

日本製薬工業協会 医薬品評価委員会 データサイエンス部会 シンポジウム
「ネットワークメタアナリシスを基礎から学ぶ：概説・事例から留意事項まで」
令和元年(2019年)11月28日

ISPORレポート



日本製薬工業協会 医薬品評価委員会
データサイエンス部会 TF4委員
藤井 陽介
(ファイザーR&D合同会社)

1. はじめに

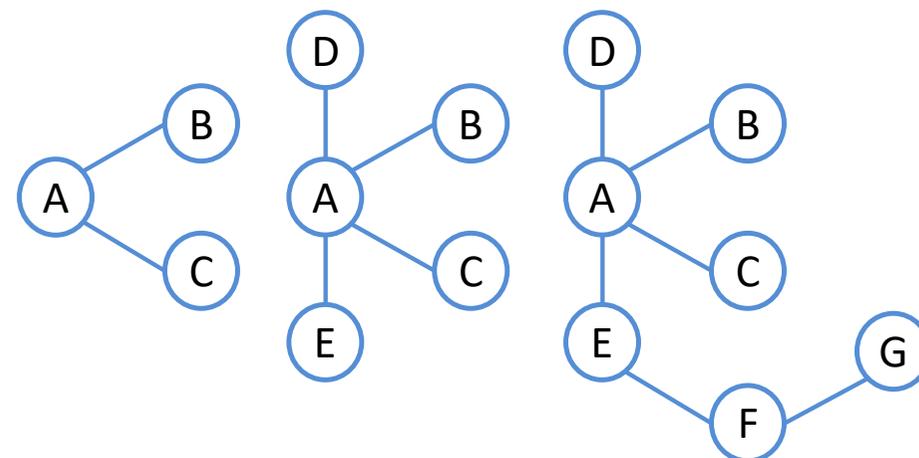
はじめに

- ネットワークメタアナリシス(NMA)とその有用性

- 興味の治療が3種以上
- 対比較のメタアナリシスを一般化
- 直接的な薬剤間比較(直接比較)のみならず
- 間接的な薬剤間比較(間接比較)も可能

- 反面, 留意すべき事項

- 治療の数が多い, 方法論が複雑
- 結果の妥当性のために満たすべき仮定
 - 類似性
 - 一貫性
 - これらの仮定が十分に満たされないとき・・・
 - 結果の解釈が困難
 - 結果にバイアスが生じている可能性



本報告書では治療間の差を「治療効果」で表現している。つまり有効性の意味合い。

報告書について

• 本報告書の方針

- 大々的に引用: ISPOR*のタスクフォース**が公表したNMAに関するレポート(2011年)
 - 本方法論を応用的観点から解説
 - NMAで得られた結果の解釈において注意を要する点
 - 本方法論の利用や結果の活用
- 我々のタスクフォースの見解についてはコラムもしくは脚注として示した
- 本報告書は一部を除いて、統計担当者以外の担当者でも理解できるような内容

* ISPOR: International Society for Pharmacoeconomics and Outcomes Research

(和名: 国際医薬経済・アウトカム研究学会)

** Indirect Treatment Comparisons Good Research Practicesタスクフォース



SCIENTIFIC REPORT

Interpreting Indirect Treatment Comparisons and Network Meta-Analysis for Health-Care Decision Making: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 1

Jeroen P. Jansen, PhD^{1,*}, Rachael Fleurence, PhD², Beth Devine, PharmD, MBA, PhD³, Robbin Itzler, PhD⁴, Annabel Barrett, BSc⁵, Neil Hawkins, PhD⁶, Karen Lee, MA⁷, Cornelis Boersma, PhD, MSc⁸, Lieven Annemans, PhD⁹, Joseph C. Cappelleri, PhD, MPH¹⁰

¹Mapi Values, Boston, MA, USA; ²Oxford Outcomes, Bethesda, MD, USA; ³Pharmaceutical Outcomes Research and Policy Program, School of Pharmacy, School of Medicine, University of Washington, Seattle, WA, USA; ⁴Merck Research Laboratories, North Wales, PA, USA; ⁵Eli Lilly and Company Ltd., Windlesham, Surrey, UK; ⁶Oxford Outcomes Ltd., Oxford, UK; ⁷Canadian Agency for Drugs and Technologies in Health (CADTH), Ottawa, ON, Canada; ⁸University of Groningen / HECTA, Groningen, The Netherlands; ⁹University of Ghent, Ghent, Belgium; ¹⁰Pfizer Inc., New London, CT, USA

