

Beginners' Training Sheet for Systematic Review

ver.5.0 by last updated on March 20, 2012
 南郷 栄秀 Eishu NANGO, MD, PhD
<http://spell.umin.jp>

このシートは初めてシステマティックレビューの論文を読むためのものです。システマティックレビューの定義と論文の構造にも触れながら、論文を読む上でのポイントを解説しました。

なお、このシートに関する質問、改善点などは、制作者まで直接お願いします。また、制作者は著作権を保持し、無断転載を禁止します。再配布に制限はしないつもりですが、再配布する際は制作者までご一報ください。

採用論文

Reviewer: _____ 年 月 日

authors : _____

title : _____

citation : _____

PubMed PMID : _____

quick check list

1. 論文の PICO は何か？
2. 全ての研究を網羅的に集めようと努力したか？
 - ① 検索に用いた文献データベースは何か？
 - ② どのような検索語で、どの期間の研究を調べたか？
 - ③ どのような種類の研究を調べたか？
 - ④ 個々の論文の参考文献まで追跡して調べたか？
 - ⑤ 個々の研究者や専門家に連絡を取ったか？
 - ⑥ 出版されていない研究も探したか？
 - ⑦ 同じ研究が複数報告されていないか？
 - ⑧ 英語以外で書かれた研究も探したか？
3. 全ての研究が網羅的に集められたか？
4. 集められた研究は、複数の評価者によって評価されたか？
5. 集められた研究は、明確な基準をもって妥当性を評価されたか？
6. 集められた研究の異質性 heterogeneity は検討されたか？
7. 結果は統合されたか (Meta-analysis)？
 - ① 最終的に何件の研究が残り、採用されたか？
 - ② 集められた研究の結果は統合されたか？
8. 結果の評価

0 システマティックレビューという研究手法

0-1) システマティックレビューとは？

システマティックレビュー Systematic review は、特定の疑問に関して、数多くの研究を網羅的に再現性のある方法に従って集め、その時点における結果のまとめを行ったもの (図 1 の b と c) である。多くは、ランダム化比較試験 Randomized controlled trial; RCT をまとめたものであるが、コホート研究や診断の研究を統合したものもある。このシートでは、RCT をまとめたシステマティックレビューの読み方について解説する。

定義：明確に定式化された疑問について、関連する研究を特定し、選択し、批判的に吟味し、レビューに組み入れられた研究から得られるデータをまとめて解析する、系統的で明示的な方法を用いて行うレビューのこと (Cochrane library, 1998)。**論文には、タイトルにシステマティックレビュー Systematic review やメタアナリシス Meta-analysis と書かれていることが多い。**

目的：対象となる疑問についての現在利用可能な全ての研究データをまとめること

特徴：① 問題とするトピックに関する研究を網羅的に探し出す膨大な努力をしていること

② 各論文を批判的に評価していること

③ 事前に定められた質評価基準を満たす研究のみを統合して結論が導かれていること

一方、システマティックレビューでない一般のレビュー（総説，図 1 の a）は**ナラティブレビューNarrative review** と呼ばれ，著者の主観的な意見を述べたものにすぎない。ナラティブレビューの多くは専門家によって書かれたものであるが，あらゆる研究が網羅されているわけではなく，著者の意見を支持する先行研究の結果だけを都合良く引用している可能性が高く，引用された研究の質もバラバラである。時にシステマティックレビューとは異なる結論が導き出されることもある。

現在利用可能なシステマティックレビューのうち最も質が高く信頼性のおけるものは，コクラン共同計画 Cochrane Collaboration によって作成されたシステマティックレビュー（Cochrane Database of Systematic Review; CDSR）であり，コクランライブラリーCochrane Library: <http://www.thecochranelibrary.com/>（有料）にデータベース化されている。

0-2) メタアナリシス Meta-analysis とは？

前述のように，システマティックレビューは，研究結果を“体系的に集める”ことに重きが置かれている。これに対して，メタアナリシス Meta-analysis とは，この用語を最初に使用した Glass (1976) の定義によれば，“研究結果を統合する目的で，個々の研究から得られた解析結果の膨大なコレクションに対して実施する統計解析”である。すなわち，集められた複数の研究の結果を，統計学的手法を用いて統合したものをメタアナリシスと呼ぶ（図 1 の c と d）。

メタアナリシスが行われないシステマティックレビュー（図 1 の b）では，個々の研究結果を羅列するのみにとどまったり，統計学的解析を行うことなく著者の直感で結論が導き出されたりする。これに対して，メタアナリシスでは，一定の統計学的手法で統合するため，誰が行っても同じ結論を得ることができる。しかし，統合の手法が優れていても，手当たり次第に研究結果を集めてきたのでは，統合された結果にはバイアスが含まれてしまう（図 1 の d）。したがって，システマティックレビューであり，かつメタアナリシスが行われているものが，結論を信頼することができる（図 1 の c）。

