

- 第1回診療ガイドライン作成のシステムティックレビューセミナー：キックオフミーティング

## ネットワークメタアナリシス

- 2022.5.14
- 公益財団法人 日本医療機能評価機構
- 客員研究主幹
- 森實敏夫

診療ガイドライン作成のためのシステムティックレビューを行う際に、ネットワークメタアナリシスが必要かどうかの判断が出来るようになることを目的に、基本的な事項を解説します。

学習者用Action Mapはこちら  
<https://info.zanet.biz/lec/srsz/0-nma-main/0-nma-stu-am.pdf>

1

## 学習ゴールと学習アウトカム

診療ガイドライン作成のためにシステムティックレビューを行う際、ネットワークメタアナリシス（NMA）の必要性を判断できるようにする。

1. NMAの基礎を説明できる。
2. NMAの結果得られるアウトプットを説明できる。
3. NMAのエビデンスの確実性評価法（GRADEアプローチ）を説明できる。

ネットワークメタアナリシスの基礎について解説し、解析の結果えられる情報すなわちアウトプットについて解説し、エビデンスの確実性の評価のGRADEアプローチについて解説します。

2

## ネットワークメタアナリシスの基礎

まずネットワークメタアナリシスの基礎です。

3

通常のペア比較メタアナリシスとネットワークメタアナリシスをシステムティックレビューの段階で比べてみると、ペア比較メタアナリシスでは、二つの介入のうちどちらの方が効果が優れているかを解析することが目的であるのに対し、ネットワークメタアナリシスでは、介入が三つ以上ある場合で、最も効果が優れているのはどれかを解析することが目的になります。

臨床決断の場面では、ペア比較メタアナリシスは二つの介入のうち、どちらを選択すべきかという問題に解答を与えてくれますが、ネットワークメタアナリシスは三つ以上の介入のうちどれを選択すべきか、言い換えると最善の介入はどれかという問

# ペア比較メタアナリシスとネットワークメタアナリシス

	ペア比較メタアナリシス	ネットワークメタアナリシス
システマティックレビュー	二つの介入の内どちらの方が効果が優れているか? Comparative effectiveness	三つ以上の介入の内、最も効果が優れているのはどれか? Extended comparative effectiveness
臨床決断	二つの介入の内どちらを選択すべきか?	三つ以上の介入の内、どれを選択すべきか? 最善の介入はどれか?
診療ガイドライン	二つの介入の内どちらを推奨すべきか?	★ 三つ以上の介入の内、どれを推奨すべきか?

ネットワーク全体が意味を持つと考える

最善の介入はどれか?      選択肢が二つしかない。      選択肢が三つ以上ある。

題に解答を与えてくれる可能性があります。いずれの場合も、それぞれのアウトカムについて、効果の大きさと、確実性に関する情報を与えてくれますが、益と害のバランス、あるいは、正味の益については、さらに複数のアウトカムに渡る解析が必要になります。

診療ガイドラインにおいては、ペア比較メタアナリシスは二つの介入のうちどちらを推奨すべきかを定めるために必要になるのに対し、ネットワークメタアナリシスは三つ以上の介入のうちどれを推奨すべきかを定めるために有用だと言えます。

また、理論的には、同じペア比較であっても、ネットワークメタアナリシスは直接比較の情報に加え、間接的な比較の情報も活用することで、より確実性の高いエビデンスが得られることが利点と考えられています。

## ネットワークメタアナリシスの前提

用語	解説	評価方法
★ Transitivity 移行性	介入の相対的効果に影響し得る要素について介入と対象者が、含める研究間で類似していること。	研究選択の段階で、 <b>研究手法、臨床的事項、介入、アウトカム</b> が類似していることを確認。効果を修飾し得る因子が類似していることを確認する。研究数が多ければ（少なくとも5件）統計学的にそれらの因子の分布を検討することも可能。
Consistency assumption 一貫性の前提	ネットワークの中のペア比較で直接のエビデンスと間接のエビデンスが統計学的に一致していること。（移行性が保たれていることが前提。一貫性の前提が確認されても移行性が保たれている証明にはならない。）	z-test, $\chi^2$ testなどが用いられるが、検出力が低い。
<b>Coherence コヒレンス</b>		$d_{XY}^{Indir} = d_{ZY}^{Dir} - d_{ZX}^{Dir}$

Chaimani A, et al: Common pitfalls and mistakes in the set-up, analysis and interpretation of results in network meta-analysis; what clinicians should look for in a published article. Evid Based Ment Health 2017;20:88-94. PMID: 28739577

ネットワークメタアナリシスを使う場合には、Transitivity 移行性の概念を理解しておく必要があります。移行性は、介入の相対的効果に影響し得る要素について 介入と対象者が、含める研究間で類似していることを表します。移行性が保たれていることはネットワークメタアナリシスの前提になります。

研究選択の段階で、研究手法、臨床的事項、介入、アウトカムが類似していることを確認します。特に、効果を修飾し得る因子が類似していることを確認します。研究数が多ければ（少なくとも5件）統計学的にそれらの因子の分布を検討することも

可能です。

さらにコヒレンスが保たれていることもネットワークメタアナリシスの前提になります。コヒレンスというのは、移行性が保たれていることを前提として、ネットワークの中のペア比較で、直接比較と間接比較の結果が一致していることを言います。ネットワークメタアナリシスのエビデンスの確実性の評価には コヒレンスが保たれていないこと、すなわちインコヒレンスの評価が必要になります。インコヒレンスの評価は通常の ペア比較のメタアナリシスで行なわれることはありません。

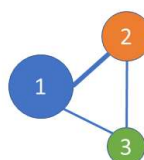
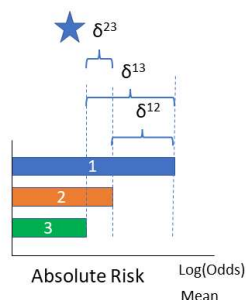
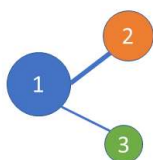
文献 <https://ebmh.bmj.com/content/ebmental/20/3/88.full.pdf>

ネットワークメタアナリシスの原理を示します。例えば、左側の図に示すように、治療1と2、治療1と3を直接比較したランダム化比較試験があったとします。その場合、治療2と3を間接的に比較することが可能になります。中央のグラフで示すように、治療1と2の絶対リスク、例えばイベント率の差は、 $\delta_{12}$ で表すことができます。そして治療1と3の間の絶対リスクの差は $\delta_{13}$ で表すことができます。中央の図に示すように、 $\delta_{13}$ から $\delta_{12}$ を引き算すると $\delta_{23}$ の値が得られます。すなわち直接比較のデータから、間接比較のデータを推定することが可能になります。この間接比較のデータの分散は、治療1、2、治療1、3の直接比較の分散の合計した値になります。

右側のネットワークの場合には、治療2と治療3の直接比較のデータもあるので、治療1 3 治療1 2から間接的に得られた、治療2 3の比較のデータと直接比較のデータを、それぞれの分散の逆数で重みづけして平均値を統合値として求めます。この統合値のことを、ネットワーク推定値と呼んでいます。

ここでいう絶対リスクの差は、例えば、それぞれの治療群のオッズの自然対数の差であれば、オッズ比の自然対数になります。リスク比、ハザード比についてもそれぞれの治療群のイベント率、あるいは、ハザード率の自然対数の差です。これらの