ABSTRACT

Evidence-based health-care decision making requires comparisons of all relevant competing interventions. In the absence of randomized, controlled trials involving a direct comparison of all treatments of interest, indirect treatment comparisons and network meta-analysis provide useful evidence for judiciously selecting the best choice(s) of treatment. Mixed treatment comparisons, a special case of network meta-analysis, combine direct and indirect evidence for particular pairwise comparisons, thereby synthesizing a greater share of the available evidence than a traditional meta-analysis. This report from the ISPOR Indirect Treatment Comparisons Good Research Practices Task Force provides guidance on the interpretation of indirect treatment comparisons and network meta-analysis to assist policymakers and health-care professionals in using its findings for decision making. We start with an overview of how networks

of randomized, controlled trials allow multiple treatment comparisons of competing interventions. Next, an introduction to the synthesis of the available evidence with a focus on terminology, assumptions, validity, and statistical methods is provided, followed by advice on critically reviewing and interpreting an indirect treatment comparison or network meta-analysis to inform decision making. We finish with a discussion of what to do if there are no direct or indirect treatment comparisons of randomized, controlled trials possible and a health-care decision still needs to be made. **Keywords:** Bayesian, decision making, comparative effectiveness, indirect treatment comparison, mixed treatment comparison, network meta-analysis.

Copyright © 2011, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

- 報告書の主な内容(全5節)

- 第2節： NMA概要・意義，ネットワーク種類，統合方法
- 第3節： NMAの結果を利用する者に向けた，批判的なレビューや解釈について
- 第4節： NMAの結果の解釈及び意思決定を行う上で考慮すべきポイント

- 有用なその他の文献

- 同タスクフォースのレポート(2014年)
 - 本報告書では補足的な位置づけとして
- PRISMAのチェックリスト
 - NMAの適切な評価・解釈に
 - 本報告書では引用のみ
- 他

2. ネットワークメタアナリシスの概要

2.3. エビデンスの統合

- 間接比較及び混合比較の妥当性を担保するための仮定
- 解析手法並びに効果の指標の選択時の留意点
- 統計モデル
- パラメータ推定の方法
- 観察研究による比較の利用

異質性(heterogeneity)

- 統合する試験の計画などが十分類似しているか
- ランダム化*は試験内の比較の内的妥当性のみ保証

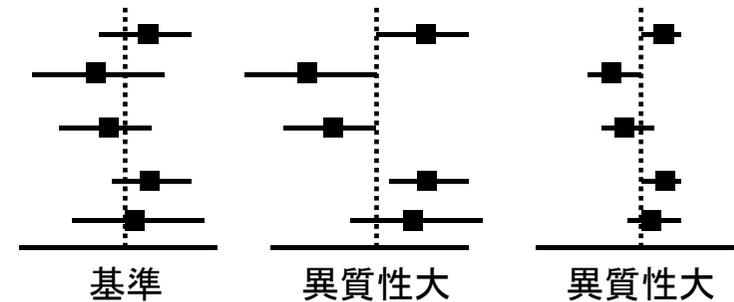
– ランダム化で群間で患者特性等がバランス

* 本報告書では基本的にランダム化比較試験(RCT)のメタアナリシスの文脈で説明しています。

- 試験間では...

