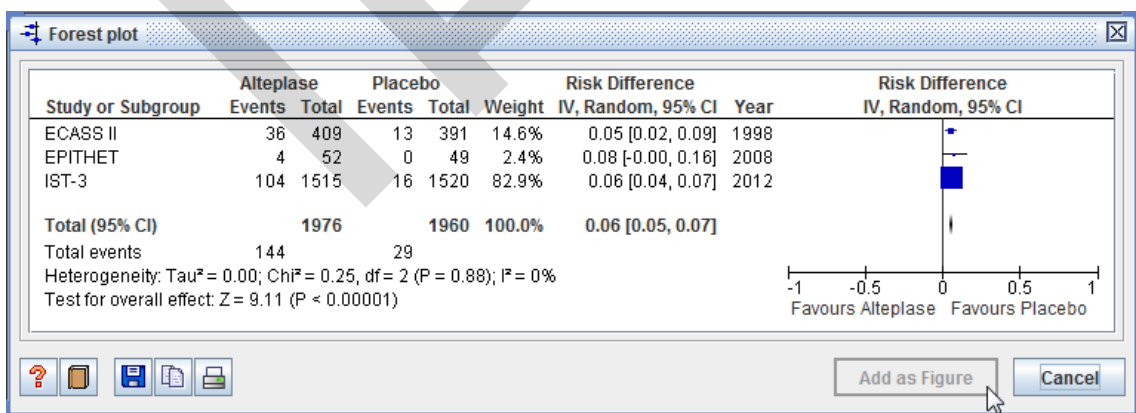
この画面で、右上のボタンによって、さまざまな変更ができるようになっている。一番左側のボタンはさらに研究を追加する場合に用いる。2つ目のボタンは現在 RD (Risk Difference) の結果を表示していることを示しており、クリックすると OR オッズ比、RR リスク比へと変更される。3つ目のボタンは RE、すなわちランダム効果モデル Random effects model を用いていることを示しており、このボタンをクリックすると FE、Fixed effects model に変更できる。これらはすぐ結果に反映され何度も変更可能である。



右下のスケールのボタンを移動させると、グラフの大きさが変わる。

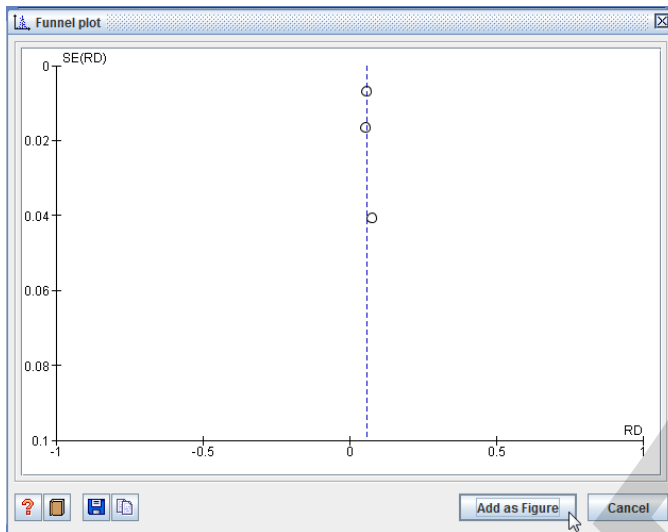


4つ目のボタンをクリックすると Forest plot を別画面に表示する。Forest plot を表示した状態で Add as Figure をクリックするとその画面の情報が Figures の情報に追加される。

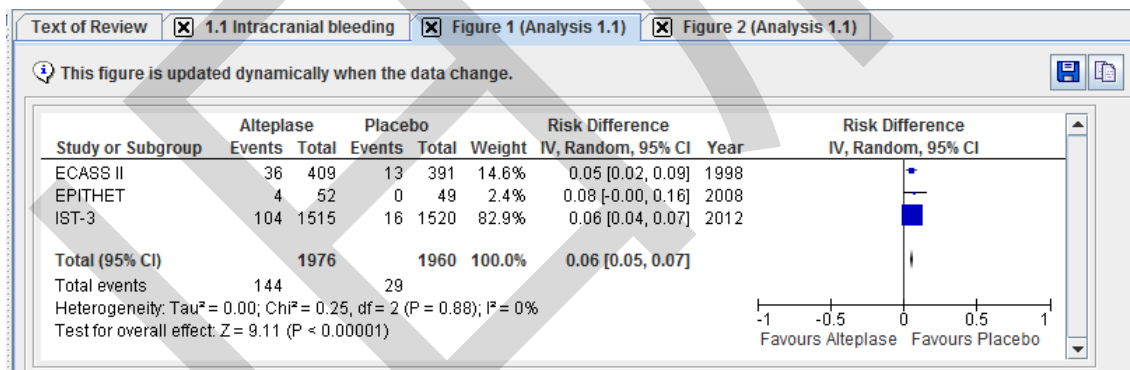


5つ目のボタンをクリックすると Funnel plot が表示される。Add as figure をクリ

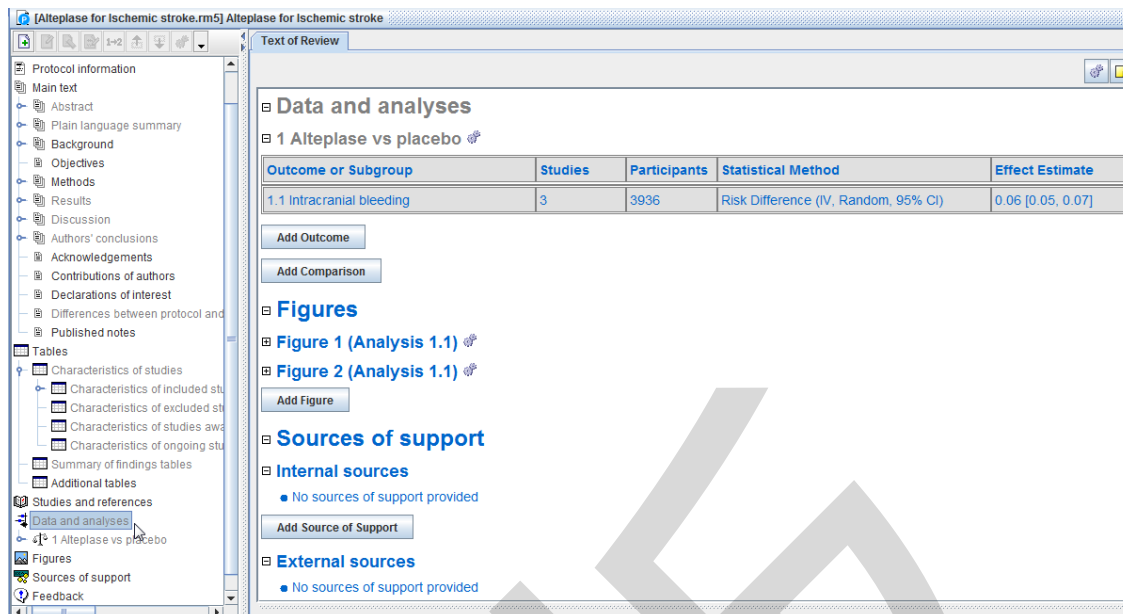
ックすると、この図を Figures の情報に追加できる。



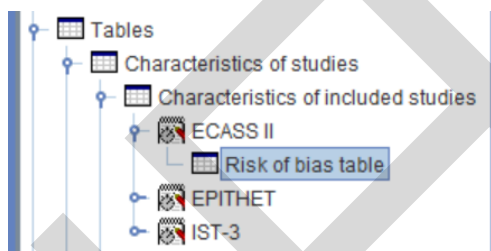
Add as figure で情報を Figures に追加すると、右画面の上のタブの部分に Figure 1、Figure 2 と表示されるようになる。クローズボックスをクリックすると閉じることができる。また、画面を保存する場合は、右上にある保存ボタンを、そしてコピーして Word などに貼り付けたい場合には、その隣のコピーボタンをクリックするとクリップボードにコピーされるので、それを貼り付けることができる。



以上の作業を済ませて、左のサイドバーから Data and analyses をクリックすると作業の結果が右の画面に表示される。右画面のたとえば Figure 1 (Analysis 1.1) の + ボタンをクリックすると Forest plot がその下に表示される。Figure 2 は Funnel plot である。また、これらの文字列の部分ダブルクリックすると Text of Review のタブの隣に Figure 1 (Analysis 1.1) というタブが表示されるとともに、上記の画面と同じように右側の画面に Figure が表示される。



各研究のバイアスリスクの評価を行いたい場合には、左のサイドバーの Characteristics of included studies で Risk of bias table をダブルクリックする。