# ネットワークメタアナリシス



治療1と2の差 $\delta^{12}$ 、1と3の差 $\delta^{13}$ 、2と3の差 $\delta^{23}$ の関係は、 $\delta^{12} + \delta^{23} = \delta^{13}$ が成立し、移項すると $\delta^{23} = \delta^{13} - \delta^{12}$ となり、治療2と3を間接比較した場合の効果推定値が得られる。分散 $V$ については、 $V^{23} = V^{12} + V^{13}$ と直接比較の分散の合計になる。

直接比較の $\delta_d^{23}$ と間接比較の $\delta_i^{23}$ をそれぞれの分散の逆数で重みづけした平均値が統合値 (NMA estimate) となる。直接比較の分散 $V_d^{23}$ 、間接比較の分散 $V_i^{23}$ となり、前者は2と3の直接比較のデータから、後者は1と2、1と3の直接比較の研究データから算出された分散から算出される (多変量正規分布の変数の加算)。<sup>6</sup>

差のエクスポネンシャルを計算すると、それぞれオッズ比、リスク比、ハザード比が得られます。

## エビデンスベースの構成：ランダム化比較試験を偏りなく収集

1. 比較すべきすべての治療 (介入) をリストアップする。
2. クリニカルクエスションの対象者 (P) で設定したアウトカム (O) についてこれらの治療の **2つまたは3つ以上を比較**した研究を探す。
3. 3つ以上の治療を比較した研究から対象外の治療アームは除く。これでつながったネットワークができたらそこでストップする。
4. もしこれにつながったネットワークでない場合：
  - a. ふたつのサブネットワークに含まれる (ステップ3で除いた治療も含め) いずれかの治療と比較された治療をさらに探す。
  - b. これらを比較治療に追加する。
  - c. ステップ2と3をより大きくなった比較治療のセットについて実行する。

ネットワークメタアナリシスの場合も、クリニカルクエスションに基づいて、包括的な文献検索を行いません。比較する治療あるいは介入は3つ以上になるので、それぞれの研究は二つまたは三つ以上の介入の効果と比較した研究になります。すべての研究が、同じ疾患あるいは病態の対象者で同じアウトカムを測定しているものである、必要があります。

まず最初に、比較すべき全ての治療をリストアップします。そして、クリニカルクエスションで設定された対象者およびアウトカムが共通のランダム化比較試験で、その治療の効果解析した研究を検索することに

なります。もし見つかった研究が、三つ以上の治療を比較した研究であって、そのうちのひとつあるいは二つ以上が目的の治療でない場合は、その介入は、解析データから除いて、その研究を解析に含めます。その上で、繋がったネットワークが出来た場合には、そこで検索を終了します。

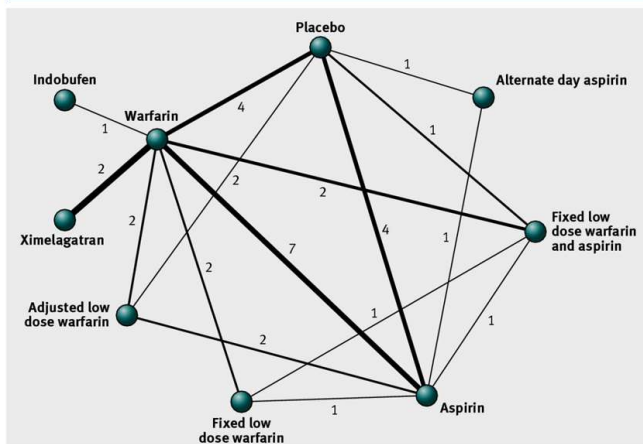
それぞれの治療を、ORで結合した検索式を用いることにはなりますが、検索結果で得られた研究で、繋がったネットワークができない場合もあり得ます。もし、最初にリストアップした治療では、ネットワークができない場合には、さらに別の治療を追加して、文献検索を進め、より拡大した治療比較のセットについて、同様に文献検索を行い、つながったネットワークができて来る場合もあります。そのように検索範囲を拡大するかどうかは、臨床的な観点からの検討も必要になり、ペア比較のメタアナリシスだけを前提とする場合と、検索戦略が異なってきます。

ここでは当該クリニカルクエスションに対応する研究の集合をエビデンスベースとよんでいます。

この図はネットワークグラフあるいはネットワークジオメトリーなどと呼ばれ、各ノードが治療法、それらを結んだ実線は直接比較、線の太さは研究の数を表しており、この例では、その線上に研究数が示されています。結合線が少ない場合には、直接比較が少ないことを表し、ネットワーク全体としては、ネットワークメタアナリシスの結果の信頼性が低いとみなされます。ネットワークグラフについて、それを解説する、定性的な記述をつけることが望ましいとされています。

### Quiz

## ネットワークグラフ (ネットワークジオメトリ)



介入比較の構造を示す。比較される介入、それぞれの介入にランダム割り付けられた症例数、特定の介入や比較が選択されたあるいは忌避されたかを提示する。ネットワークはさまざまな形をとり、結合線が少ない場合は、直接比較が少ないことを表し、そのようなネットワークのメタアナリシスは信頼性が低いとみなされる。ネットワークジオメトリの定性的な記述が必要とされ、ネットワークグラフに付随されるべきである。確立したネットワークジオメトリの見方はまだない。

Hutton B, et al: The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med 2015;162:777-84. PMID:26030634

## ネットワークメタアナリシスの結果をまとめる

それぞれの研究から、メタアナリシスに必要なデータすなわち症例数とイベント数、あるいは症例数と平均値、標準偏差、ハザード比と標準誤差などのデータを抽出し、それらを解析することで、ネットワークメタアナリシスのさまざまな結果が、アウトプットとして得られます。

## ネットワークメタアナリシスのアウトプット

Cochrane Handbook v.6.0

- **ネットワークグラフ**
- ネットワーク構造の表による提示
- 直接比較の貢献度の表による提示：各ペア比較とネットワーク全体（研究数含む）
- **Forest plot**
- **Rankograms**：各介入がそれぞれの順位になる確率をグラフ化（積み重ね棒グラフ、折れ線グラフ）
- 異質性HeterogeneityとインコヒレンスIncoherence： $\tau^2$ 値、表
- **Surface Under the Cumulative Ranking Curve SUCRA**：各介入の順位確率を累積確率として折れ線グラフ化して表示する。折れ線下の面積が1の場合は最善、0の場合は最悪の介入となる。

Cochrane Handbook for Systematic Reviews of Interventions version 6.3によると、ネットワークメタアナリシスのアウトプットとして、ここにリストアップした情報が得られます。これらのうち、ネットワークグラフ、フォレストプロット、ランコグラム、SUCRA (Surface Under the Cumulative Ranking Curve)が重要です。

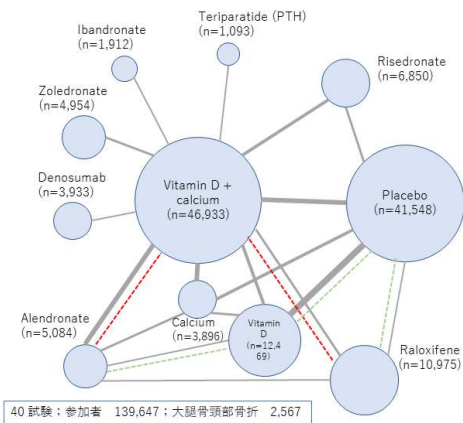
また、ここには含まれていませんが、直接比較と間接比較の推定値と95%確信区間あるいは信頼区間、および、ネットワーク推定値とその95%確信区間あるいは信頼区間、さら