- 患者特性がバランスがしているとは限らない, 試験特性も
- それらが治療効果の修飾因子の場合, 統合される試験は異質であるとされ, 推定値が異質性を有する

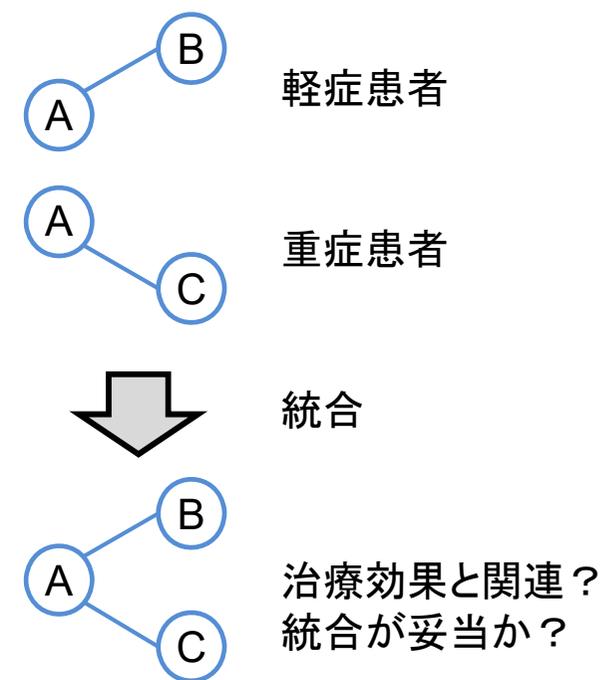
- 対比較のメタアナリシスでも問題になる性質



[患者特性や試験特性]
年齢(結果, 組入基準)
アウトカムの定義や測定方法
併用可能な薬剤
観察期間

間接比較及び混合比較の妥当性を担保するための仮定 「類似性(similarity)」(transitivityとも)

- NMA特有の性質です
 - 間接比較の妥当性に関する性質。異質性の間接比較版。
- 治療A-BとA-Cの比較間で試験特性に差異
かつその差異が治療効果の修飾因子であるとき
 - 治療Aをアンカーとした治療B-Cの間接比較の推定値はバイアス
 - 間接比較の類似性(similarity)の仮定が未成立(右図)
 - アンカーを治療を通じて統合という文脈ではtransitivityとも



間接比較及び混合比較の妥当性を担保するための仮定

「治療効果の修飾因子」(赤下線は報告書に記載の因子)

患者特性

交絡因子
※ Ecological biasについては後ほど

試験特性(試験の実施計画上の特性)

- アウトカム定義や測定方法, 測定時期
- 併用可能な薬剤
- 観察期間
- 試験実施時期
- Estimand
- 中間解析

同じ名称の評価項目でも、定義・測定方法が異なることも。測定時期が用法と関連することもある。

直接的に修飾因子として寄与しているわけではないが、例えば医療環境などを説明するものとして

それぞれのポリシーで解析方法が異なることも

推定の精度は有意性を示すためのギリギリの程度

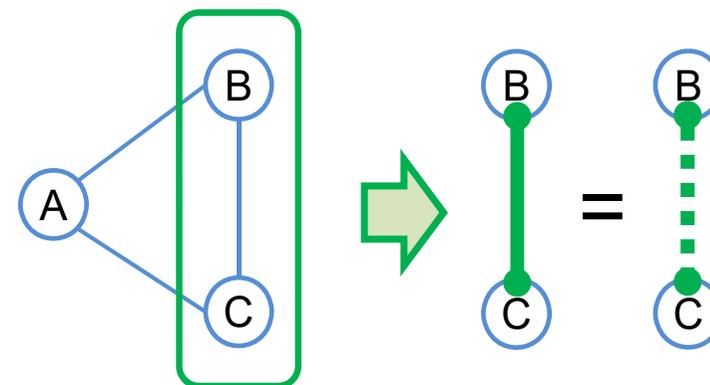
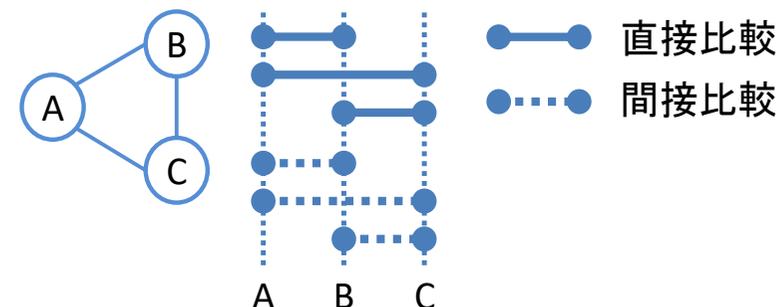
間接比較及び混合比較の妥当性を担保するための仮定 「一貫性(consistency)」

- NMA特有の性質です

- 類似性と似ているが...
- エビデンスループ全体での直接・間接比較に関する性質

- 直接比較と間接比較の推定値に乖離がない

- 直接比較(D_{BC})と間接比較($D_{AC} - D_{AB}$)の間の一貫性
 - $D_{BC} = D_{AC} - D_{AB}$ が満たされる(右図)
 - ※ 実際は誤差があるので、等しくないからといって、不一致であるとみなすわけではない
- 3組の対比較のうち1つ以上で、相対的な治療効果の修飾因子の分布に偏りがある場合不一致となる(右図)
- 一貫性はエビデンスのループに対して適用される概念
 - 上では治療B-Cの比較に限定して説明したが、本質的には全ての対比較で一貫性を満たすかに興味