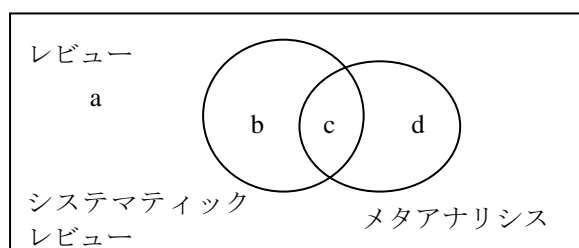


図 1：システマティックレビューとメタアナリシスの関係

論文の構造

要約 abstract, summary

緒言 introduction

方法 methods

←チェックすべき項目はほとんどここにある！

結果 results

考察 discussion

研究が扱っている題材は，「要約 abstract, summary」に記載されており，これは PICO でまとめることができる。ただし，「要約 abstract, summary」の部分だけでは情報が不十分なことが多く，論文の「方法 methods」の項で詳細を確認することが必要である。

1 論文の PICO は何か？

PICO とは疑問を定式化したものであり，どんな患者が (P; Patient)，どんな治療や検査を受けるのは (I; Intervention)，何と比べて (C; Comparison)，どうなるか (O; Outcome) を一文にまとめたものである。

患者 P(Patient)

介入 I(Intervention)：

比較 C(Comparison)：

結果 O(Outcome)：

システマティックレビューには複数の研究結果が集められており，個々の研究の PICO（年齢や病期などの患者背景，投与薬剤の用量など）は互いに異なることが多い。したがって，システマティックレビューの批判的吟味で PICO を探る際は，おおよその内容を把握するだけでよい。

よく分からなければ，導入 Introduction の一番最後をみると良い。この研究で検証しようとしている研究仮説に書いて簡潔に書かれている。おおよその PICO を把握するためのキーワードは，“the aim of this study”である。

2 全ての研究を網羅的に集めようと努力したか？

研究はどのようにして集められたか？（研究を集めるフローチャートがあれば参照する）

① 検索に用いた文献データベースは何か？

- MEDLINE
- EMBASE
- CENTRAL (Cochrane Central Register of Controlled Trials)
- Google scholar
- その他のデータベース()

② どのような検索語で、どの期間の研究を調べたか？

③ どのような種類の研究を調べたか？

- ランダム化比較試験 (RCT), 準ランダム化比較試験 (quasi-RCT), 臨床対照試験 (non-RCT, CCT)
- システマティックレビュー (SR), メタアナリシス (MA), コクランレビュー (CDSR)
- コホート研究
- 症例対照研究
- 診断研究 (横断研究)
- その他の種類の研究()

④ 個々の論文の参考文献まで追跡して調べたか？

- 参考文献まで調べた
- 参考文献は調べなかった, 不明

⑤ 個々の研究者や専門家に連絡を取ったか？

- 連絡を取った
- 連絡を取らなかった, 不明

⑥ 出版されていない研究も探したか？

- 探した
- 探さなかった, 不明

⑦ 同じ研究が複数報告されていないか？

- 複数報告されている研究は排除されている
- 複数報告されている研究が排除されていない, 不明

⑧ 英語以外で書かれた研究も探したか？

- 探した
- 探さなかった, 不明

研究を集める際には、通常以下の流れに従う。

- ① 研究テーマである PICO からキーワードを抜き出し、それを用いてデータベースを検索する。
- ② ①を補完する形で参考文献を調べたり、専門家に直接連絡を取ったりして、データベースに含まれていない研究や未発表のデータを探す。
- ③ 集めた研究のうち、質の低い研究や他の研究とデータが重複しているもの、PICO が異なるものなどを除外する。

システマティックレビューでは、研究の集め方のルールとして inclusion criteria と exclusion criteria を定めていることがある（これは個々の研究で研究参加者を集めるために定めている inclusion / exclusion criteria とは異なるので混同しないこと）。右の図のようなフローチャートが本文中にあれば、それを参考にとすると良い。この研究では、キーワードを用いてデータベースから論文を検索した後、RCT でないものを 209 件、重複報告を 70 件、現在進行中の試験を 5 件除外して固有の RCT のみ残し、さらに exclusion criteria に該当するものを除外して、最終的に 13 件をレビューの対象としている。