に直接比較と間接比較の有意差検定の結果を表形式あるいはグラフ表示したデータはインコヒレンスの評価に有用です。

なお、ベイジアンアプローチを用いた場合は、95%確信区間 Credible Interval、頻度論派アプローチを用いた場合は、95%信頼区間Confidence Intervalという用語が用いられます。

これは、骨粗鬆症の患者さんが対象で、大腿骨頸部骨折をアウトカムとした、11種類の治療法のネットワークグラフの例です。結合線は直接比較 (head-to-head comparison)を示しますが、間接比較される二つの治療の間に別の治療が一つ介在す

# ネットワークグラフとループ



★ 一次ループ First order loops : -----  
比較される二つの治療の間に別の治療が二つ介在する。

① vs ② ; ② vs ③ → ①と③の間接比較

二次ループ Second order loops : -----  
比較される二つの治療の間に別の治療が二つ介在する。

① vs ② ; ② vs ③ ; ③ vs ④ → ①と④の間接比較

**GRADEアプローチでは一次ループだけを対象とする Loop-specific approach**

Puhan MA, et al: A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ 2014;349:g5630. PMID: 25252733

11

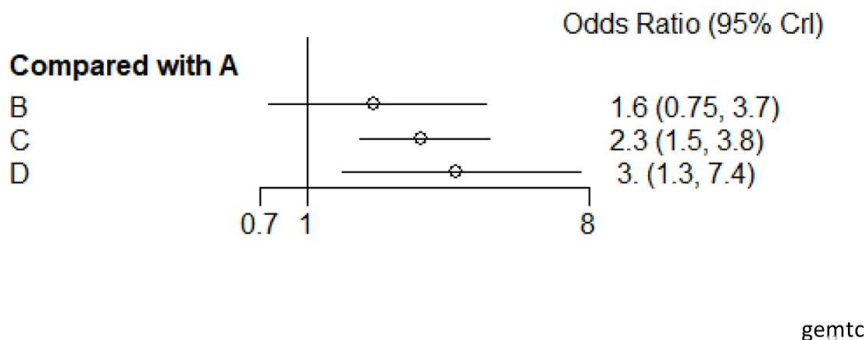
る場合は一次ループと呼び、間接比較される二つの治療の間に別の治療が二つ介在する場合は二次ループと呼びます。赤の点線は一次ループの例で、緑の点線が二次ループの例です。

GRADEアプローチでは、原則として、一次ループだけをエビデンス評価の対象とすることになっています。すなわち、エビデンスの確実性の評価の対象は、一次ループまでとなっています。エビデンス評価の対象になる場合は、間接比較のみの場合と直接比較と間接比較の両方ある場合とがあります。後者は、直接比較と間接比較の間の不一致、すなわ

ちインコヒレンスの程度の評価が必要になります。

結合線の太さは研究数を表します。また、このネットワークグラフの例ではノードの大きさが症例数を表しています。

## Forest plot



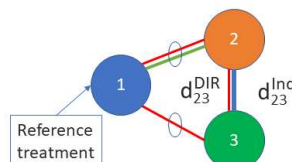
これは、ABCD、四つの治療を比較したネットワークメタアナリシスのフォレストプロットの例です。治療Aを対照として、BCDに対するオッズ比と95%確信区間をグラフ表示しています。

通常のペア比較のメタアナリシスの場合は、点推定値と95%確信区間あるいは信頼区間のバーは、一つの研究の効果指標の値に対応していますが、ネットワークメタアナリシスの場合は、複数の研究の統合値とその確信区間あるいは信頼区間を表しています。統合値は間接比較の場合の値と、直接比較の場合と、それらを統合した“ネットワーク推定値”（あ

るいはネットワークメタアナリシス推定値）です。

## Node-splitting method 結節分割法

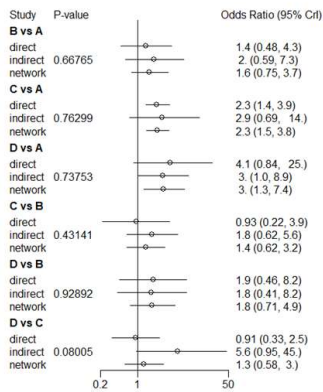
- 直接比較と間接比較のデータが得られるすべてのコントラスト（ペア）について  $d_{23}^{Dir}$  と  $d_{23}^{Ind}$  を算出する。
- $d_{23}^{Ind}$  は一次ループだけでなくネットワーク全体から算出する。
- Rのパッケージgemtcは自動で実行し、結果のグラフ表示、検定結果も得られる。



直接比較と間接比較の両方の推定値が得られる場合、例えば、この図で治療2と治療3の比較のような場合、両方を統合したネットワーク推定値 network estimate だけでなく、直接比較の推定値と間接比較の推定値、全部で3種類の推定値を算出することができます。それを自動的に、直接比較と間接比較のデータが得られるすべてのペアについて算出する方法が結節分割法 Node-splitting method です。

Rのパッケージgemtcでは、この方法を実行し、結果のグラフ表示、直接比較と間接比較の有意差検定の結果も得られるので、それをインコヒレ

ンスの評価に用いることができます。



## Node-splitting analysisの結果のプロット

これはNode-splitting analysisの結果をグラフ表示したものです。それぞれのペア比較に対して、直接比較、間接比較それぞれの統合値、ネットワーク推定値とそれらの95%確信区間が示されています。

さらに、直接比較と間接比較のそれぞれの統合値の有意差検定のP値が示されています。

14

## 順位確率→効果が最大の治療を明らかにする

```
> rank.p=rank.probability(res,preferredDirection=1,covariate=NA)
> print(rank.p)
Rank probability; preferred direction = 1
      [,1]  [,2]  [,3]  [,4]
治療 A 0.000030 0.002265 0.105390 0.892315
治療 B 0.058115 0.178120 0.660630 0.103135
治療 C 0.233275 0.596800 0.169735 0.000190
治療 D 0.708580 0.222815 0.064245 0.004360
> dev.new();plot(rank.p)
```

治療	1位の確率	2位の確率	3位の確率	4位の確率
A	0.000030	0.002265	0.105390	0.892315
B	0.058115	0.178120	0.660630	0.103135
C	0.233275	0.596800	0.169735	0.000190
D	0.708580	0.222815	0.064245	0.004360