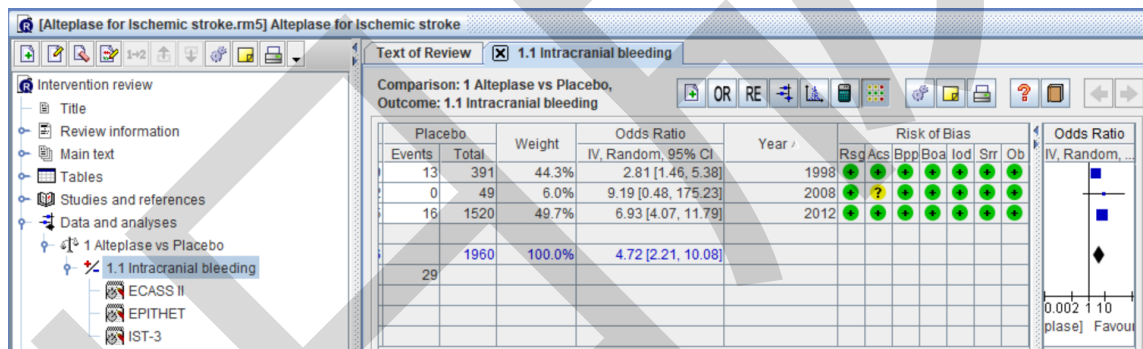


その研究のバイアスリスクの判定を入力する画面が出るので、それぞれ Authors' judgement のポップアップメニューから選択する。さらにその根拠を Support for judgements に入力する。その結果は Data and analyses の画面に反映される。

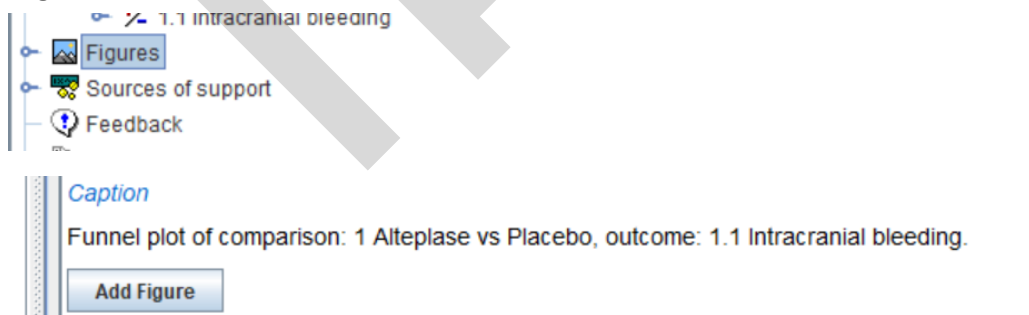
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	
Allocation concealment (selection bias)	Unclear risk	
Blinding of participants and personnel (performance bias)	Unclear risk	
Blinding of outcome assessment (detection bias)	Unclear risk	
Incomplete outcome data (attrition bias)	Unclear risk	
Selective reporting (reporting bias)	Unclear risk	
Other bias	Unclear risk	

作業を終了するとき、Fileメニューからファイルとして保存する。

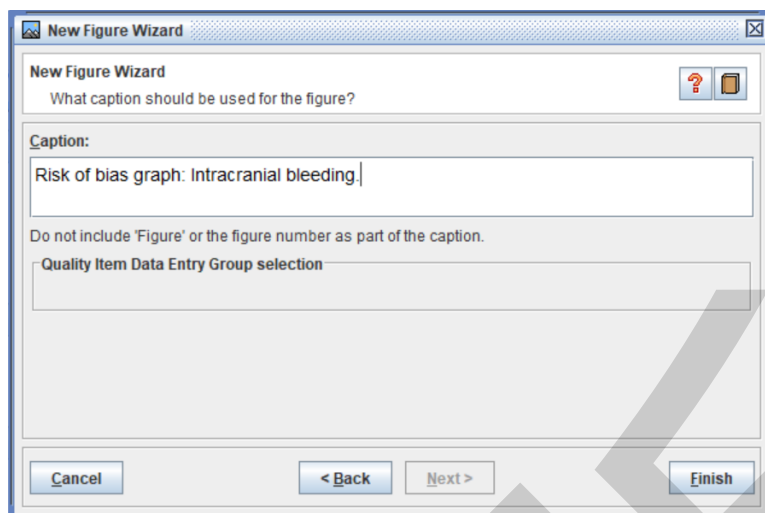
Risk of bias tableの設定を、すべての研究について終了後、保存したファイルを開き、Data and analysesの1.1 Intracranial bleedingをダブルクリックすると、右上のRisk of bias figureのボタンが押されて状態であれば、右の画面にRisk of bias figureが追加されて表示される。



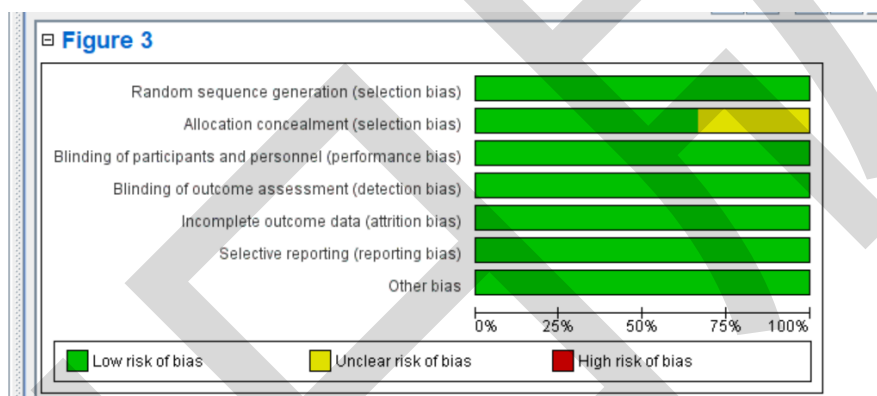
Risk of bias graphは左のサイドバーでFiguresをクリックして、右の画面でAdd Figureをクリックする。



Risk of bias graph を選択して Next をクリックし、Caption を入力して、Finish をクリックする。

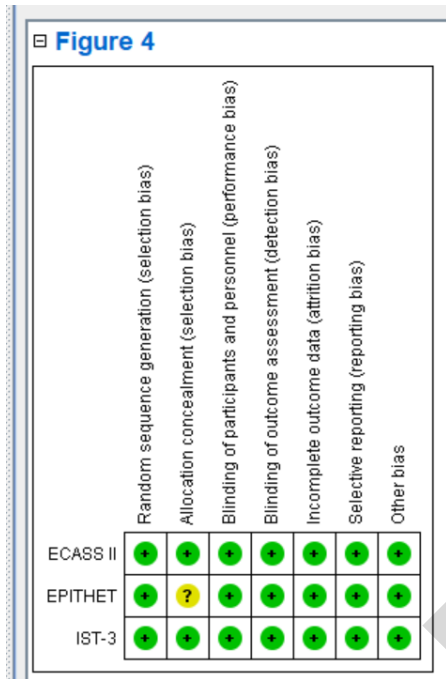


右の画面に表示される。



同じ手順で Risk of bias summary を選択すると、Risk of bias figure が右の画面に表示される。ここで言う Summary の意味は対象となった複数の研究の評価結果を、研究全体をまとめて表示するという意味で、研究ごとのバイアスリスクのまとめや、エビデンス総体のバイアスリスクを一つの値で表したものではない。

Fileメニューからファイルとして保存すると、追加・変更箇所がすべて保存される。



以上。

RevMan 使用時の注意点

RevMan (Review Manager)はコクランシステマティックレビュー作成を目的に作られているため、診療ガイドライン作成のための利用時には機能の一部を利用することになる。上記のようにRisk of bias summary figure およびRisk of bias graph figure を作成することができるが、研究ごとのsummary of risk of biasは、図には含まれていないこと、また、エビデンス総体の確実性を評価する際にはサンプルサイズによる重みづけが必要であるが、図にはそれが表示されていないことに注意し、適切な判定をすることが望まれる。また、Cochrane risk of bias tool の個別のバイアスリスクの評価の基準については92ページを参照。

また、表に示すように、コクランのバイアスリスクのHigh, Unclear, Lowの判定からGRADEのVery serious limitation, Serious limitation, No serious limitationの判定への変換作業が必要になる。

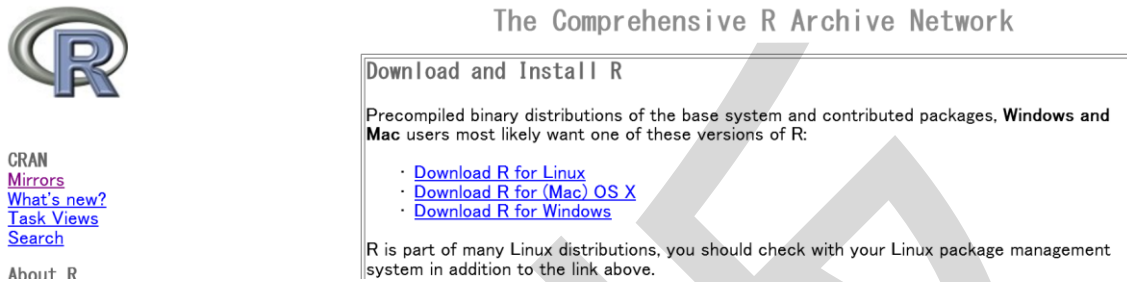
Risk of bias	Across studies	Interpretation	Considerations	GRADE assessment of study limitations
Low risk of bias.	Most information is from studies at low risk of bias.	Plausible bias unlikely to seriously alter the results.	No apparent limitations.	No serious limitations, do not downgrade.
Unclear risk of bias.	Most information is from studies at low or unclear risk of bias.	Plausible bias that raises some doubt about the results.	Potential limitations are unlikely to lower confidence in the estimate of effect.	No serious limitations, do not downgrade.
			Potential limitations are likely to lower confidence in the estimate of effect.	Serious limitations, downgrade one level.
High risk of bias.	The proportion of information from studies at high risk of bias is sufficient to affect the interpretation of results.	Plausible bias that seriously weakens confidence in the results.	Crucial limitation for one criterion, or some limitations for multiple criteria, sufficient to lower confidence in the estimate of effect.	Serious limitations, downgrade one level.
			Crucial limitation for one or more criteria sufficient to substantially lower confidence in the estimate of effect.	Very serious limitations, downgrade two levels.