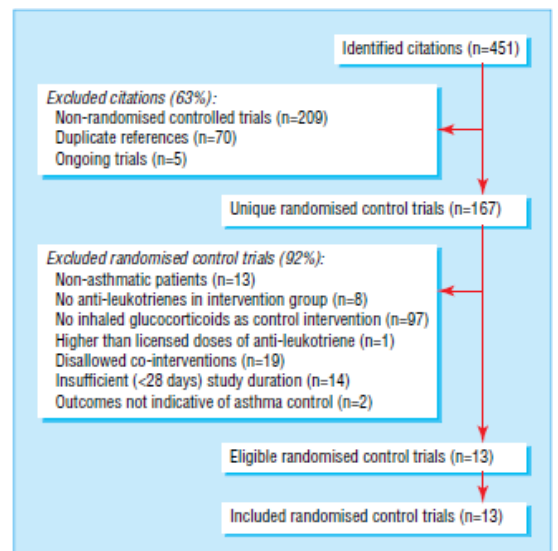


Fig 1 Selection process of eligible randomised controlled trials from all identified citations

研究の検索の手順は **Methods** の冒頭に記載がある。以下の点をチェックする。

- ① 検索に用いた文献データベースは何か？：主として Medline だが、これだけでは不十分である。他に EMBASE, CINAHL, CDSR (Cochrane Database of Systematic Review), CENTRAL (Cochrane Library に収載されている RCT のデータベース), さらに個々の専門領域のデータベースなども併用されるべきである。
- ② どのような検索語で、どの期間の研究を調べたか？：検索範囲は広すぎても狭すぎても良くない。
- ③ 個々の論文の参考文献を追跡したか？：データベースにはない研究が参考文献にあるかも知れない。“screened bibliographies” とか “search reference literature” などと記載されていることが多い。
- ④ 研究者や専門家に連絡を取ったか？：論文には書かれていないデータも集めたか？ “contact to researchers” や “contact to authorities” などと記載されていることが多い。
- ⑤ 出版されていない研究も探したか？：出版バイアス publication bias (治療効果が認められなかった場合など、インパクトの弱い研究結果は発表されにくい傾向にある) を排除する。ClinicalTrials.gov (<http://clinicaltrials.gov>) などの臨床試験の登録データベースが参照されることもある (→次項ファンネルプロットを参照)。
- ⑥ 同じ研究が複数報告されていないか？：同じ研究が複数報告されている (サブ解析や二重投稿など) とその結果に引きずられる可能性がある。このチェック項目については記載されていない論文も多い。論文では、「二重投稿」は “duplicate reports” と書かれていることが多い。また、論文選択のフローチャートの図中に除外された旨が書かれていることもある。
- ⑦ 英語以外で書かれた研究も探したか？：言語バイアス language bias を排除する。例えば、結果に差がなかった論文はインパクトが低いので英語で発表されず、日本語で発表されているかもしれない。英語の文献のみを抽出したものは、効果が過大評価されていると考えるべきである。“no language restrictions” などと記載されていることが多い。

3 全ての研究が網羅的に集められたか？

網羅的に集められている

- ファンネルプロットを用いて出版バイアスの有無を検討している
- ファンネルプロットは用いられていない

網羅的には集められていない

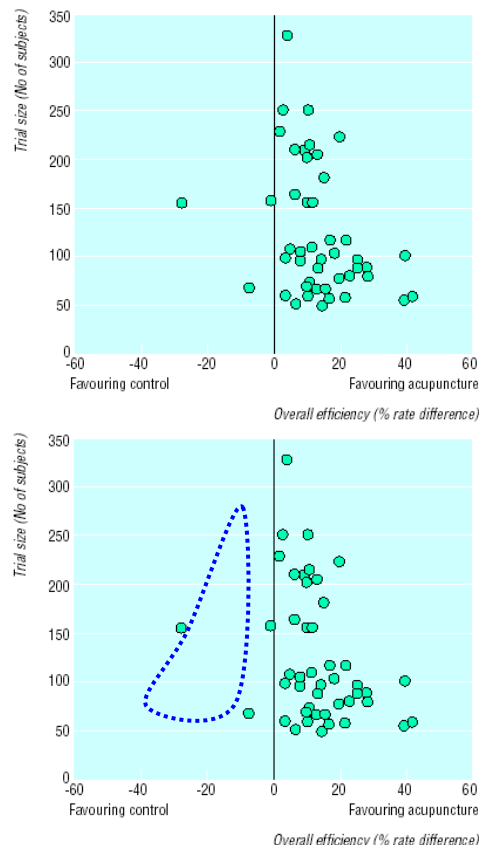
不明

研究を網羅的に集めようとするのと、実際に洩れなく集めることができたかどうかは、別の問題である。集められた研究が本当に網羅的であるかどうか、本来存在するはずの研究結果が抜けていないかどうかを検討されているか確認する。網羅的に集められているかどうかの記載箇所を探すためのキーワードは、“funnel plot”, “publication bias”, “Egger’ regression test” である。