- Rankogram
- SUCRA
- 順位中央値と95%CI

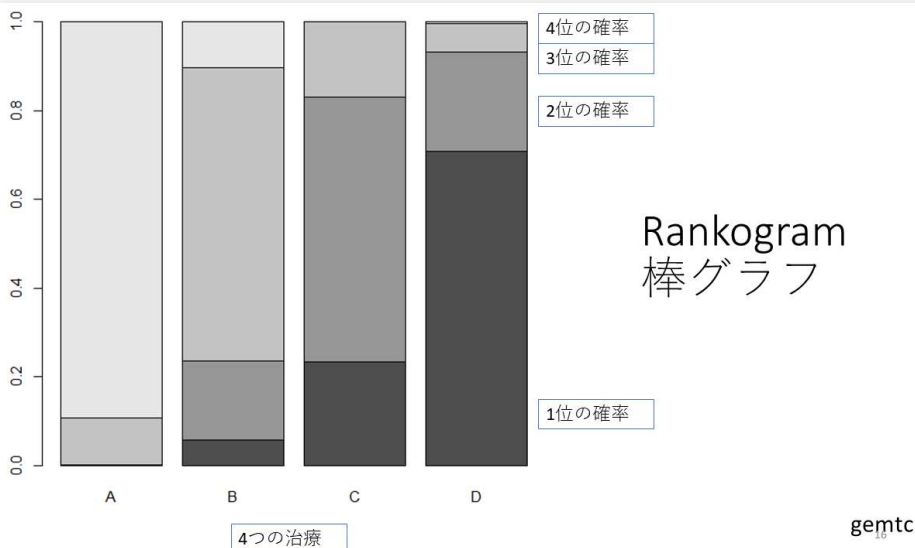
列および行の合計は1.0

gemtc

ネットワークメタアナリシスでは、それぞれの治療がそれぞれの順位になる確率が計算されます。例えば、ここに示す例では、治療Dの効果の大きさが1位になる確率0.7、水平方向に見て、2位になる確率が0.22、3位になる確率が0.06、4位になる確率が0.004です。

また、垂直方向に見ると、例えば、1位になる確率は高い順に、治療D、C、B、Aであることが分かります。

これら順位確率の値をグラフ化したものとして、ランコグラムとSUCRAがあります。



これはランコグラムですが、それぞれの治療がそれぞれの順位になる確率を積み上げ棒グラフの形で表したものです。1位の確率は黒で表されており、1位になる確率が最も高いのは治療Dであることがわかります。一方、治療Aは、4位になる確率が最も高いことがわかります。なお、ランコグラムは、このような積み重ね棒グラフだけでなく、折れ線グラフとすることもできます。

また、1位か2位のどちらかになる確率を見てみると、治療CとDは大きな差が無いことがわかります。もし、1位の治療Dと2位の治療Cの効果の大きさがわずかな場合は、1位の確率が

一番高い、治療Dを用いても、2位の確率が一番高い治療Cを用いても、アウトカムに大きな差がないことが推定されます。

これは4つの治療を比較した別の例ですが、順位確率の値から、順位の中央値と95%信頼区間（または確信区間）を算出しています。

# Rank and 95% CI

順位 b	治療j			
	1	2	3	4
1	0.25	0.50	0.25	0
2	0.25	0.25	0.50	0
3	0.25	0.25	0.25	0.25
4	0.25	0	0	0.75
順位中央値(95%CI)	2.5 (1,4)	1.5 (1,3)	2 (1,3)	4 (3,4)

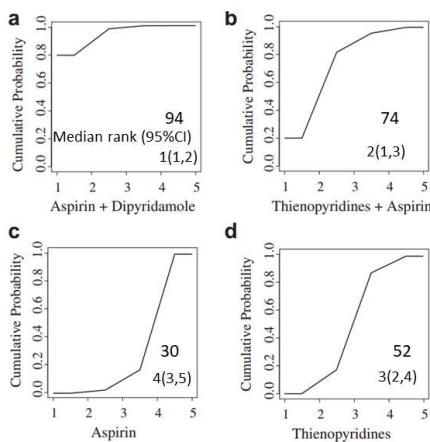
1を1番目から25番目まで25個並べる。続けて、2を26番目から75番目まで50個並べる、続けて3を76番目から100番目まで25個並べる。

50番目は2、51番目も2なので中央値は2。

2番目も3番目も1なので、95%CIの下限値は1、97番目も98番目も3なので、95%CIの上限値は3。

Rでなら：  
`rnkp=c(0.25,0.25,0.25,0.25)`  
`rnknum=trunc(A*1000)`  
`rnk=c(rep(1,rnknum[1]),rep(2,rnknum[2]),rep(3,rnknum[3]),rep(4,rnknum[4]))`  
`median(rnk)`  
`quantile(rnk,c(0.025,0.975))`

```
> median(rnk)
[1] 4
> quantile(rnk,c(0.025,0.975))
 2.5% 97.5%
    3     4
```



SUCRA  
Surface Under  
the  
Cumulative  
Ranking  
Curves

Salanti G, et al: Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol 2011;64:163-71. PMID: 20688472

SUCRAも、ネットワークメタアナリシスの結果について治療効果の順位を表す一つの指標です。累積順位曲線下面積という意味になります。それぞれの治療について累積順位確率を縦軸に、順位を横軸にして描かれる曲線で、治療の効果の大きさの順位付けに、曲線下の面積を用います。

SUCRAの値は、最下位の順位をKとした場合、(K-順位の数)を順位確率で重みづけして、合計値を求め、最大値が100、最小値が0になるように標準化した値に相当します。つまり、1位の値に最大値を設定し、等間隔でそれより小さな値を設定し、そ

れらを順位確率で重みづけして、合計値を求め、最大値が100、最小値が0になるように標準化した値であれば、どのような最大値を設定しても同じ結果が得られます。

順位の数値は絶対効果の数値そのものではなく、その順位です。したがって、SUCRAの数値も順位判断に用いることは問題ありませんが、1位と2位の絶対効果の大きさの差と、2位と3位の絶対効果の差は同じとは限りません。また、1位の治療と2位の治療の絶対効果の差がわずかしかない場合もあり得ます。

さらに、一つのアウトカムに対する効果の大きさの順位なので、他のアウトカム、特に害のアウトカムに対する効果の大きさを考慮する場合は、絶対効果で評価しないと誤った結論を出す可能性があります。

ネットワークメタアナリシスの結果に基づく、SoFテーブル(Summary of Findings Table)には2種類のフォーマットが提案されています。これはその一つですが一行に一つのペア比較の結果を記述するように構成されています。比較する治療の数が多い場合にはこの形式のほうがまとめやすいでしょう。

各行の一番左側に、ネットワークグラフの対応するノードの色などを提示すると、ネットワークグラフとの関係がわかりやすくなります。

各ペアの、介入なし、介入ありの欄に、それぞれペア比較される治療の名称を記入します。

クリニカルクエスチョンを記述する欄に、ネットワークグラフ(ネットワークメタアナリシスGeometry plot)を提示するようになっています。また、順位と95%確信区間、SUCRAの値を記入するようになっています。

効果推定値はオッズ比などの相対効果の指標と各群の絶対リスクとその群間の差、すなわち絶対効果の大きさを記入するようになっています。

通常のSoFテーブルには無い項目として、順位と95%確信区間およびSUCRA値があります。

NMA-SoF table-1

益/害

患者/参加者(P):

介入(I):

対照(参照)(C): ネットワークメタアナリシス Geometry plot\*

アウトカム(O):

セッティング(S):

総研究数 参加者数	相対効果** (95% CrI)	予測される絶対効果*** (95% CrI)			エビデンスの 確実性	順位**** (95% CrI)	結果の解釈
		介入なし	介入あり	差			
		治療1	治療2				
		治療1	治療3				
		治療2	治療3				

\* 実線は直接比較

\*\* 推定値はオッズ比, CrIは確信区間(ベイズアプローチによる).

\*\*\* 予測される絶対効果: 介入群と対照群の絶対リスクの差.

\*\*\*\* Surface under the cumulative (SUCRA) rankingと効果に対する確信区間を提示. 順位の統計値はその治療が1位, 2位, ..., n位になる確率である.