統計モデル

「変量効果モデル，メタ回帰，患者レベルのデータのNMA」

- メタアナリシスの目的は，治療効果を統合して推定することだが・・・
- 異質性などの状況下でどのようにそれを測り，治療効果を統合するのか？

変量効果モデル： 異質性をモデル化（異質性パラメータを導入）

- 治療効果を言い当てようとしているが，治療効果の修飾因子が原因で，試験単位でユレは生じる
- 同じ薬剤比較の治療効果パラメータに関し交換可能性を仮定。でも，原因そのものを説明しない

メタ回帰： 試験特性の影響をモデル化

※ 部分集団解析で限定するアプローチもある

- 試験単位でのユレの原因をモデルで表現する（治療と共変量の交互作用をモデルに含める）
- しかし，共変量は試験レベルの変量で，Ecological biasが生じる

患者レベルのデータのメタアナリシス： 患者レベルのデータをモデル化

- ある意味究極的
- メタアナリシスの対象の全試験の患者レベルのデータが入手可能か？ハイブリッドなアプローチある

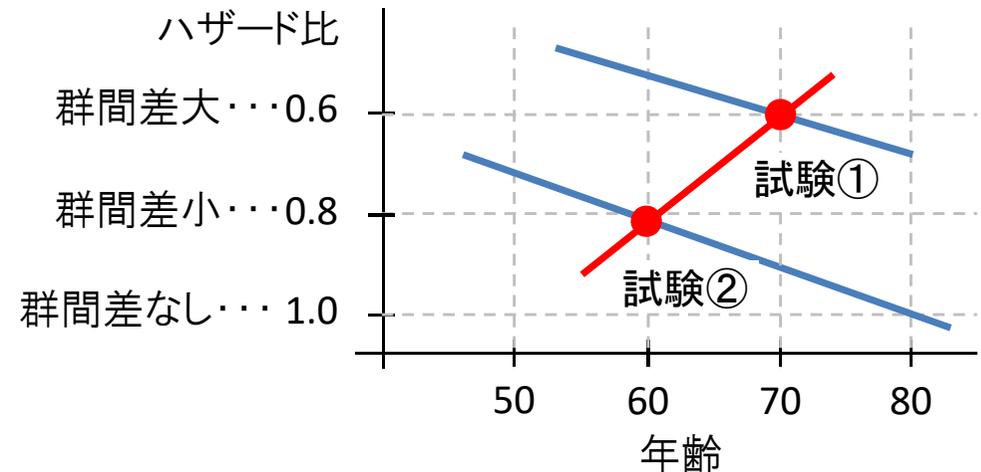
Ecological Bias

- 試験間に認められた関連性の方向が、各試験で観察された関連性の方向と反対となる状況を引き起こす可能性のあるバイアス

- メタ回帰で用いられる変量は試験レベルの代表値である
- その共変量や変量のアウトカムに対する真の影響を誤って推定してしまうことがある

- 例

- アウトカム・・・全生存(指標・・・全生存率)
- 年齢中央値, ハザード比
 - 試験①・・・70歳, 0.6
 - 試験②・・・60歳, 0.8
- 年齢は治療効果の修飾因子
 - 年齢高→ハザード比が1に近づく
 - 疾患, 薬効とは関係なく, 生存率低下



「NMAに観察研究を含めること」

- NMAに観察研究を含め、ネットワークを充実させたい
- 観察研究を含めることは可能。だが、バイアスの懸念が高まる
 - 治療効果の修飾因子による問題に加え、観察研究由来の特有の問題
 - 未観測の交絡因子：ただでさえ「各研究でバイアス」が懸念されているのに、統合してしまうと…
 - 観測した交絡因子：メタアナリシスでは試験特性でのみ調整可能。交絡調整は困難(ecological bias)
- 観察研究を含めるのであれば
 - バイアスのリスク* を評価する
 - 選択バイアス, 交絡バイアス, 曝露変数やアウトカムの測定に関連したバイアス
 - 観察研究の対比較のメタアナリシスに関するチェックリスト(Stroupら)
 - 感度分析** をする
 - 観察研究のエビデンスを除いたときの統合結果の変化を確認(RCT + 観察研究のNMAの場合)
 - 交絡の影響の観点から、観察研究の結果を変動させたときの統合結果の変化を確認