ファンネルプロット funnel plot

ファンネルプロット funnel plot は、システマティックレビューにおいて出版バイアスの存在を評価するために用いられる。右図の例に示すように、ファンネルプロットは通常、横軸に試験の治療効果 (例えばオッズ比や有効率の差) を、縦軸には治療効果の精度 precision (多くは症例数や分散によって規定される) を示し、この 2 次元上に個々の臨床試験の結果がプロットされる。

プロットした結果は理想的には、漏斗を逆さまにしたような左右対称の形になるはずである。しかし、一般的に研究から得られた治療効果が小さいほど、その研究結果が発表される可能性は低くなる。このような出版バイアスが存在する場合、ファンネルプロットの下方の片側は空白となる。本来そこには、介入を無効とする少数例の研究の結果が存在するはずであるが、これらが発表されていないために、出版済みの研究のみが統合された結果は、真の効果よりも高い評価となってしまふ。ファンネルプロットの図から、統合された結果が有効か無効のどちらの方に傾くかを予想しておくことが重要である (右図では左下が欠けているので、集められた研究を統合した結果は真の効果よりも効果がある方に傾いている)。ファンネルプロットは、論文に示されていなくても、個々の論文のデータをもとに自分で作成することができる。なお、出版バイアス以外にファンネルプロットで左右が対象とならない原因としては、研究の質のバラツキがある。



Funnel plot of overall efficacy of acupuncture in treatment of stroke (49 trials), according to trial size
ファンネルプロットの例 (BMJ 1999;319:160 より)

4 集められた研究は、複数の評価者によって評価されたか？

- 複数の評価者 → 何人で評価されたか？
 - 評価者間で評価のくい違いが生じた場合
 - 合意を形成して最終的に評価を下している
 - 合意を形成せず、各評価者の判断を個別に記載している
 - その他
- 単独の評価者
- 不明

評価者の数は Methods の欄に記載がある。

集められた研究を単独の評価者が評価すると、研究の妥当性の評価に偏りが出る可能性がある。このため、集められた研究の評価は、**複数の評価者で独立**に行うことが望ましい。評価者間で評価にくい違いが生じた場合は、話し合いにより最終的な判断が下されることが多い。

5 集められた研究は、明確な基準をもって妥当性を評価されたか？

- 評価基準として、明確な基準を持って評価を行った
 - Jadad score
 - それ以外の基準
 - どのような評価か？
 - 集められた個々の介入研究論文の Cochrane Handbook for Systematic Reviews of Interventions 5.0.1⁸⁾における評価基準
 - ランダム割付けの方法
 - 割付け方法の隠蔽化 concealment
 - 割付け内容の盲検化（患者，研究者，outcome 評価者，解析者）
 - 脱落者，除外者の扱い
 - 研究前に予め決められたアウトカムで結果を報告しているか
 - それ以外の起こりうるバイアス
- 明確な基準はない

論文評価の基準は Methods の欄に記載がある。

集められた研究は、明確な基準を持って評価される必要がある。最も良い評価方法は、それぞれの論文の良否をスコア化して行うことである。

RCT の質を評価するスコアのうち唯一妥当性が証明されているものは、Jadad score（ハダッド・スコア）と呼ばれるものである。Cochrane Library のシステマティックレビューでは以前はこのスコアが用いられていたが、チェック項目についての記載をされていない方が、不適切に記載されている場合よりもスコアがよくなることから、評価の妥当性に疑問があり、最近では用いられなくなってきている。代わりに RCT の批判的吟味を行う際の個々のチェック項目を評価基準としている。ただし、スコアを合計するのも無意味とされている。

Jadad score⁹⁾

1. その研究はランダム割付けと明示されているか？（1=はい，0=いいえ）
 2. ランダム割付けの方法について明示されていて，かつ適切か？（0=明示されていない，1=明示されていてかつ適切），-1=明示されているが不適切）
 3. その研究はダブルブラインドと明示されているか？（1=はい，0=いいえ）
 4. ダブルブラインドの方法について明示されていて，かつ適切か？（0=明示されていない，1=明示されていてかつ適切），-1=明示されているが不適切）
 5. 試験の最初と最後の各治療群の患者数を明らかにするために，投与中止 withdrawals や脱落 dropout が記載されているか？（1=はい，0=いいえ）
- 0～5 点で表記。概ね 3 点以上であれば，比較的質が高いと判定される。

RCT 以外の研究デザインのシステマティックレビューでは，チェック項目も異なる。

6 集められた研究の異質性 heterogeneity は検討されたか？

- 異質性は検討された
 - 用いられた統計手法は？
 - Cochran Q (カイニ乗検定)
 - I² 統計量
 - その他
- 異質性は検討されていない
- 不明