GRADEによるエビデンスの確実性

高い質: 真の効果は効果推定値に近いことに非常に確信がある.

中等度の質: 効果推定値に中等度の確信がある: 真の効果は効果推定値に近い可能性が高いがかなり異なる可能性がある.

低い質: 効果推定値に対する確信は限られている: 真の効果は効果推定値とかなり異なるかもしれない.

非常に低い質: 効果推定値にほとんど確信が持てない: 真の効果はかなり異なる可能性がある.

解説脚注:

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

19

SoF table format-1

絶対効果の大きさは、ベースラインリスクの値を疾患レジストリなどから得て、オッズ比、リスク比、ハザード比から計算することも可能ですし、ベースラインリスクに、高・中・低などあり得る推定値を設定して、計算する場合があります。

NMA-SoF table-2

益/害

患者/参加者(P):

介入(I):

対照(参照)(C): ネットワークメタアナリシスGeometry plot\*

アウトカム(O):

セッティング(S):

アウトカム	効果推定値と確信区間	治療1	治療2	治療3	コメント	
アウトカム1						
参照治療 (%)***	ORなど相対効果 (95%CrI) ネットワーク推定値	絶対効果 (95%CrI)	ORなど相対効果 (95%CrI) ネットワーク推定値	絶対効果 (95%CrI)	ORなど相対効果 (95%CrI) ネットワーク推定値	絶対効果 (95%CrI)
順位(95%CrI)	効果推定値に対する確信度	順位(95%CrI)	効果推定値に対する確信度	順位(95%CrI)	効果推定値に対する確信度	順位(95%CrI)
参加者総数(研究数)		参加者総数(研究数)		参加者総数(研究数)		参加者総数(研究数)
アウトカム2						
参照治療 (%)***	ORなど相対効果 (95%CrI) ネットワーク推定値	絶対効果 (95%CrI)	ORなど相対効果 (95%CrI) ネットワーク推定値	絶対効果 (95%CrI)	ORなど相対効果 (95%CrI) ネットワーク推定値	絶対効果 (95%CrI)
順位(95%CrI)	効果推定値に対する確信度	順位(95%CrI)	効果推定値に対する確信度	順位(95%CrI)	効果推定値に対する確信度	順位(95%CrI)
参加者総数(研究数)		参加者総数(研究数)		参加者総数(研究数)		参加者総数(研究数)

NMA-SoF表の定義

\* 実線は直接比較

\*\* 推定値はオッズ比, CrIは確信区間(ベイズアプローチによる).

\*\*\* 想定されるリスク(例えば、研究横断的な対照群の中央値)は脚注に記載.

CrI: Credible Interval, OR: Odds Ratio.

GRADEによるエビデンスのグレーディング

高い質: 将来の研究が効果推定値に対する我々の確信度を要する可能性は非常に低い.

中等度の質: 将来の研究が効果推定値に対する我々の確信度へ重要な影響を持つ可能性があり、推定値を変えるかもしれない.

低い質: 将来の研究が我々の効果推定値に対する確信度へ重要な影響を持つ可能性がかなりあり、推定値を変える可能性がある.

非常に低い質: 推定値に対して我々は非常に不確かである.

解説脚注:

1 ベースラインリスク(想定される対照群のリスク)はNational Cancer Instituteの集約プログラムから得た.

2 95%確信区間が同等の値と交差するので非常に重要な不確実性があり、広い確信区間から害の可能性が高いことを示す.

3 概念的にはありうる修飾因子分布は異なる化学療法を用いる研究間で差がないので、明らかな非移行性はない.

4 順位は中央値(順位1~10)と95%確信区間を示す.

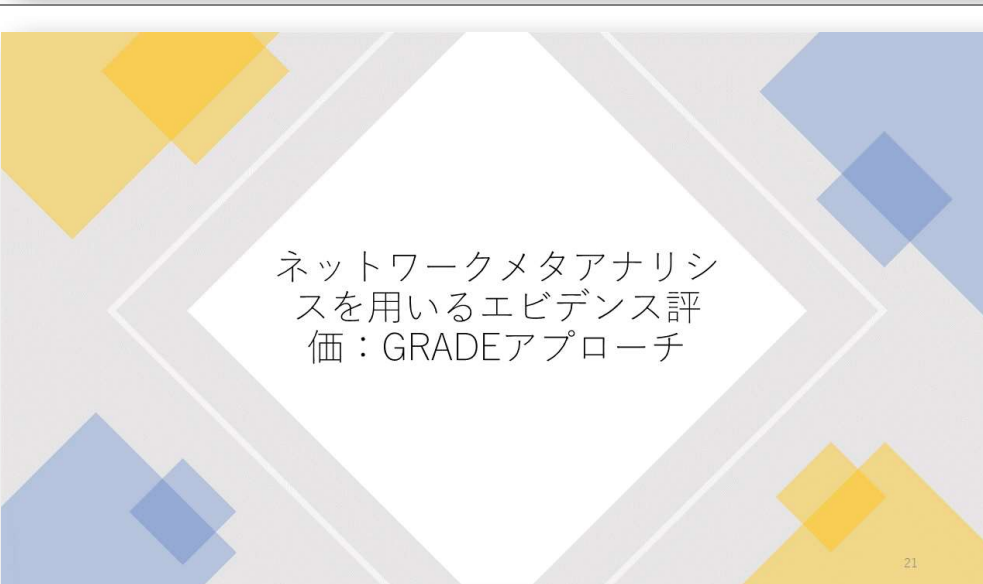
20

SoF table format-2

SoFテーブルの、もう一つのフォーマットがこちらです。このフォーマットは比較する治療の数が少ない場合に使うことができます。一つの行に比較される治療のそれぞれの効果推定値を記入するようになっています。複数のアウトカムに対する効果を一覧しやすいことが利点です。

以上2つのフォーマットのいずれも、ネットワーク推定値を記入するようになっていますが、間接比較、直接比較だけの場合は、その旨を記述する必要がありますでしょう。

Quiz



ネットワークメタアナリシスのエビデンス評価のGRADEアプローチについて解説します。

ランダム化比較試験のバイアスリスクの評価には 選択バイアス、実行バイアス、検出バイアス、症例減少バイアス、その他のバイアスの5つのドメインの分類に基づいて、評価が行われてきました。2019年の8月に、コクランが risk of bias tool version 2.0を発表し、現在はそちらへの移行が行われています。RoB 2 では、シグナリングクエスチョンに答えることによって、バイアスリスクをHigh、Some Concerns、Lowの3段階で評価するようになっています。

RoB 2 tool A revised Cochrane risk of bias tool for randomized trials (August 2019)  
<https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool>