Cochrane handbook for systematic reviews for interventions (Version 5.1) Table 12.2.d (Higgins 2011a)

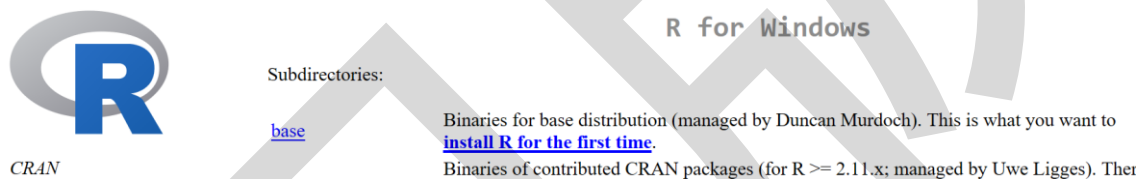
4.7 (参考) R と metafor パッケージを用いたメタアナリシス

R のインストール

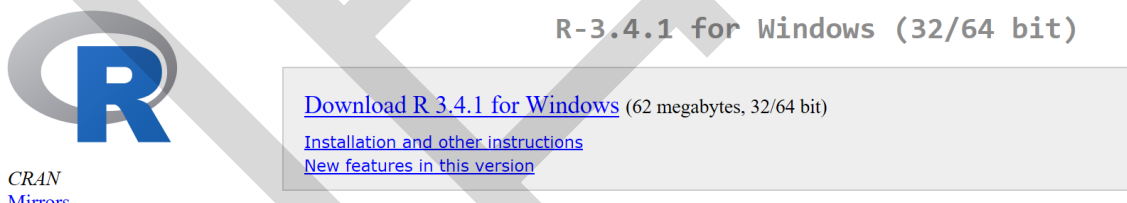
R のプログラムは The Comprehensive R Archive Network (CRAN) のウェブサイト <http://cran.r-project.org/> からダウンロード可する。OS のプラットフォームとして Windows、Linux、Mac に対応したバージョンが用意されている。



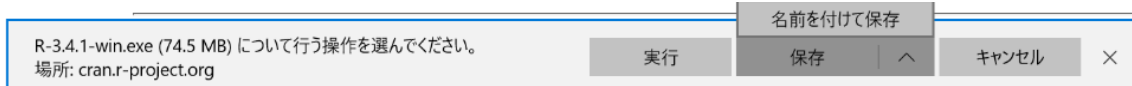
Windows の場合であれば、Download for Windows をクリックし、次の画面で install R for the first time をクリックする。



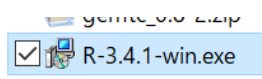
次の画面で Download R “バージョン No.” for Windows をクリックして R-3.0.1-win.exe ファイルをダウンロードする。



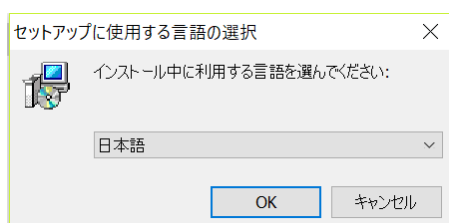
ブラウザーの下の方に保存ボタンが現れるので、名前を付けて保存を選び、ダウンロードのフォルダなどにファイルを保存する。



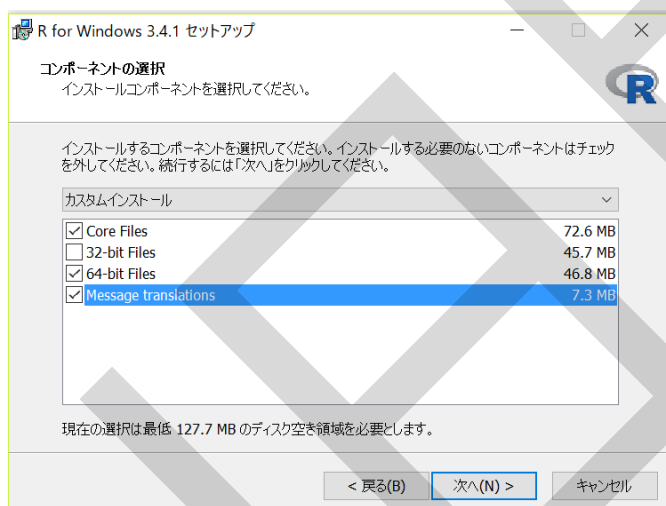
ダウンロードしたファイルをダブルクリックして、インストールを開始する。



コンピュータへの変更を確認する画面が出るので、OK をクリックし、言語を選択する画面となるので、日本語を選択し、インストール中の表示を日本語にする。

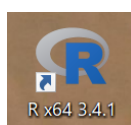


あとはセットアップウィザードが開始されるので指示に従って、インストールする。途中でコンポーネントの選択画面が出るので、Message translation にチェックを入れておくと、インストールされた R のメニューが日本語表示となる。



32 ビット版と 64 ビット版が選択できるので、各自の PC に合わせていずれかをインストールすればよいが、両方をインストールすることもできる。スタートメニューにアイコンを置くか、デスクトップにアイコンを置くかなどの設定はインストールの過程で選択できるのは他のソフトウェアの場合と同様である。

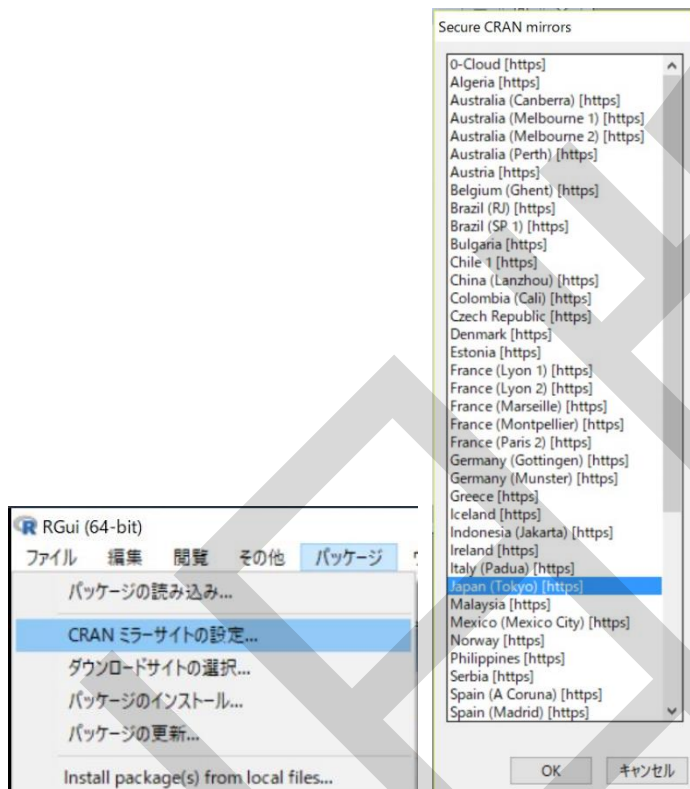
デスクトップにおかれた R のアイコン。



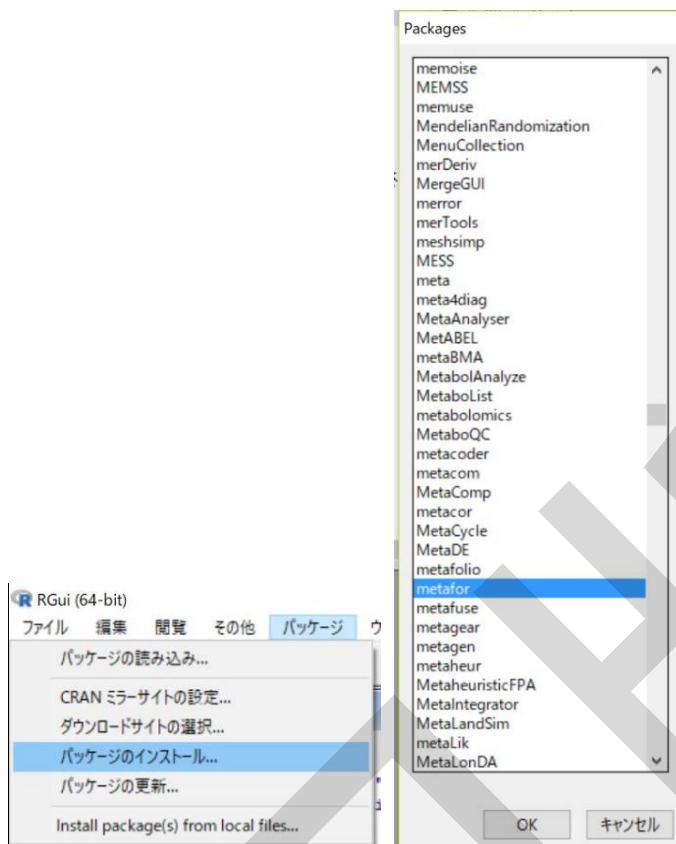
パッケージのインストール

R 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。Viechtbauer W 氏が開発したメタアナリシスのためのパッケージ metafor (Viechtbauer 2017) を用いてメタアナリシスを行い、Gordon M 氏の開発した Forest plot のためのパッケージ forestplot (Gordon 2017) 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探し、選択して、OK をクリックする。