*および** いずれも後の「NMAの事例」で議論。ただし観察研究というよりNMA全般の話題として。

3. ネットワークメタアナリシスの報告のレビュー

3.1. 内的妥当性と外的妥当性

- NMAの結果に対して吟味が必要。それも批判的に
 - 自分たちにとって〇〇な結果
 - 魅力のある結果……そのまま受け止めてしまいたくなる
 - 不都合な結果……蓋をして追いやってしまいたくなる
 - 著名な論文誌の論文だから信頼できる？……そうとは限らないかも
 - 自分たちのリサーチクエスチョンと照らして参考に値するか？
 - 参照：治療効果の修飾因子のスライド
 - 例：どういう患者集団を対象にして、どういう評価項目を測っているか？
- 批判的レビューの観点
 - NMAの内的妥当性と外的妥当性
 - NMAの評価や解釈に役立つチェックリスト

- 「観察された相対的な治療効果が交絡なく治療間の差によるものとし、それにより因果が治療によるものであるとできる程度」
- NMAにおける解析の内的妥当性は、以下の3つの要素を条件として成立する

ネットワークを構成する
試験の適切な同定

各RCTの質

類似性や一致性の逸脱による
交絡を原因とする
バイアスの程度

- 文献検索の適切さ
 - 適切な検索, 選択
 - 公表バイアスの検討
 - 特定の比較が多いことによる不釣り合いなネットワーク
- 各試験の内的妥当性
 - ネットワークに含まれる各試験の内的妥当性
 - ランダム化はバイアスがないことを完全に保証しているわけではない
 - 治療群の盲検の欠如
 - ランダム化後に試験中止する患者数の群間での不均衡の原因となり得る。
 - アウトカムの盲検の欠如
 - 治療効果の過大評価をもたらすかもしれない

- 試験特性(試験計画など)の類似性
 - 治療効果の修飾因子(右記再掲)
 - この因子の分布に相違がある場合, NMAの結果は交絡による影響を受ける
 - 前述の通り, 共変量は試験レベルの変量であり, ecological biasが生じうる
 - 一方, ベースラインリスクの差やプラセボ反応の差が, 試験レベルの変量の差を反映したものである可能性があることもある

患者特性

試験特性(試験の実施計画上の特性)

- アウトカム定義や測定方法, 測定時期
- 併用可能な薬剤
- 観察期間
- 試験実施時期
- Estimand
- 中間解析

- 「一般化可能性ともいう。NMAで得られた結果を、NMAの対象とした集団より広い集団において、一般性を失わず適用できることを意味する。」
- NMAの結果が興味のある母集団に外挿できるかをレビューする必要がある。
 - NMAの外的妥当性は、各試験の外的妥当性によって制限される
 - 治験は外的妥当性をある程度犠牲にした集団となる傾向がある
 - 幅広い集団でネットワークが構築でき、興味の対象の集団と相違ないとしても
 - 類似性や一致性に影響を及ぼすこともあるかもしれない(統合することの妥当性に影響)



3. ネットワークメタアナリシスの報告のレビュー

3.2. ネットワークメタアナリシスの報告

チェックリスト

- チェックリストの位置づけ
 - NMAの実施者として, NMAの報告に含めるべき項目のリスト
 - NMAの評価者として, 意思決定を補助するリスト
- レポートの質の評価に焦点
 - NMAの内的妥当性や外的妥当性そのものを評価することは意図していない

導入	試験実施の根拠や目的が明確に示されているか。
方法	方法の節に以下が含まれているか。 <ul style="list-style-type: none">- 組入・除外基準- 情報源- 検索方針- 試験選択の手順- データ抽出(個々の試験の妥当性や品質の評価)- アウトカム指標が記載されているか。 解析方法や統合方法の記載があるか。また, 示されている方法には以下が含まれているか。 <ul style="list-style-type: none">- 解析方法やモデルの記載- 潜在的なバイアスや不一致性への対応- 解析のフレームワーク 感度分析について示されているか。
結果	ネットワークに含まれた試験の要約が示されているか。 <ul style="list-style-type: none">- 個々の試験のデータ- 試験のネットワーク モデル適合度の評価について記載があるか。また, 競合するモデルが比較されているか。統合したエビデンス(間接比較や直接比較)の結果が明確に示されているか。感度分析及びシナリオに基づいた解析
考察	考察には以下が含まれているか。 <ul style="list-style-type: none">- 主な結果について記載され, 要約されているか。- 解析に関する内的妥当性- 外的妥当性- 対象者に対する結果の意味合い