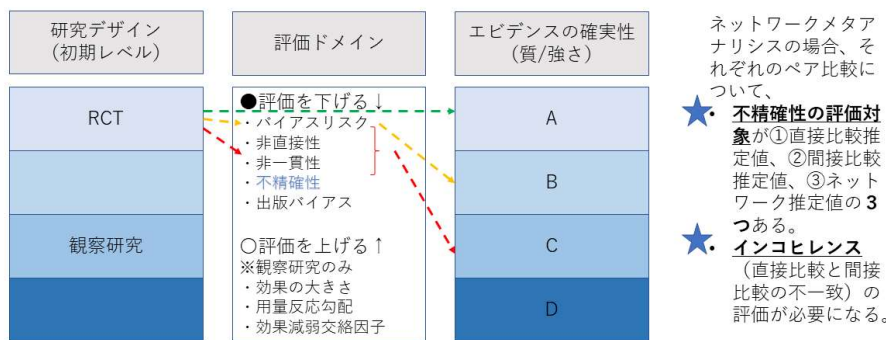
# 個別研究のバイアスリスクの ドメインと項目: RCT

コクランRoB2に基づくバイアスリスク評価のためのウェブサイト

ドメイン	評価項目	Cochrane RoB v.2.0
選択バイアス	①ランダム化： ランダム配列の生成、ベースラインの不均衡	ランダム化の過程から生じるバイアス (AHRQ) ・ランダム化の過程あるいは交絡で生じるバイアス ・研究への参加者の選択におけるバイアス
	②割り付けの隠蔽 (コンシールメント)	
実行バイアス	③参加者と医療提供者の盲検化	治療企図からの乖離によるバイアス
検出バイアス	④アウトカム測定者の盲検化	アウトカム測定におけるバイアス
症例減少バイアス	⑤ITT解析 ⑥不完全アウトカムデータ	アウトカムデータの欠損のため生じるバイアス
その他のバイアス	⑦選択的アウトカム報告 ⑧早期試験中止 ⑨その他のバイアスの可能性：交差試験のキャリーオーバー、COIによるバイアス、その他	報告結果の選択におけるバイアス

[https://stat.zanet.biz/sr/crobt2\\_j.htm](https://stat.zanet.biz/sr/crobt2_j.htm)

## エビデンス総体のエビデンスの確実性 (質/強さ) 評価 枠組み (アウトカム毎)



ランダム化比較試験の場合、複数の研究をまとめたエビデンス総体の評価ドメインは、バイアスリスク、非直接性、非一貫性、不精確性、出版バイアスの5ドメインですが、ネットワークメタアナリシスの場合は、それぞれのペア比較について不精確性の評価対象が、直接比較推定値、間接比較推定値、ネットワーク推定値の3つになります。そして直接比較と間接比較の不一致、すなわちインコヒレンスの評価が必要になってきます。

## 4-ステップGRADEアプローチ

Puhan MA, et al: A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ 2014;349:g5630. PMID: 25252733

1. 直接比較と間接比較の効果推定値をそれぞれの比較について提示する。
  - ・直接比較の効果推定値はHead-to-headの比較 (A vs B)のRCTから、間接比較の効果推定値は共通の対照をもつ2つ以上のHead-to-headのRCTから (例: A vs C, B vs CからA vs B) 得る。
2. 直接比較と間接比較の効果推定値の質をレーティングする。
3. エビデンスネットワークのそれぞれの比較についてネットワークメタアナリシス効果推定値 (NMA estimate) を提示する。
4. それぞれのネットワークメタアナリシス効果推定値の質をレーティングする。

ネットワークメタアナリシスに対するGRADEアプローチは4つのステップからなっています。まず、直接比較と間接比較の効果推定値をそれぞれ提示します。二番目に 直接比較と間接比較の効果推定値の質をレーティングします。三番目に エビデンスネットワークのそれぞれの比較についてネットワークメタアナリシス効果推定値 (ネットワーク推定値) を提示します。四番目にそれぞれのネットワーク推定値の質をレーティングします。

Puhan MA, et al: A GRADE Working Group approach for rating the quality of treatment

effect estimates from network meta-analysis. BMJ 2014;349:g5630. PMID: 25252733

<https://www.bmj.com/content/349/bmj.g5630>

ネットワークメタアナリシス中のペア比較のエビデンスの確実性の評価のひとつの例を示します。

# NMA中のペア比較のエビデンスの確実性評価

比較①は、直接比較のエビデンスの確実性が低、ABCDのCの評価です。コメントdの意味は、“深刻な不精確性”となっています。すなわち、オッズ比の95%確信区間が1をはさんで、しかも0.04から5.45と幅広くなっているため“深刻”と判定したと考えられます。この低の評価は、エビデンス総体の評価なので、おそらく、“深刻な不精確性”とそれ以外のドメイン、例えば、バイアスリスクが“深刻”との判断があって、Aから2段階レートダウンして、低 (C) になったのだと考えられます。

比較①の間接比較のエビデンスの確実性は中 (B) となっており、コメン

比較		直接比較のエビデンス		間接比較のエビデンス		ネットワークメタアナリシス	
介入	対照	オッズ比 (95%CI)	エビデンスの確実性	オッズ比 (95%CI)	エビデンスの確実性	オッズ比 (95%CI)	エビデンスの確実性
①	アレンドロネット	ラロキシフェン 0.49 (0.04-5.45)	⊕⊕○○ 低 <sup>d</sup>	0.53 (0.30-0.90)	⊕⊕⊕○ 中 <sup>e</sup>	0.51 (0.29-0.87)	⊕⊕⊕○ 中
②	アレンドロネット	デノスマブ -	-	0.90 (0.45-1.78)	⊕⊕○○ 低 <sup>g*</sup>	0.90 (0.45-1.78)	⊕⊕○○ 低
③	ラロキシフェン	プラセボ 0.84 (0.63-1.13)	⊕⊕⊕○ 中 <sup>c</sup>	0.96 (0.53-1.78)	⊕⊕○○ 低 <sup>g*</sup>	0.87 (0.63-1.22)	⊕⊕⊕○ 中
④	リセドロネット	プラセボ 0.17 (0.05-0.59)	⊕⊕○○ 低 <sup>ah</sup>	0.54 (0.36-0.75)	⊕⊕○○ 低 <sup>f</sup>	0.48 (0.31-0.66)	⊕⊕○○ 低
⑤	カルシウム	カルシウム+ビタミンD 1.21 (0.89-1.66)	⊕⊕○○ 低 <sup>g*</sup>	3.43 (0.26-160.40)	⊕○○○ とても低 <sup>gh</sup>	1.40 (1.03-1.95)	⊕⊕○○ 中 <sup>i</sup>
⑥	ビタミンD+カルシウム	デノスマブ 1.67 (1.02-2.70)	⊕⊕⊕○ 中 <sup>c</sup>	評価不能 <sup>g</sup>	評価不能 <sup>g</sup>	1.64 (0.97-2.87)	⊕⊕⊕○ 中 <sup>c</sup>

a. バイアスリスク; b. 非一貫性; c. 不精確性; d. 深刻な不精確性; e. その直接比較のエビデンスが中の質; f. その直接比較のエビデンスが低または非常に低の質; g. エビデンスネットワーク中でループによるつながりが無いため評価できない; h. 対象者の非直接性または非移行性; i. 不精確性が小さい

ここに注目

トeは、“その直接比較のエビデンスが中の質”となっています。ここでいう、その直接比較というのは、上述した直接比較のエビデンスのことではなく、この間接比較の元になった直接比較の意味です。それが、バイアスリスク、非直接性、非一貫性、出版バイアスのいずれか、あるいは、非移行性に深刻な問題があると判定したため、AからBへレートダウンしたと考えられます。不精確性は“深刻でない”と判定したと考えられます。