他の 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 を用いている場合は、Program Files フォルダの中の R フォルダの中の 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。インストール作業は最初に一度行うだけである。

メタアナリシスのためのデータの準備

必要なデータは Excel で用意できるので、評価シートの一部を用いて用意することができる。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。

アウトカムが二分変数で効果指標がリスク比の場合の例を以下に示す。同じデータで RR を RD に変え、Risk Ratio を Risk Difference に変えれば、リスク差を効果指標としてメタアナリシスを行う。OR, Odds Ratio に変えれば、オッズ比を効果指標とする。

アウトカム 治癒					
研究コード	対照群		介入群		ラベル
	総例数	アウトカム +例数	総例数	アウトカム +例数	
author	nc	cpos	nt	tpos	label
Salminen P 2015	273	272	257	186	Author Year
Vons C 2011	119	104	120	72	Surgery
Hansson J 2009	167	142	202	83	Antibiotics
Styrud J 2006	124	124	128	97	Healing
Eriksson S 1995	20	20	20	13	Risk Ratio
					RR

この行のラベル名は変更不可
 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	対照群		介入群		ラベル
	総例数	効果指標 値	総例数	標準誤差	
author	nt	nc	yi	sei	label
Doffoel M 2008	62	61	-0.0726	0.1972	Author Year
Lo CM 2002	40	40	-0.6931	0.245	BSC
Llovet JM 2002	40	35	-0.755	0.4661	TACE
Pelletier G 1998	37	36	-0.0834	0.2659	Death
Groupe d'Etude 1995	50	46	-0.3425	0.2303	Hazard Ratio
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
Styrud 2006	124	2.6	1.2	128	3	1.4	Appendectomy
Turhan 2009	183	2.4	0.14	107	3.14	0.1	Antibiotics
Vons 2011	119	3.04	1.5	120	3.96	4.87	Days in hospital
							Mean Difference
							MD

略称(RR, OR, RD, HR,MD)

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画面を開き、次のスクリプトを書き込む、あるいはこの画面からコピーして貼り付ける。後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。以後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。

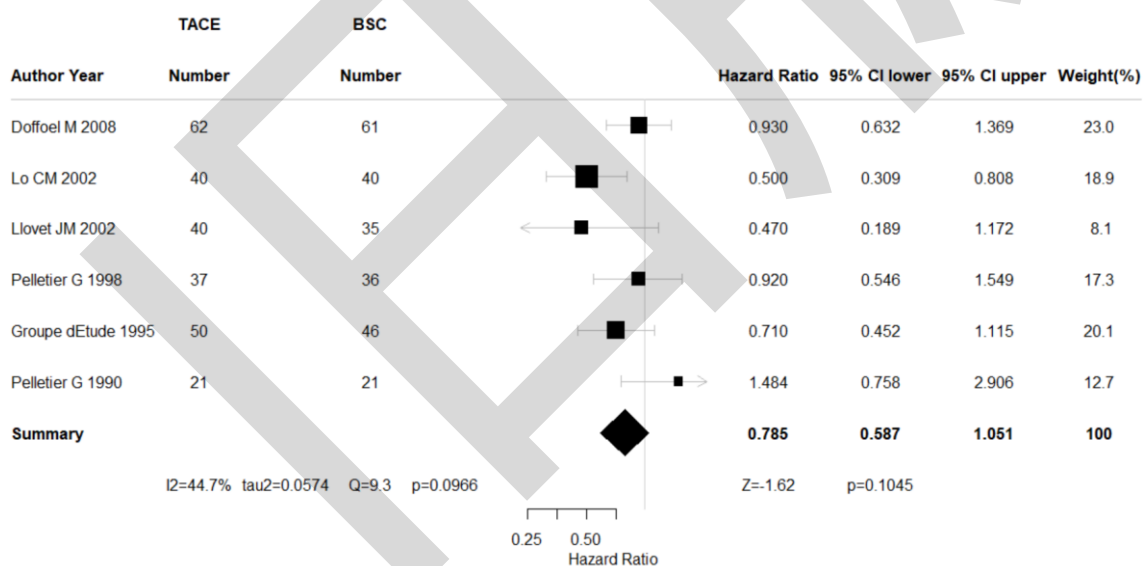
metaclip.R

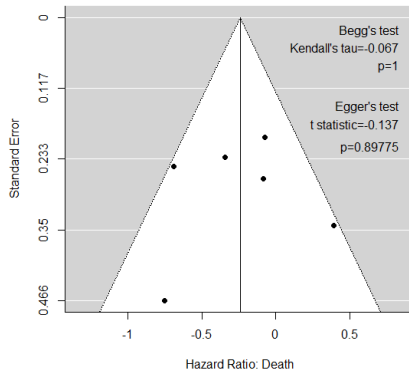
```
#1. クリップボード経由でデータ読み込み：ヘッダーあり
exdat=read.delim("clipboard", sep="¥t", header=TRUE)      # Windows
exdat=read.delim(pipe("pbpaste"), sep="¥t", header=TRUE)  # Mac
#2. スクリプトを読み込んでメタアナリシスを実行。
source("http://zanet.biz/med/useRj/scripts/forest_metafor_b.R")
```

Excel を起動して、データファイルを開き、上記の太枠の範囲を選択し、コピー操作を行う。

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)。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。

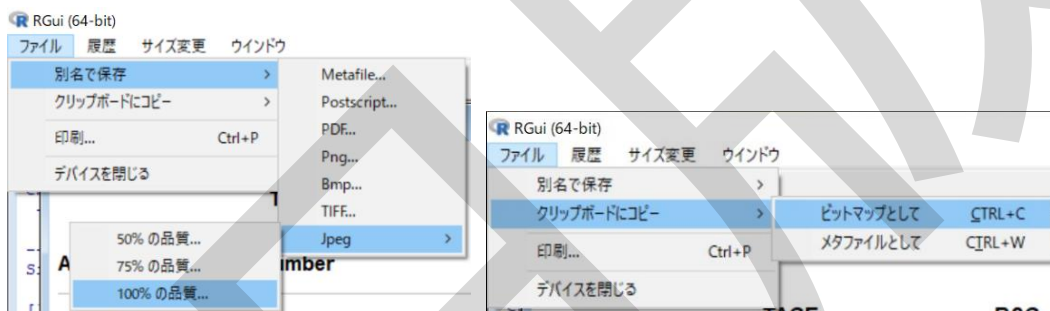
続いて、#2のスクリプト source(……)の行にカーソルを置いて、同様に実行させるとForest plotとFunnel plotが描画される。





Forest plot はウィンドウの大きさを変更して、全部が見えるようにする。

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して Excel、Word や PowerPoint に貼り付けたりすることができる。



注)

リスク比、リスク差、オッズ比、ハザード比、平均値差を効果指標としたメタアナリシスを metafor で実行し、Forest plot は forestplot、Funnel plot は metafor で作成するスクリプト metafor_forest_b.R はサーバにおいてあり、それをインターネットを介して、R で読み込んで実行させている。メタアナリシスの手法はランダム効果モデルを用い、Restricted maximum likelihood method で処理を行っている。RevMan で用いられている DerSimonian-Laird method とはわずかに統合値が異なる場合がある。

スクリプトファイルをローカルに置いて使用したい場合は、以下の URL から ZIP ファイルをダウンロードして解凍し、metaclip.R と forest_metafor_b.R の二つのファイルをどこかのフォルダに置く。R を起動後、ファイルメニューからディレクトリの変更 を選択し、そのフォルダを作業ディレクトリとして選択する。ファイルメニューからスクリプトを開く を選択し、metaclip.R を開いて、#1 のスクリプトで Excel からデータを読み込み、#3 のスクリプトを実行する。

URL: http://zanet.biz/med/useRj/scripts/forest_metafor_b.zip

また、スクリプトファイルを使用時にウェブサイトから読み込んで解析を実行することも可能である。この場合は、スクリプトを各自管理する必要がなく、解析用のデータを Excel で用意するだけで済む。