異質性の検定についての記載は Methods の Statistical analysis の欄に記載がある。

「同質」とは、集められた全ての研究にバラツキがないことを指す。異質性 heterogeneity は統計学的異質性 statistical heterogeneity ともいい、研究の PICO のバラツキである臨床的異質性 clinical heterogeneity (臨床的多様性 clinical diversity) と、研究手法の質にバラツキをもたらす方法論的異質性 methodological heterogeneity (方法論的多様性 methodological diversity) の両方が含まれる。

各研究における効果の大きさが大きくバラついており、例えば信頼区間が重ならない場合は異質性があると考えられる。このため、**Cochran Q 統計量**と呼ばれる各研究間の結果のバラツキの度合いを計算し、これが有意なものかどうか、**カイニ乗検定**を用いて統計学的に検討する。ここでは、異質性の検討がなされたかだけを確認し、異質性の検定の結果については「8. 結果の評価」の項で確認する。

7 結果は統合されたか (Meta-analysis) ?

- ①最終的に何件の研究が残り、採用されたか？
- | | | |
|------------|-------|-----------------|
| 合計 | () 件 | 症例数 () 例 |
| ランダム化比較試験 | () 件 | コホート研究 () 件 |
| 準ランダム化比較試験 | () 件 | 症例対照研究 () 件 |
| 非ランダム化比較試験 | () 件 | 診断(横断)研究 () 件 |
| SR・MA・CDSR | () 件 | その他の種類の研究 () 件 |
- ②集められた研究の結果は統合されたか？
- 統合されている
 - 統合する際に用いたモデルは何か？
 - Fixed effect model 固定(母数)効果モデル
 - Random effects model ランダム(変量)効果モデル
 - その他のモデル
 - 統合されていない

最終的に何件の研究が残ったか、また研究の結果が統合されたかどうかは Results の欄に記載がある。

最終的に研究結果が統合されていない場合、対象となる疑問に対する結論は曖昧になる。複数の研究結果について統計学的手法を用いて統合し、1つの総合的な統計量で示したものをメタアナリシス Meta-analysis という。

統計学的手法で統合する際には、**Fixed effect model** や **Random effects model** が用いられる。集められた研究結果のバラツキはもっぱら偶然誤差であると仮定するのが **Fixed effect model** であり、この偶然によるバラツキ以外に、研究間に、プロトコールの違い、患者の違い、地域の違いなどといった無視できない違い (heterogeneity) としての各研究の偏りも関与していると考えるのが **Random effects model** である。**Random effects model** の方が信頼区間の幅が若干広く、有意差が出にくくなる。

具体的な統計手法としては、Outcome 指標の種類に応じて、**Fixed effect model** では、Mantel-Haenszel method, Peto method, General variance-based method が、**Random effects model** では DerSimonian-Laird method が用いられる。

Fixed effect model と **Random effects model** のどちらが用いられるべきかについては議論が多く、どちらが優れているということはない。異質性 heterogeneity が低いと判定された場合には、いずれの方法を用いても同じ結論に達するので、どの方法が用いられているかは重要ではなくなる。しかし、異質性が高い場合は、**Random effects model** を用いるべきである。

異質性が極めて高い場合は、結果を統合しないという決断が下されることもある。あるいは、patient を性別、重症度、年齢など、サブグループに分けることで異質性がなくなる場合には、サブグループ解析 subgroup analysis を行って統合することができる場合もある。

元の研究にバイアスが含まれている可能性が高い場合や、元の研究同士が著しく異なる場合は、統合してはいけない。

8 結果の評価

結果はどのように示されたか？フォレストプロットが示されていれば、これを読む。

- 票数計算 vote counting
- オッズ比 odds ratio
- リスク比 risk ratio (相対危険度 relative risk)
- リスク差 risk difference
- 平均値の差 mean difference
- その他

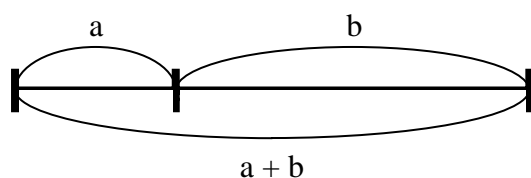
結果の大きさはどの程度か？：

1) 結果の示し方

票数計算 vote counting は、効果ありの研究の数と効果なしの研究の数を単純に数えて比較したもの（有効：無効=○：△のようなもの）である。この場合、いずれの研究の質も結果も、重みが等しいとみなされている。票数計算では、その介入が実際にはどの程度の効果を示すのかが分からないので、この方法で効果の有無を表現してはいけない。