比較①のネットワークメタアナリシスのエビデンスの確実性は、中 (B) となっています。コメントは特に付けられていません。ここでは、直接比較と間接比較の効果推定値の点推定値、すなわちオッズ比は0.49と0.53で、ほぼ同じであり、95%確信区間も重なっているため、インコヒレンスは“深刻でない”と判定され、ネットワーク推定値の95%信頼区間が0.29~0.87で1を十分下回っているため、不精確性も深刻でないと判定され、間接比較の貢献度が直接比較の貢献度より大きい（信頼区間の幅から間接比較の重みが大きいと考えられた）ので、間接比較のエビデンスの確実性をネットワーク推定値のエビデンスの確実性に設定したと考えられます。

比較②は直接比較のエビデンスがない場合です。

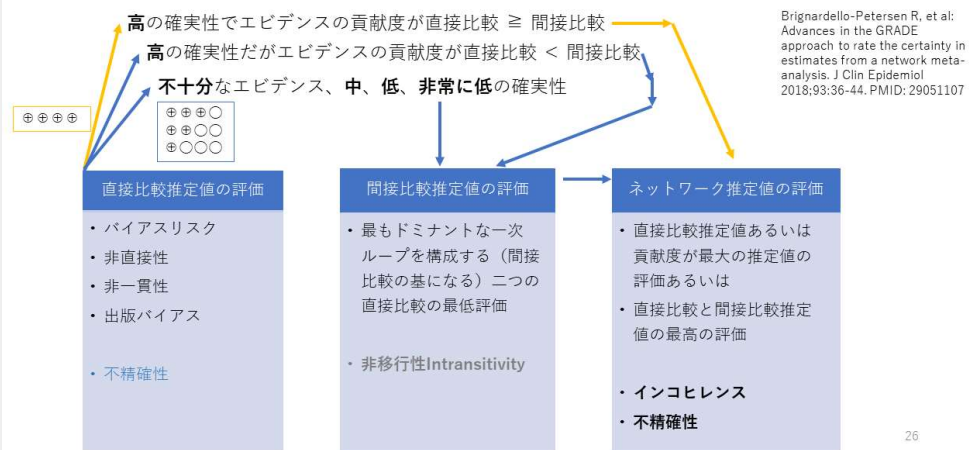
Puhan MA, et al: A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ 2014;349:g5630. PMID: 25252733 <https://www.bmj.com/content/349/bmj.g5630> より一部抜粋翻訳 (Table 1)。

## ネットワーク推定値の確実性評価プロセス



GRADE working groupが提言している、ネットワークメタアナリシスで得られるネットワーク推定値の、エビデンスの確実性の評価の、より詳細な手順を示します。

一つの例を説明すると、比較するペアに直接比較のデータがある場合、そのエビデンスの確実性が高であれば、ネットワーク推定値のエビデンス評価を行います。直接比較の貢献度が大きければそのエビデンスの確実性をネットワーク推定値のエビデンスの確実性とします。もし、直接比較と間接比較の貢献度が同じ場合は、高い方のエビデンスの確実性をネットワーク推定値のエビデ



スの確実性とします。その際にインコヒレンスとネットワーク推定値の不精確性を考慮し、これらが深刻な場合は、レートダウンすることを考慮します。

もう一つの例として、間接比較しかない場合、直接比較の不十分なエビデンスに該当するとみなし、間接比較推定値の評価へ進みます。二つの直接比較のエビデンスの確実性の低い方を採用して、非移行性も評価し、ネットワーク推定値の評価へと進みます。そこで、不精確性を考慮し、これらが深刻な場合は、レートダウンすることを考慮します。

## インコヒレンスIncoherenceと異質性Heterogeneity

- 異質性：研究間の効果修飾因子（研究デザイン、研究実施、バイアス、非直接性など）の分布の違いによって生じる効果の違い。より上位の概念。
- **インコヒレンス**：直接比較と間接比較の効果修飾因子（研究デザイン、研究実施、バイアス、非直接性など）の分布の違いによって生じる効果の違い。
- インコヒレンスはネットワークのクローズループ内で比較を構成する3つ以上の独立したエビデンスソースがないと評価できない。
- 点推定値の類似性、信頼区間の重なり、点推定値間の統計学的検定で評価する。

27

同じ疾患の患者さんを対象にし、同じ介入の効果を調べた研究でも、まったく同じ効果推定値が得られることはありません。研究結果は、偶然による偏りに加え、バイアス、非直接性の影響を受けています。複数の研究の間で、偶然によるばらつき以上の、異なる効果推定値が得られた場合、研究間に異質性 Heterogeneityがあると考えられます。

インコヒレンスは異質性に含まれますが、直接比較と間接比較の効果修飾因子（研究デザイン、研究実施、バイアス、非直接性など）の分布の違いによって生じる効果の違いに限

定して用います。通常のパラ比較のメタアナリシスでは問題になることはありませんが、ネットワークメタアナリシスは間接比較のデータと直接比較のデータの両方を用いるため、それらが乖離する場が起きうることになり、効果推定値の確実性に影響を及ぼす因子になります。

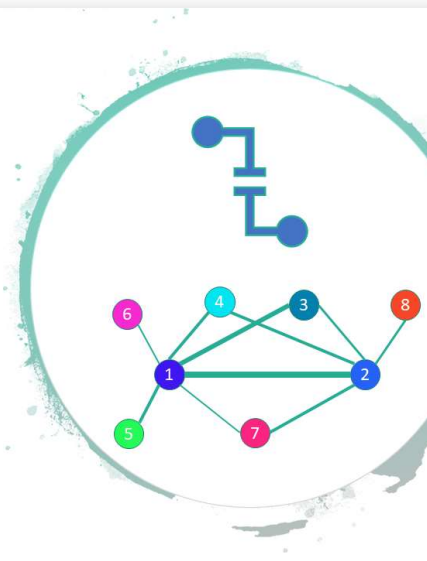
インコヒレンスはネットワークのクローズループ内で比較を構成する3つ以上の独立したエビデンスソースがないと評価できません。

## インコヒレンス

少なくとも3つの独立したエビデンスソース（RCTなど）があり、クローズドループの中で3つあるいはそれ以上の比較をする場合に生じうる。

マルチアーム（同時に3つ以上の治療を比較する）試験（のメタアナリシス）ではインコヒレンスはない。（それ自身での不一致のみ）。

ネットワークの中にマルチアーム試験が混ざっているとICDFとインコヒレンスはより複雑になる。



インコヒレンスは、少なくとも3つの独立したエビデンスソース（RCTなど）があり、クローズドループの中で3つあるいはそれ以上の比較をする場合に生じる可能性があります。

これは、GRADE working groupが提唱している、インコヒレンスの評価フローチャートです。

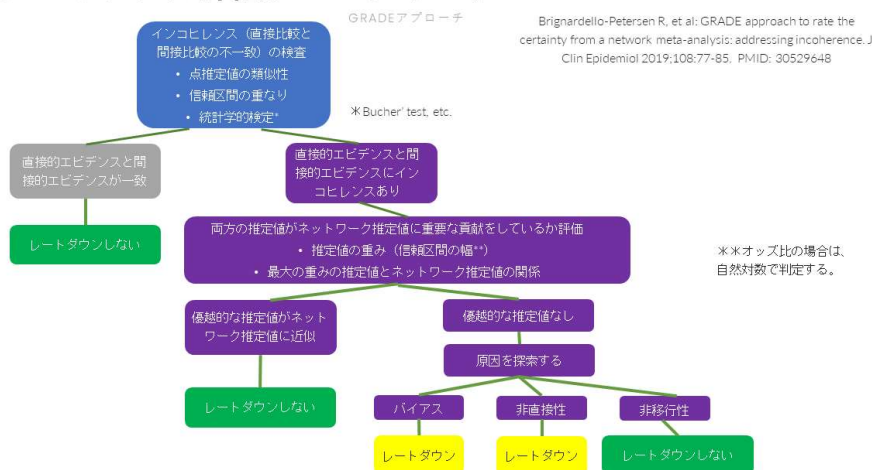
まず、直接比較と間接比較の効果推定値の点推定値が類似しているか、信頼区間が重なっているか、そしてブッチャーの検定などの統計学的検定の結果から、インコヒレンスの有無を判定します。インコヒレンスがないと判断した場合には、レートダウンはしません。