URL: <http://zanet.biz/med/useRs/>



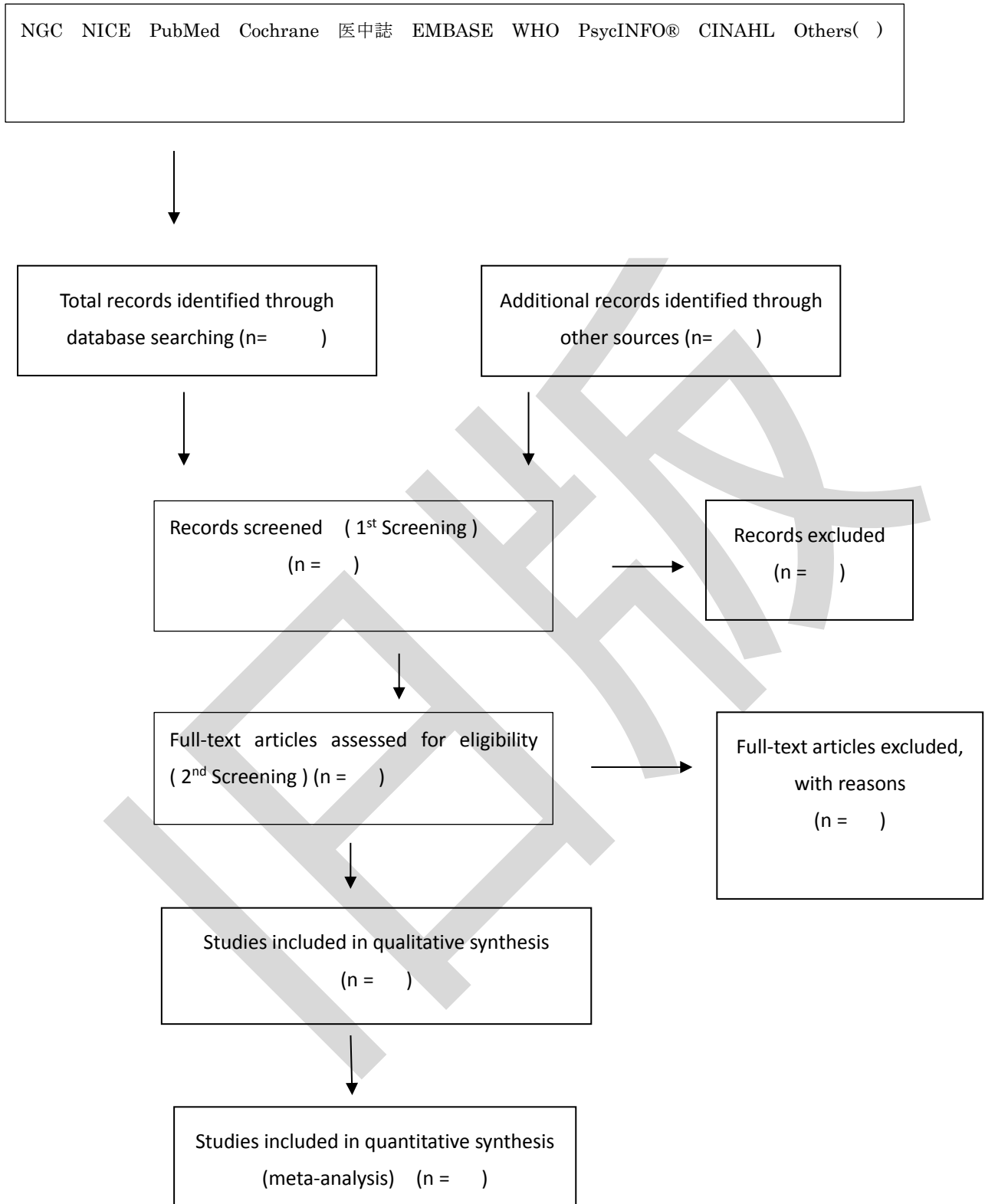
4章 テンプレート

【4-1 データベース検索結果】

タイトル:
CQ:
データベース:
日付:
検索者:

＃	検索式	文献数

【4-2 文献検索フローチャート】 PRISMA 声明を改変



【4-3 二次スクリーニング後の一覧表】

文献	研究デザイン	P	I	C	O	除外	コメント

【4-4 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		

【4-5 評価シート 介入研究】

【4-5 評価シート 介入研究】

診療ガイドライン
対象
介入
対照

* 各項目の評価は“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
各アウトカムごとに別紙にまとめ

アウトカム	個別研究	バイアスリスク*	バイアスリスク*						非直接性*			リスク人数(アウトカム率)						効果指標(種別)	効果指標(値)	信頼区間								
			ランダム化	選択バイアス	実行バイアス	検出バイアス	症状減少バイアス	不全アウトカムデータ	選択的アウトカム報告	早期試験中止	他のバイアス	対照	介入	アウトカム	まとめ	対照群母	対照群子				介入群母	介入群子	(%)	(%)				
研究デザイン			コンシームント	盲検化	盲検化	盲検化	盲検化	ITT	不完全アウトカムデータ	選択的アウトカム報告	早期試験中止	他のバイアス	対照	介入	アウトカム	まとめ	対照群母	対照群子	介入群母	介入群子	(%)	(%)						
研究コード																												

コメント(該当するセルに記入)

絶対効果指標の結果を記入する場合の表。

評価シート エビデンス総体用																							
対象	介入	対照	研究デザイン/研究数	ハイアリスリスク	非一貫性	不精確性	非直接性	上界要因(観察研究)	その他(出版バイアスなど)	対照群分母	介入群分母	介入群分子	介入群分子 (%)	効果指標(種類)	効果指標値	95%信頼区間	エビデンスの強さ	重要性	コメント	絶対効果	効果指標値	95%信頼区間	
CO		エビデンスの強さはROTC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(-1)”、“底(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的重要性(1~9)																					
エビデンス総体		リスク人数(アウトカム率)										対照群分子	介入群分子	介入群分子 (%)	効果指標(種類)	効果指標値	95%信頼区間	エビデンスの強さ	重要性	コメント	絶対効果	効果指標値	95%信頼区間
アウトカム		SIO																					

【4-8 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

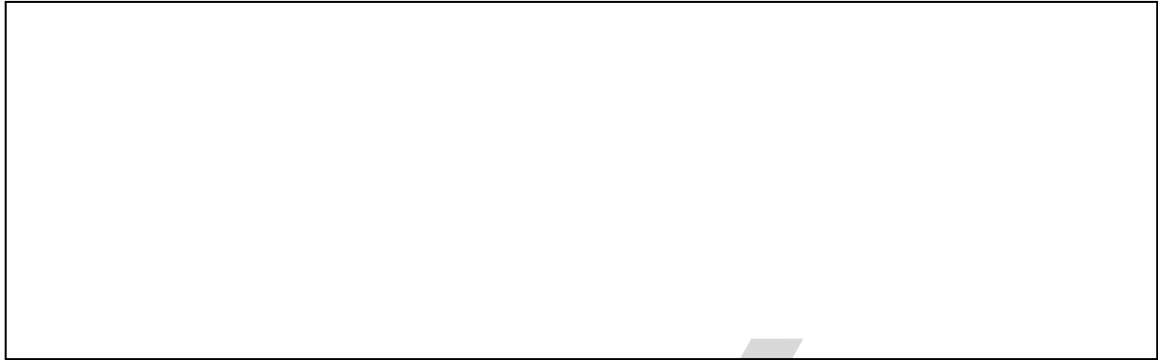
03

:
:

【4-9 メタアナリシス】

GQ			
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=
Forest plot			
	コメント:		
Funnel plot			
	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			

【4-10 SR レポートのまとめ】



【4-10-2 結果のまとめ (SoF summary of findings)】

結果のまとめ (SoF) 表						
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% CI)	期待される絶対効果* (95% CI)			エビデンス確 実度	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					⊕⊕⊕⊕ 高	
対象者数(研究数)	%	% (~)	% 少ない (~ 少ない)	Due to		
対象者1					⊕⊕⊕○ 中	
対象者数(研究数)	%	% (~)	% 少ない (~ 少ない)	Due to		
アウトカム2: 経過観察期間:					⊕⊕○○ 低	
対象者数(研究数)	%	% (~)	% 少ない (~ 少ない)			
アウトカム3: 経過観察期間:					⊕○○○ 非常に低	
対象者数(研究数)	%	% (~)	% 少ない (~ 少ない)	Due to		
アウトカム4: 経過観察期間:					⊕○○○ 非常に低	
対象者数(研究数)	%	% (~)	% 少ない (~ 少ない)			
	*					
	**					
解説						
	1					
	2					
	3					
	4					
	5					
	6					
	7					
	8					
	9					
	10					
	11					
	12					

【4-11 Future Research Question】

--



4章 記入方法

【4-1 データベース検索結果 記入方法】

タイトル:	混同のないようすべてのテンプレートにタイトルを記載する
CQ:	混同のないようすべてのテンプレートに CQ 番号と CQ 文を記載する
データベース:	データベース名とその検索期間を記載
日付:	検索日を記載
検索者:	2名の検索者を記載。公開時には削除または、イニシャルに変更すること。