2) オッズ比とリスク比

Meta-analysis で統合される Outcome 指標には、オッズ比、リスク比、リスク差、平均値の差などがあるが、それぞれの 95%信頼区間 95%CI や p 値 p-value も評価すべきである。



	Outcome		
	(+)	(-)	
介入群	a	b	a+b
対照群	c	d	c+d
	a+c	b+d	a+b+c+d

右上の表で、outcome 発生率と、介入群と対照群の

outcome 発生率の比であるリスク比は、以下のように計算される。

$$\begin{aligned} \text{介入群の Outcome 発生率} &= \frac{a}{a+b} & \text{対照群の Outcome 発生率} &= \frac{c}{c+d} \\ \text{リスク比 Risk ratio; RR} &= \frac{a(c+d)}{(a+b)c} \end{aligned}$$

オッズは率と似て非なるものであり、ある事象が起こったものと起こらなかったものの比である。また、介入群のオッズと対照群のオッズの比をオッズ比といい、以下のように計算される。

$$\begin{aligned} \text{介入群の Outcome 発生オッズ} &= \frac{a}{b} & \text{対照群の Outcome 発生オッズ} &= \frac{c}{d} \\ \text{オッズ比 Odds ratio; OR} &= \frac{ad}{bc} \end{aligned}$$

リスク比とオッズ比の式を見比べると分かるように、 $a+b$ を b と、 $c+d$ を c と置き換えることができれば、両者の値は一致する。 a や c が小さい場合、つまり有病率の低い疾患ほど、オッズ比はリスク比に近似できる。

最もよく用いられるのはオッズ比 odds ratio: OR である。治療群と対照群の効果に差が無ければ、RR, OR は 1 に等しくなり、治療に効果がある場合は RR と OR は共に 1 未満、効果がない場合には RR, OR は 1 を越える。また、95%信頼区間 95%CI が 1 をまたいでいるかどうかによって、その結果が統計学的に有意かどうか判断できる。対数オッズ logarithmic odds で示される場合には、ゼロ点から両側に距離の等しい対数スケールにプロットされる。

オッズ比と患者イベント予想発生率 (PEER) から NNT が計算できる (はじめてトライアルシート参照)。この患者イベント予想発生率は、個々の患者でどれくらいの高さになるかを見積もるものである。

$$\text{NNT} = \frac{1 - \{\text{PEER} \times (1 - \text{OR})\}}{(1 - \text{PEER}) \times \text{PEER} \times (1 - \text{OR})}$$

4) 結果の評価の仕方

メタアナリシスされている場合は、フォレストプロット forest plot を中心に結果の評価を行う。

PICO 毎に、統合された結果を評価する

→フォレストプロット毎の PICO を確認する（タイトルと横軸を見ると分かりやすい）

- ①統合された結果に有意差があるか？
- ②統合された結果の大きさ（点推定値と信頼区間）は？
- ③集められた研究に異質性 heterogeneity はあるか？

異質性は検討されていない

異質性は検討されている

異質性 heterogeneity の検定

○Cochran Q（カイニ乗検定）

有意差なし →同質

有意差あり →異質

○I² 統計量

0～25% →異質性が低い

25～50% →異質性が中程度

50～75% →異質性が高い

75～100% →異質性が極めて高い

- ④異質性が高い場合には、高い原因は何か？

○患者背景が異なる

○介入内容が異なる

○アウトカムの定義が異なる

○研究の質が異なる

- ⑤異質性の高いものに対してサブ解析、感度分析が行われていれば、その結果を評価する

結果は PICO 毎に評価する。特に、全体を統合した結果と、Patient, Intervention/Comparison, Outcome の内容によってサブグループ解析が行われることが一般的であるので、どの PICO のメタアナリシスで有意な差が出ているのかを確認する。

異質性は、「6. 集められた研究の異質性 heterogeneity は検討されたか？」の項に示したとおり、**Cochran Q 統計量**と**カイニ乗検定**を用いて統計学的に検討する。カイニ乗検定の結果はフォレストプロット（結果の図）の下の方に記載されていることが多い（8 ページ参照）。但し、Cochran Q 統計量のカイニ乗検定では検出力（power）が小さいため、集められた研究数が少ないときには有意水準（異質か同質かの判定基準）には 5%ではなくて 10%が用いられることもある（すなわち $p < 0.10$ で有意差あり）。最近では、この Cochran Q の欠点を補うため **I² 統計量**が用いられる。これは Cochran Q から自由度を引いたものを 100%表示したもので、負の値は 0%と見なすことから、I² 統計量は 0～100%の値の間を取る。25%以下では異質性が低く、25～50%は中程度、50～75%は高く、75～100%は極めて高いと評価する。これにより、統合する研究数が少なくても多くても、比較的正しく評価できる。実際のフォレストプロットでは、Cochran Q 統計量のカイニ乗検定と I² 統計量の両方が併記されていることが多い。