もしインコヒレンスがあると判断した場合には、間接比較と直接比較の効果推定値のそれぞれの信頼区間の幅から、推定値の重みを判断し、最大の重みの推定値とネットワーク推定値の関係をみます。重みの大きな推定値とネットワーク推定値が近似している場合にはレートダウンする必要はありません。ネットワーク推定値をほぼ決定付けているような優越的な推定値がなかった場合には、そのインコヒレンスの原因を探索し、バイアスあるいは非直接性の影響と判断した場合には、レートダウンします。もし非移行性が原因と考えられた場合には、通常レートダウンをしません。

# インコヒレンス評価フローチャート



Quiz



実際にRのパッケージgemtc [https://cran.r-](https://cran.r-project.org/web/packages/gemtc/index.html)

## GeMTCでNMAを実行

30

[project.org/web/packages/gemtc/index.html](https://cran.r-project.org/web/packages/gemtc/index.html) を用いてネットワークメタアナリシスを実行します。

## NMAのためのソフトウェア

- OpenBUGS (Bayesian inference Using Gibbs Sampling)
- OpenBUGS via BRugs with R
- JAGS via rjags with R (Just Another Gibbs Sampler)
  - gemtc → Contrast-based models
  - pcnetmeta → Arm-based models
- netmeta with R
- Others

ネットワークメタアナリシスを実行するためのソフトウェアには、ベイジアンアプローチを用いるもの、頻度論派のアプローチを用いるものがあります。ベイジアンアプローチの場合には、OpenBUGS、あるいは、RからBRugsなどを介してOpenBUGS動かす方法を用いることが一般的です。特に gemtcは、Rでrjagsを介してJAGSを動かして、ベイジアンアプローチによるContrast-based modelによるネットワークメタアナリシス解析を行うソフトウェアです。また、pcnetmetaはベイジアンアプローチによるArm-based modelによるネットワークメタアナリシスを、Rからrjagsを介してJAGS

31

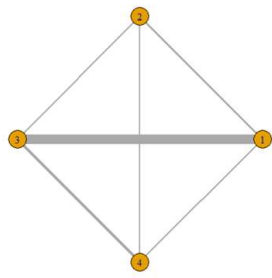
を動かして、実行します。

頻度論派のアプローチを用いるソフトウェアとしてnetmetaが広く用いられています。

ここでは、gemtcで禁煙治療のサンプルデータを用いて、ベイジアンアプローチによるContrast-based modelによるネットワークメタアナリシス解析を実際にデモでお見せします。

# ネットワークメタアナリシス用のデータ

: 二値変数アウトカム



4つの治療；1行に1研究

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

これは4つの治療を比較し、アウトカムが二値変数で測定されているデータの例です。このようなデータをExcelで用意します。

24件の研究で、1つの研究のデータが1行に入力されています。治療は1, 2, 3, 4の整数で表し、1がレフアレンスの治療として用いられています。

例えば、2行目のデータを見ると、r[,1]のカラムの2行目の9は、群1でアウトカムが9人に起きたことを意味します。n[,1]は群1の症例数で140人だったことを意味します。群1の受けた治療は、t[,1]で示されてお

り、治療1であることを示しています。つまり、群1は治療1を140人が受け、その内9人でアウトカムが起きたことを示しています。アウトカムは治療開始後1年時点での禁煙です。

同じく2行目のr[,2]は23ですが、群2で23人にアウトカムが起きたことを示しており、n[,2]は140ですが、群2の症例数が140人であったことを示します。群2の治療はt[,2]が3なので、治療3を受けたことを示します。

r[,3]、n[,3]、t[,3]も同様です。2行目の研究は、3つのアームがあり、治療1, 3, 4を比較するランダム化比較試験です。

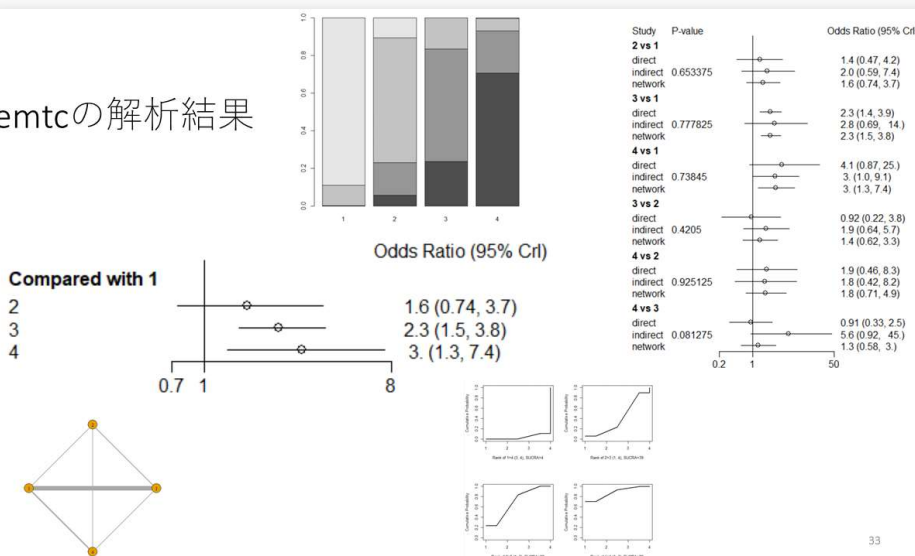
4行目の研究は、r[,3]、n[,3]はNA、t[,3]もNAなので、群1が治療2、群2が治療3を受け、これら2群を比較したランダム化比較試験であることが分かります。

最後の列はラベルがna[]となっていますが、the number of armsすなわち、各研究で直接比較された治療の数を示しています。

これらのデータを各研究から抽出しExcelで準備します。準備ができれば、RとJAGSをインストールし、Rでrjagsとgemtcパッケージをインストールして、Excelでデータ範囲をコピーし、インターネット接続の状態、Rで以下のスクリプトを実行すると結果が得られます。

```
exdat=read.delim("clipboard",sep="¥t",header=TRUE);source("http://zanet.biz/med/stat/ma/nma_gemtc_simpleH
```

## gemtcの解析結果



ネットワークグラフ、フォレストプロット、Rankogram、Node-splitting plot、SUCRA plot、SUCRA値、順位と95%確信区間の値が得られます。

これは3つの治療を比較し、アウトカムが連続変数で測定されているデータの例です。

一つの研究のデータが1行に入力されています。

# ネットワークメタアナリシスのデータ

: 連続変数アウトカム

症例数 平均値 標準偏差 治療番号      ダミー      治療アーム数

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

3つの治療；1行に1研究

これだけ3アーム ←