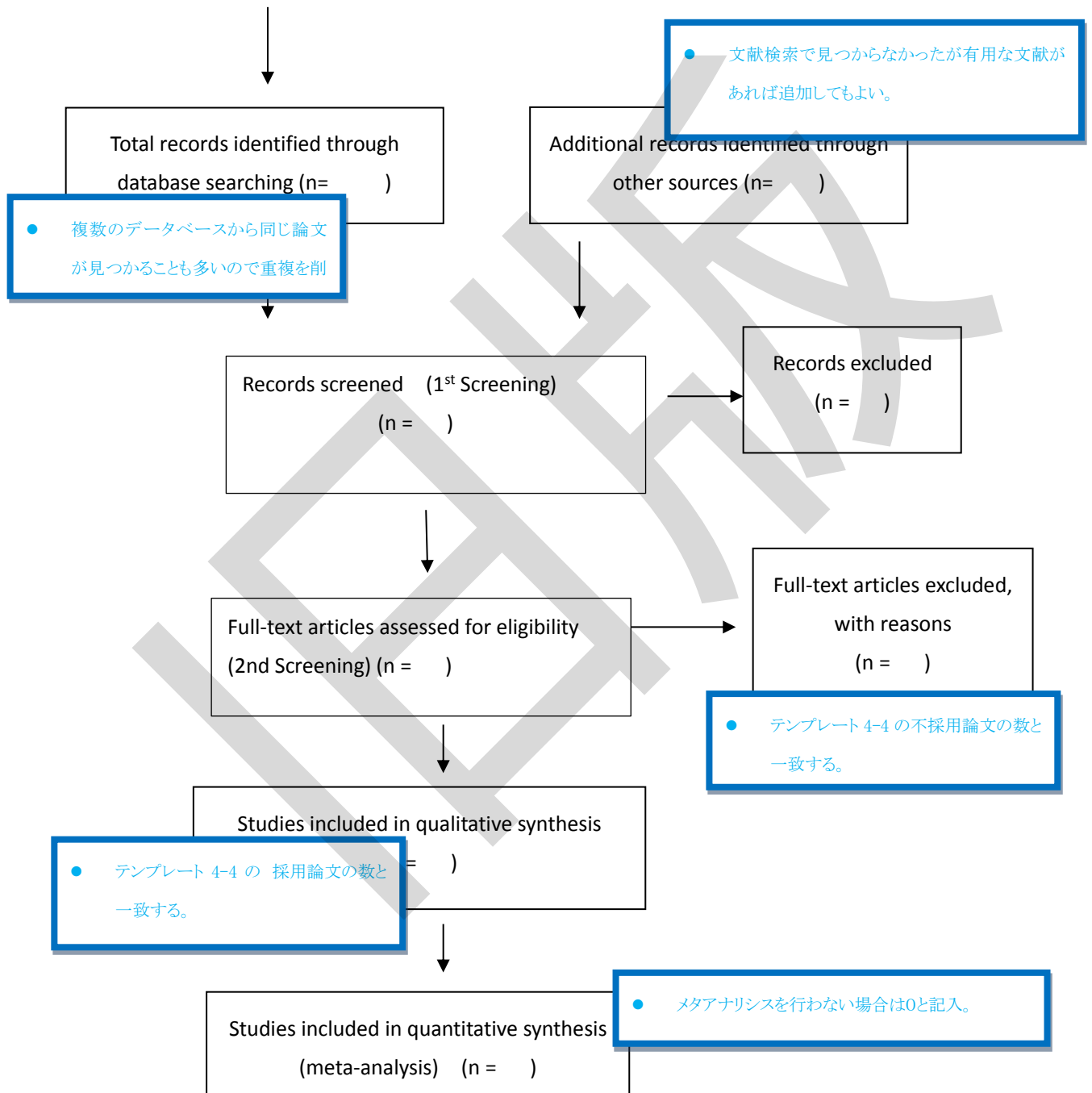
#	検索式	文献数
	検索式を AND/OR、ti/ab/kw などを含めて記載	文献数を記載
	最終検索結果は上でも下でもよいが統一して記載する	

● 検索を行ったDBの数だけ作成されるが、代表的なもののみ診療ガイドライン作成に記載してもよ


【4-2 文献検索フローチャート 記入方法】 PRISMA 声明を改変

NGC NICE PubMed Cochrane 医中誌 EMBASE WHO PsycINFO® CINAHL Others()

- 各 CQ 毎に使用した文献データベース別に、検索した研究デザインの合計文献数を記載。使用しなかったものは NA とする。CQ の数だけできるが、代表的なものを診療ガイドライン巻末に資料として掲載する。



【4-3 二次スクリーニング後の一覧表 記入方法】

文献	研究デザイン	P	I	C	O	除外	コメント
著者名、出版名 で一元的に管理		サンプル数、 セッティン グ、Pの特徴 を記載	介入の特徴	比較の 特徴	記載されて いるOとそ の形式		
		<ul style="list-style-type: none"> 二次スクリーニングには残ったが、バイアスリスクの検討などで後に除外された論文がある場合はここに明記し、コメントに理由を記載する。 					

【4-4 引用文献リスト 記入方法】

	文献 ID	書誌情報
採用論文	ハーバード式による文献整理 ID で、本文と対応させる	4-3 二次スクリーニング後の一覧表に残った、システマティックレビューに使用した論文の書誌情報
不採用論文	ハーバード式による文献整理 ID で、本文と対応させる	4-3 二次スクリーニング後の一覧表からその後のステップで削除された論文の書誌情報
その他の引用論文	ハーバード式による文献整理 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

【4-8 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。
P	CQの文章を補則する十分詳細な記述。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。	
C	比較は1:1の比較が望ましいが、必要に応じて複数の比較を	
臨床的文脈	診療のプロセスのどこに位置づけられるか、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わ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と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

:

【4-10-1 SR レポートのまとめ 記入方法】

【4-8 定性的システマティックレビュー】、【4-9 メタアナリシス】の結果をまとめる。

【4-11 Future Research Question 記入方法】

現在まで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臨床的クエスチョンを作成する。以下の項目を記述する。

- ・ リサーチクエスチョン
- ・ 背景
- ・ 可能な研究計画の概略

4章 記入例

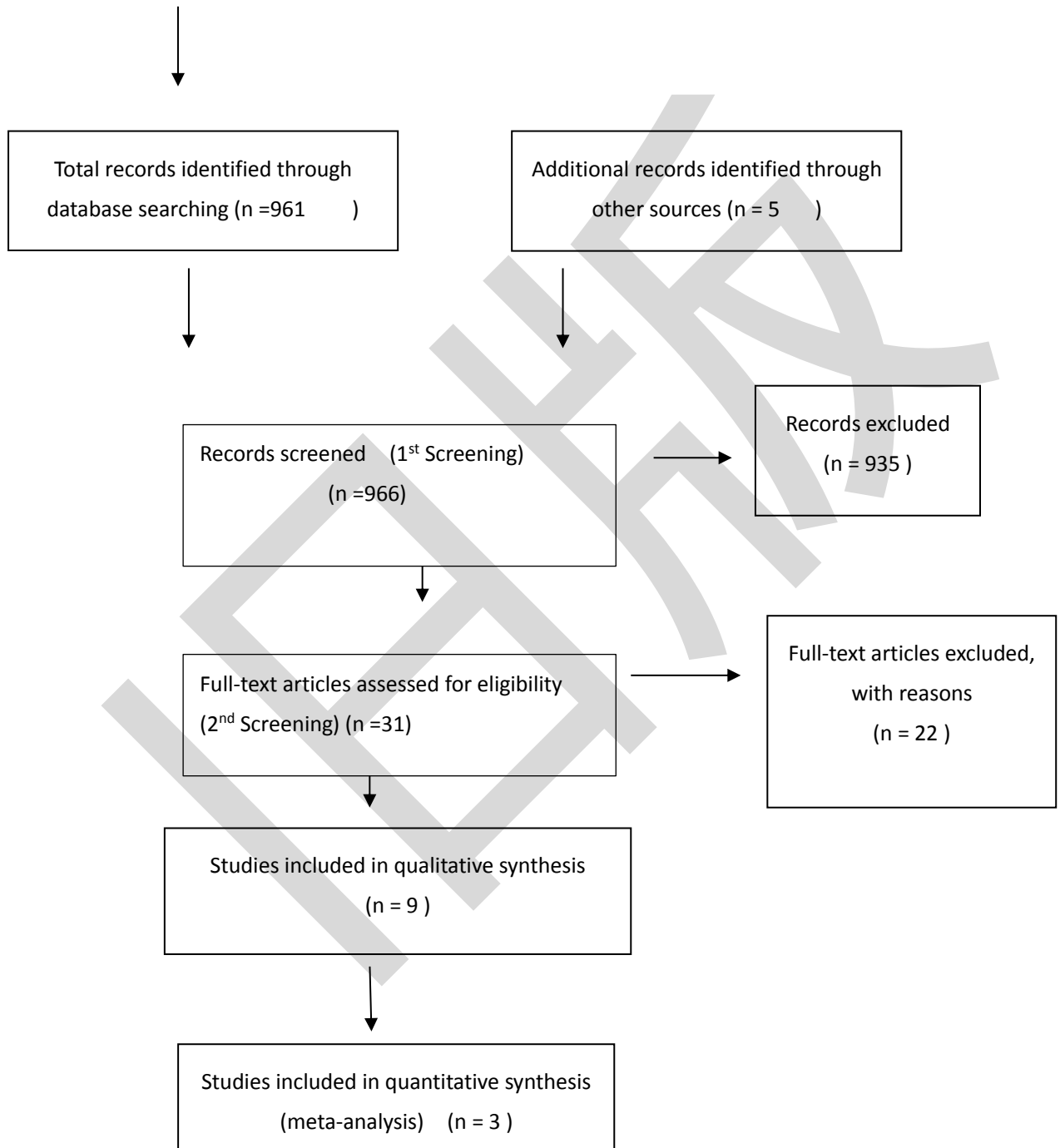
【4-1 データベース検索結果 記入例】

タイトル:	Stroke and t-PA
CO1	発症 6 時間の脳梗塞患者に対して t-PA 静脈投与は未治療に比べて総死亡を減少させるか。
データベース:	The Cochrane Library
日付:	2013/11/02
検索者:	NK/FK