異質性が高い場合には、研究間のバラツキがあることを示しているなので、メタアナリシスされた結果は信頼性が低くなる。そのため、異質性が高い結果では PICO の違い（例えば、患者の重症度や薬剤の量）で分けてメタアナリシスし直したり（サブ解析）、フォレストプロットで他のものから外れた研究を排除してメタアナリシスし直したり（感度分析）して、異質性が低くなるか検討する。サブ解析や感度分析により異質性が低くなれば、その結果は信頼性が高くなる。

以上をまとめると、メタアナリシスの結果を評価するための手順としては、フォレストプロットの統合された結果（一番下にある菱形）を見て、①有意差の有無と②効果の大きさを評価し、その後、③異質性の評価のために Cochran Q のカイニ乗検定に有意差があるかどうか、また I² 統計量を確認し、異質性が高いと判断されるなら、④その原因は何かを、PICO の違いや研究の質のバラツキを見て検討する。⑤異質性の高いものに対してサブ解析、感度分析が行われていれば、その結果を評価する。

参考文献

- 1) Oxman AD, Cook DJ, Guyatt GH, for the Evidence-Based Medicine Working Group: Users' guides to the Medical Literature. IV: How to Use an overview? JAMA 1994;272:1367-1371.
- 2) 開原成允, 浅井泰博, 治療や予防に関する文献の使い方, JAMA 医学文献の読み方, 中山書店 2001 年, 95-109.
- 3) Sackett DL et al. Evidence-Based Medicine, How to Practice and Teach EBM. Churchill Livingstone 2000, 133-138.
- 4) Bedenoch D 他著, 斉尾武郎監訳, EBM の工具箱. 中山書店 2002 年, 33-38.
- 5) 柳川敏彦, Ian Roberts, 津谷喜一郎. META とは何かー未公表臨床試験を探す「アムネスティ」の試みー http://www.sphere.ad.jp/cont/27_3/META/report.html
- 6) Moher D, Cook DJ, Eastwood S, et. al. Improving the quality of reports of meta-analysis of randomised controlled trials: the QUOROM statement. The Lancet 1999;354:1896-900.
- 7) 丹後俊郎, メタアナリシス入門, エビデンスの統合を目指す統計手法, 朝倉書店 2002 年.
- 8) Cochrane Handbook for Systematic Reviews of Interventions 5.0.1. <http://www.cochrane-handbook.org/>.
- 9) Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, and others. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996;17(1):1-12.
- 10) 森實敏夫, わかりやすい医学統計学, メタアナリシス (4), 統計学的手法 (1), Medical Tribune 2000 年, <http://www.medical-tribune.co.jp/mtbackno3/3305/05hp/M3305281.htm>.

改訂履歴

- 1.0→1.1
- ・書き込み用 CAT sheet を作成
- 1.1→2.0
- ・用語を統一
 - ・結果の評価の項の説明を変更
- 2.0→2.1 (2004.3.21)
- ・書き込み用 CAT sheet の改変 (1 ページ化)
 - ・論文の PECO を探るの項を改変
- 2.1→3.0 (2006.3.10)
- ・全体の構成を大幅に改訂
 - ・Narrative review との違いについて言及
 - ・Jadad score についての解説を追加
 - ・統計学的手法について言及
 - ・ファンネル・プロットの記載場所を変更
 - ・リスク比とオッズ比についての解説を追加
 - ・フォレストプロットの解説を図示
- 3.0→3.1 (2006.7.23)
- ・メタアナリシスの説明を充実化
 - ・細かいデザインの変更
- 3.1→3.2 (2006.9.23)
- ・複数の評価者による評価の項目に選択肢を追加
 - ・heterogeniety の説明を改変
- 3.2→3.3 (2008.1.28)
- ・複数報告されている研究の有無についてのチェック項目を新設
 - ・集められた研究の結果は統合されたか (Meta-analysis) ? の項の記載を訂正
 - ・ブロボグラム brobogram の用語を追加
- 3.3→3.4 (2009.2.20)
- ・一部の項目で選択肢の順序を変更
 - ・研究が網羅的に集められたかの判定の参考になる文章を追加
 - ・集められた研究の妥当性評価と Jadad score の解説を改正
 - ・異質性の検定の評価方法についての解説を追加
- 3.4→3.5 (2010.4.24)
- ・ファンネル・プロットの図の評価の仕方についての解説を追加
 - ・集められた研究の妥当性評価の説明を修正
 - ・異質性の評価の項目の並びを変更
 - ・異質性の項に Cochran Q 統計量と I^2 検定の記述を追加
 - ・研究の結果の統合の項に, 統合してはいけない場合の記述を追加
 - ・forest plot の解説に用いる例を差し替え
- 3.5→3.6 (2010.10.3)
- ・誤字訂正
- 3.6→4.0 (2010.11.22)
- ・PECO を PICO に変更
 - ・「システムティック・レビュー」を「システムティックレビュー」に, 「メタ・アナリシス」を「メタアナリシス」に, 「ファンネル・プロット」を「ファンネルプロット」それぞれ変更
 - ・全ての研究を網羅的に集めようと努力したか? の項に, どのような種類の研究を調べたか? を新たに追加. また, 検索に用いた文献データベースに Google scholar を追加
- 4.0→5.0 (2012.3.20)
- ・どのような種類の研究を調べたか? の項の選択肢を修正
 - ・結果は統合されたか (Meta-analysis) ? の項に, 最終的に何件の研究が残り, 採用されたか, の質問を追加
 - ・異質性の検討のうち, 結果の部分を「結果の評価」に移動.