#	検索式	文献数
10	#3 and #9	415
9	#6 or #7 or #8	2430
8	alteplase:ti, ab, kw	373
7	rt-pa :ti, ab, kw	501
6	tissue plasminogen activator:ti, ab, kw	2133
5	#3 and #4	4156
4	stroke:ti, ab, kw	19227
3	#1 or #2	10673
2	ischaemic:ti, ab, kw	2530
1	ischemic:ti, ab, kw	8427

【4-2 文献検索フローチャート 記入例】 PRISMA 声明を改変

NGC	NICE	PubMed	Cochrane	医中誌	EMBASE	WHO	PsycINFO®	CINAHL	Others()
78	69	354	415	74	204	NA	NA	NA	NA



【4-3 二次スクリーニング後の一覧表 記入例】

文献	研究 デザ イン	P	I	C	O	除 外	コメン ト
The IST-3 collaborat ive group, 2012	RCT	3035 patients were enrolled by 156 hospitals in 12 countries	0.9 mg/kg intraven ous rt- PA	control	independent, as defined by an Oxford Handicap Score (OHS) of 0-2 at 6 months		1617 (53%) were older than 80 ys.

【4-4 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. Lancet. 2012. 379 (9834):2352-2363.
不採用論文		
その他の引用論文		

【4-5 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																				
アウトカム		発症6カ月後の総死亡																				
	個別研究	バイアスリスク*																				
研究コード		選択バイアス	検出バイアス	症例減少バイアス	その他	非直接性*						リスク人数(アウトカム率)			効果指標(値)	信頼区間						
		ランダム化	コンシメント	盲検	盲検	盲検	盲検	盲検	盲検	盲検	盲検	盲検	対照群母	対照群母	介入群母	介入群母	介入群母	対照群母	対照群母	介入群母	効果指標(種類)	信頼区間
IST-3	RCT	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	26.8	1515	408	26.9	RR	1.006	0.89 - 1.13
EPITHET	RCT	0	-1	0	0	0	0	0	0	0	0	-1	0	-1	7	14.3	52	13	25	RR	1.75	0.76 - 4.02
ECASS II	RCT	0	0	0	0	0	0	0	0	0	0	0	-1	0	42	10.7	409	43	10.5	RR	0.979	0.65 - 1.46
コメント(該当するセルに記入)					選択的アウトカム報告	早期試験中止	他のバイアス	対象 介入 対照						アウトカム			効果指標(種類)			信頼区間		
IST-3					80歳以上が4分の1程度			用量が日本より多い						90日時点で判定								
EPITHET		4名のプロクランダム化			脱落例あり			用量が日本より多い						90日時点で判定								
ECASS II		1-3時間と3-6時間混在だが差がなし						用量が日本より多い						90日時点で判定								

* 各項目の評価は“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
 各アウトカムごとに別紙にまとめ

【4-7 評価シート エビデンス総体 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ゼ0.6mg/kg静注																																																																																						
対照	プラセボ																																																																																						
エビデンス総体	エビデンスの強さは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)																																																																																						
アウトカム	研究デザイン/研究数	バイアスリスク	非一貫性	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)	効果指標値	信頼区間	エビデンスの強さ	重要性	コメント																																																																							
6か月後の総死亡	RCT/3	0	0	0	-1	0	1960	456	0.233	1976	464	0.23	RR	1.010	0.905-1.139	A	9	差がない。																																																																					
6か月後の依存性(mRS0-2)	RCT/3	0	0	0	-1	0	1960	1226	0.626	1976	1176	0.6	RR	0.923	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	-1	0	1961	29.5	0.015	1977	145	0.07	RR	4.437	2.013-9.777	A	9	脳出血例のほとんどが死亡																																																																					
総死亡	<table border="1"> <thead> <tr> <th>Study or Subgroup</th> <th>Alteplase Events</th> <th>Alteplase Total</th> <th>Placebo Events</th> <th>Placebo Total</th> <th>Total Weight</th> <th>IV, Random, 95% CI</th> <th>Risk Ratio</th> <th>IV, Random, 95% CI</th> </tr> </thead> <tbody> <tr> <td>IST-3</td> <td>408</td> <td>1515</td> <td>407</td> <td>1520</td> <td>90.5%</td> <td>1.01 [0.89, 1.13], 2012</td> <td>1.01</td> <td>[0.89, 1.13]</td> </tr> <tr> <td>EPITHET</td> <td>13</td> <td>52</td> <td>7</td> <td>49</td> <td>1.8%</td> <td>1.75 [0.76, 4.02], 2008</td> <td>1.75</td> <td>[0.76, 4.02]</td> </tr> <tr> <td>ECASS II</td> <td>43</td> <td>409</td> <td>42</td> <td>391</td> <td>7.7%</td> <td>0.98 [0.65, 1.46], 1998</td> <td>0.98</td> <td>[0.65, 1.46]</td> </tr> <tr> <td>Total (95% CI)</td> <td>464</td> <td>1976</td> <td>456</td> <td>1960</td> <td>100.0%</td> <td>1.01 [0.91, 1.13]</td> <td>1.01</td> <td>[0.91, 1.13]</td> </tr> <tr> <td colspan="9">Total events: 464 (Alteplase), 456 (Placebo)</td> </tr> <tr> <td colspan="9">Heterogeneity: Tau² = 0.00; Chi² = 1.70, df = 2 (P = 0.43); I² = 0%</td> </tr> <tr> <td colspan="9">Test for overall effect: Z = 0.24 (P = 0.81)</td> </tr> </tbody> </table>															Study or Subgroup	Alteplase Events	Alteplase Total	Placebo Events	Placebo Total	Total Weight	IV, Random, 95% CI	Risk Ratio	IV, Random, 95% CI	IST-3	408	1515	407	1520	90.5%	1.01 [0.89, 1.13], 2012	1.01	[0.89, 1.13]	EPITHET	13	52	7	49	1.8%	1.75 [0.76, 4.02], 2008	1.75	[0.76, 4.02]	ECASS II	43	409	42	391	7.7%	0.98 [0.65, 1.46], 1998	0.98	[0.65, 1.46]	Total (95% CI)	464	1976	456	1960	100.0%	1.01 [0.91, 1.13]	1.01	[0.91, 1.13]	Total events: 464 (Alteplase), 456 (Placebo)									Heterogeneity: Tau ² = 0.00; Chi ² = 1.70, df = 2 (P = 0.43); I ² = 0%									Test for overall effect: Z = 0.24 (P = 0.81)								
Study or Subgroup	Alteplase Events	Alteplase Total	Placebo Events	Placebo Total	Total Weight	IV, Random, 95% CI	Risk Ratio	IV, Random, 95% CI																																																																															
IST-3	408	1515	407	1520	90.5%	1.01 [0.89, 1.13], 2012	1.01	[0.89, 1.13]																																																																															
EPITHET	13	52	7	49	1.8%	1.75 [0.76, 4.02], 2008	1.75	[0.76, 4.02]																																																																															
ECASS II	43	409	42	391	7.7%	0.98 [0.65, 1.46], 1998	0.98	[0.65, 1.46]																																																																															
Total (95% CI)	464	1976	456	1960	100.0%	1.01 [0.91, 1.13]	1.01	[0.91, 1.13]																																																																															
Total events: 464 (Alteplase), 456 (Placebo)																																																																																							
Heterogeneity: Tau ² = 0.00; Chi ² = 1.70, df = 2 (P = 0.43); I ² = 0%																																																																																							
Test for overall effect: Z = 0.24 (P = 0.81)																																																																																							
頭蓋内出血	<table border="1"> <thead> <tr> <th>Study or Subgroup</th> <th>Alteplase Events</th> <th>Alteplase Total</th> <th>Placebo Events</th> <th>Placebo Total</th> <th>Total Weight</th> <th>IV, Random, 95% CI</th> <th>Risk Ratio</th> <th>IV, Random, 95% CI</th> </tr> </thead> <tbody> <tr> <td>ECASS II</td> <td>36</td> <td>409</td> <td>13</td> <td>391</td> <td>44.8%</td> <td>2.65 [1.43, 4.92], 1998</td> <td>2.65</td> <td>[1.43, 4.92]</td> </tr> <tr> <td>EPITHET</td> <td>4</td> <td>52</td> <td>0</td> <td>49</td> <td>6.2%</td> <td>8.49 [0.47, 153.70], 2008</td> <td>8.49</td> <td>[0.47, 153.70]</td> </tr> <tr> <td>IST-3</td> <td>104</td> <td>1515</td> <td>16</td> <td>1520</td> <td>49.0%</td> <td>6.52 [3.87, 10.99], 2012</td> <td>6.52</td> <td>[3.87, 10.99]</td> </tr> <tr> <td>Total (95% CI)</td> <td>144</td> <td>1976</td> <td>29</td> <td>1960</td> <td>100.0%</td> <td>4.43 [2.07, 9.48]</td> <td>4.43</td> <td>[2.07, 9.48]</td> </tr> <tr> <td colspan="9">Total events: 144 (Alteplase), 29 (Placebo)</td> </tr> <tr> <td colspan="9">Heterogeneity: Tau² = 0.24; Chi² = 4.95, df = 2 (P = 0.08); I² = 60%</td> </tr> <tr> <td colspan="9">Test for overall effect: Z = 3.83 (P = 0.0001)</td> </tr> </tbody> </table>															Study or Subgroup	Alteplase Events	Alteplase Total	Placebo Events	Placebo Total	Total Weight	IV, Random, 95% CI	Risk Ratio	IV, Random, 95% CI	ECASS II	36	409	13	391	44.8%	2.65 [1.43, 4.92], 1998	2.65	[1.43, 4.92]	EPITHET	4	52	0	49	6.2%	8.49 [0.47, 153.70], 2008	8.49	[0.47, 153.70]	IST-3	104	1515	16	1520	49.0%	6.52 [3.87, 10.99], 2012	6.52	[3.87, 10.99]	Total (95% CI)	144	1976	29	1960	100.0%	4.43 [2.07, 9.48]	4.43	[2.07, 9.48]	Total events: 144 (Alteplase), 29 (Placebo)									Heterogeneity: Tau ² = 0.24; Chi ² = 4.95, df = 2 (P = 0.08); I ² = 60%									Test for overall effect: Z = 3.83 (P = 0.0001)								
Study or Subgroup	Alteplase Events	Alteplase Total	Placebo Events	Placebo Total	Total Weight	IV, Random, 95% CI	Risk Ratio	IV, Random, 95% CI																																																																															
ECASS II	36	409	13	391	44.8%	2.65 [1.43, 4.92], 1998	2.65	[1.43, 4.92]																																																																															
EPITHET	4	52	0	49	6.2%	8.49 [0.47, 153.70], 2008	8.49	[0.47, 153.70]																																																																															
IST-3	104	1515	16	1520	49.0%	6.52 [3.87, 10.99], 2012	6.52	[3.87, 10.99]																																																																															
Total (95% CI)	144	1976	29	1960	100.0%	4.43 [2.07, 9.48]	4.43	[2.07, 9.48]																																																																															
Total events: 144 (Alteplase), 29 (Placebo)																																																																																							
Heterogeneity: Tau ² = 0.24; Chi ² = 4.95, df = 2 (P = 0.08); I ² = 60%																																																																																							
Test for overall effect: Z = 3.83 (P = 0.0001)																																																																																							

【4-7 評価シート エビデンス総体 観察研究 記入例】

評価シート エビデンス総体用		エビデンス総体										コメント(該当するセルに記入)						
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																	
エビデンスの強さは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)"、"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)																		
アウトカム	研究デザイン/研究数	バイアスリスク	非一貫性	不精確	非直接性	その他(出版バイアスなど)	上昇要因(観察研究)	対照群母	対照群子	(%群分)	介入群母	介入群子	効果指標(種類)	効果指標(統合値)	信頼区間	エビデンスの強さ	重要性	コメント
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	NA	NA	NA	弱(C)	8	PSL+AZPまたはPSL単独は寛解導入に有効と考えられる。
リスク人数(アウトカム率)																		
軽症例のみ。寛解率は81%。																		
コメント(該当するセルに記入)																		
Czajka 2008の研究介入群は1つにまとめらる。																		

【4-9 メタアナリシス 記入例】

CQ	CQ19 虚血性脳卒中患者で発症後6時間以内のt-PA投与は非投与と比べ6ヵ月後の自立性を改善するか？																																																														
P	発症後6時間以内の虚血性脳卒中	I	t-PA静脈内1回投与																																																												
C	プラセボ	O	6ヵ月後の自立性																																																												
研究デザイン	RCT	文献数	3																																																												
コード	IST-3, EPITHET, ECASS II																																																														
モデル	ランダム効果	方法	Inverse-variance method (RevMan5.2)																																																												
効果指標	リスク比	統合値	0.92 (0.84 - 1.03) P= 0.17																																																												
Forest plot	<table border="1"> <thead> <tr> <th rowspan="2">Study or Subgroup</th> <th colspan="2">Alteplase</th> <th colspan="2">Placebo</th> <th rowspan="2">Weight</th> <th colspan="2">Risk Ratio</th> <th rowspan="2">Year</th> </tr> <tr> <th>Events</th> <th>Total</th> <th>Events</th> <th>Total</th> <th>IV, Random, 95% CI</th> <th>IV, Random, 95% CI</th> </tr> </thead> <tbody> <tr> <td>IST-3</td> <td>961</td> <td>1515</td> <td>986</td> <td>1520</td> <td>60.1%</td> <td>0.98</td> <td>[0.93, 1.03]</td> <td>2012</td> </tr> <tr> <td>EPITHET</td> <td>28</td> <td>52</td> <td>29</td> <td>49</td> <td>8.3%</td> <td>0.91</td> <td>[0.65, 1.28]</td> <td>2008</td> </tr> <tr> <td>ECASS II</td> <td>187</td> <td>409</td> <td>211</td> <td>391</td> <td>31.6%</td> <td>0.85</td> <td>[0.74, 0.97]</td> <td>1998</td> </tr> <tr> <td>Total (95% CI)</td> <td></td> <td>1976</td> <td></td> <td>1960</td> <td>100.0%</td> <td>0.93</td> <td>[0.84, 1.03]</td> <td></td> </tr> <tr> <td>Total events</td> <td>1176</td> <td></td> <td>1226</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>Heterogeneity: Tau² = 0.00; Chi² = 3.62, df = 2 (P = 0.16); I² = 45% Test for overall effect: Z = 1.37 (P = 0.17)</p> <p>コメント: $NNT=1/(1-0.92)=12.5$程度の効果が期待されるという結果であるが、有意水準に達していない。</p>			Study or Subgroup	Alteplase		Placebo		Weight	Risk Ratio		Year	Events	Total	Events	Total	IV, Random, 95% CI	IV, Random, 95% CI	IST-3	961	1515	986	1520	60.1%	0.98	[0.93, 1.03]	2012	EPITHET	28	52	29	49	8.3%	0.91	[0.65, 1.28]	2008	ECASS II	187	409	211	391	31.6%	0.85	[0.74, 0.97]	1998	Total (95% CI)		1976		1960	100.0%	0.93	[0.84, 1.03]		Total events	1176		1226					
Study or Subgroup	Alteplase		Placebo		Weight	Risk Ratio		Year																																																							
	Events	Total	Events	Total		IV, Random, 95% CI	IV, Random, 95% CI																																																								
IST-3	961	1515	986	1520	60.1%	0.98	[0.93, 1.03]	2012																																																							
EPITHET	28	52	29	49	8.3%	0.91	[0.65, 1.28]	2008																																																							
ECASS II	187	409	211	391	31.6%	0.85	[0.74, 0.97]	1998																																																							
Total (95% CI)		1976		1960	100.0%	0.93	[0.84, 1.03]																																																								
Total events	1176		1226																																																												
Funnel plot	<p>コメント: 出版バイアスを示唆する分布は認められない。</p>																																																														
その他の解析	施行せず。		コメント:																																																												
メタリグレーション																																																															
感度分析																																																															

【4-10-2 結果のまとめ表 SoF 記入例】

結果のまとめ (SoF) 表						
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もなわないもの					
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+クラブリ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% CI)	期待される絶対効果 (95% CI)			エビデンスの 確実性	何が起きるか？
		対照	介入	差		
虫垂炎の再発					⊕⊕⊕⊕⊕	
1年					高	
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)	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ヶ月以内に手術を受けずに済んでも、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ヶ月以内)					⊕⊕⊕⊕⊖	
1年					中	
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)	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ヶ月間手術を受けずに済む者が90%程度いる。
主要な合併症					⊕⊕⊕⊖⊖	
1年					低	
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)	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見ると、抗菌薬投与群のほうが手術を受けずに済む者が多いので、合併症は少なくなる。
その他の合併症					⊕⊕⊕⊖⊖	
1年					中	
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)	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見ると、抗菌薬投与群のほうが手術を受けずに済む者が多いので、合併症は少なくなる。
入院期間					⊕⊕⊕⊕⊖	
0~14日					中	
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)	バイアスリスク、 非直接性 ^{1,2} が認められるが、1日以上の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					低	
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)	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、 ^{1,2} 統計量が大きく非一貫性に問題がある。						