Critically appraised topic for Systematic review

Reviewer: _____

年 月 日

authors : _____

title : _____

citation : _____

PubMed PMID : _____

1. 論文の PICO は何か？

P : _____

I : _____

C : _____

O : _____

2. 全ての研究を網羅的に集めようと努力したか？

- ①データベースは？ MEDLINE EMBASE CDSR CENTRAL Google scholar その他 (_____)
- ②検索語 (_____), 期間 (_____)
- ③どのような種類の研究を調べたか？ RCT/quasi-RCT/non-RCT/CCT SR/MA/CDSR コホート研究
症例対照研究 診断研究 その他の種類の研究 (_____)
- ④参考文献まで調べたか？ 参考文献まで調べた 参考文献は調べなかった 不明
- ⑤個々の研究者や専門家に連絡を取ったか？ 連絡を取った 連絡を取らなかった 不明
- ⑥出版されていない研究も探したか？ 探した 探さなかった 不明
- ⑦同じ研究が複数報告されていないか？ 複数報告の研究は排除されている 排除されていない 不明
- ⑧英語以外で書かれた研究も探したか？ 探した 探さなかった 不明

3. 全ての研究が網羅的に集められたか？

- 網羅的に集められている 網羅的には集められていない 不明
○ファンネルプロットを用いて出版バイアスの有無を検討している
○ファンネルプロットは用いられていない

4. 集められた研究は、複数の評価者によって評価されたか？

- 複数の評価者→ _____ 人 単独の評価者
→ 評価者間で評価のくい違いが生じた場合
○合意を形成して最終的に評価を下している ○合意を形成せず、各評価者の判断を個別に記載している
○その他 (_____)
- 不明

5. 集められた研究は、明確な基準をもって妥当性を評価されたか？

- 評価基準として、明確な基準を持って評価を行った ○Jadad score ○それ以外の基準 (_____)
→ どのような評価か？
集められた個々の介入研究の Cochrane Handbook for Systematic Reviews of Interventions 5.0.1 における評価基準
○ランダム割付けの方法 ○割付け方法の隠蔽化 concealment
○割付け内容の盲検化 (患者, 研究者, outcome 評価者, 解析者) ○脱落者, 除外者の扱い
○予め決められたアウトカムを報告しているか ○それ以外の起こりうるバイアス

明確な基準はない

6. 集められた研究の異質性 heterogeneity は検討されたか？

- 異質性は検討された → 用いられた統計手法は？ ○Cochran Q (カイ二乗検定) ○I² 統計量 ○その他
異質性は検討されていない
不明

7. 結果は統合されたか (Meta-analysis)？

- ①最終的に何件の研究が残り、採用されたか？
RCT () 件, quasi-RCT () 件, non-RCT () 件, SR/MA/CDSR () 件, コホート研究 () 件,
症例対照研究 () 件, 診断 (横断) 研究 () 件, その他の種類の研究 () 件
- ②集められた研究の結果は統合されたか？
統合されている → 統合する際に用いたモデル？ ○Fixed effect model ○Random effects model ○その他
統合されていない

8. 結果の評価 (PICO 毎に評価)

- ①統合された結果に有意差があるか？ ②統合された結果の大きさは (点推定値と信頼区間)？
- ③集められた研究に異質性 heterogeneity はあるか？ 異質性は ○検討されている ○検討されていない
異質性 heterogeneity の検定 ○Cochran Q (カイ二乗検定) 有意差なし→同質 有意差あり→異質
○I² 統計量 0~25% 25~50% 50~75% 75~100%
- 同質である (有意差なし, I² 統計量 < 50%) 異質である (有意差あり, I² 統計量 ≥ 50%) 不明