導入 試験実施の根拠や目的が明確に示されているか。

• チェックの方針

- 興味の対象の患者集団や競合する治療が何であるか。選択した根拠も示す
- 意思決定上のリサーチクエスチョンは何か
- 参考： ISPOR TFのレポート(2014)
 - アウトカムを示すとともに、それが意思決定者にとって意味のあるものか。評価時期も示す
 - 各試験の背景(状況や環境)と興味に相違がないか
 - 例： 各試験の実施年は、疾患の医療背景が劇的に変わったときや、標準治療が変わったときに興味の対象となりうる

チェックリスト 「方法」

方法	方法の節に以下が含まれているか。 - 組入・除外基準 - 情報源 - 検索方針 - 試験選択の手順 - データ抽出(個々の試験の妥当性や品質の評価) - アウトカム指標が記載されているか。
	解析方法や統合方法の記載があるか。また、示されている方法には以下が含まれているか。 - 解析方法やモデルの記載 - 潜在的なバイアスや不一致性への対応 - 解析のフレームワーク
	感度分析について示されているか。

• チェックの方針

- システムティックレビューの方法の事前規定: 検索ロジック, 試験の選択のための組入・除外基準など
- 試験選択後のデータ抽出方法: データ抽出方法(二者によるデータ抽出, その不一致に対する解決法, 欠測データの扱い)
- 統計解析法: 一致性の考察, 類似性や一致性の向上のための試験レベルの共変量の導入

結果	ネットワークに含まれた試験の要約が示されているか。 - 個々の試験のデータ - 試験のネットワーク
	モデル適合度の評価について記載があるか。また、競合するモデルが比較されているか。
	統合したエビデンス(間接比較や直接比較)の結果が明確に示されているか。
	感度分析及びシナリオに基づいた解析

• チェックの方針

- 試験のリストの提示: システマティックレビューで同定した試験, NMAの解析対象とした試験
 - データが不適切(例:推定値が欠損)な試験があるとき, 2つのリストには相違が生じる
 - 試験の選択方法を説明するフロー図は有用(PRISMAステートメント)
- 主な患者特性及び試験特性の各試験のリスト: 治療効果の修飾因子の探索
 - 観測された治療効果の修飾因子について試験間で均衡がとれていたとしても, 未知もしくは未観測の治療効果の修飾因子があるリスク
 - 治療効果の修飾因子の同定もできるだけ, 事前に規定しておく必要がある
 - 事前に得られている知見や各試験で報告されたサブグループ解析の結果をもとに
- エビデンスネットワークの図

結果	ネットワークに含まれた試験の要約が示されているか。 - 個々の試験のデータ - 試験のネットワーク
	モデル適合度の評価について記載があるか。また、競合するモデルが比較されているか。
	統合したエビデンス(間接比較や直接比較)の結果が明確に示されているか。
	感度分析及びシナリオに基づいた解析

• チェックの方針

- **各試験の結果**: ①各試験の点推定値と信頼区間。②各試験の治療効果
 - ① 試験間で共通の対照治療の結果を比較することで、試験間の相違を評価するのに役立つ
 - ② 図(フォレストプロット)や表は、NMAの統合結果と各試験の結果間の比較することができる
- **NMAの結果**: 共通の対照治療(or アンカー)に対する治療効果の推定値や信頼区間(信用区間)
 - 全ての対比較の結果を示すことが推奨される
- **順位**: ベイズ流の方法を用いた場合には、当該治療の順位を求めることもできる
 - 単に、1位となる確率にのみ基づいて順位を示すのは誤解する可能性
 - 全ての治療に対して効果の順位の期待値を算出することが有用である(SUCRA)

SUCRA (surface under the cumulative ranking)

順位付けのための指標

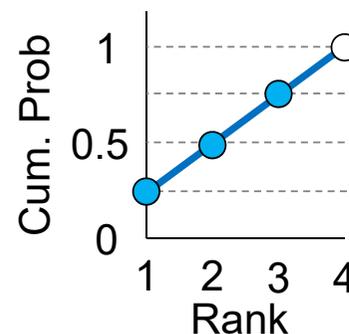
- ランキングプロット
- SUCRA [Salanti et al.]

$$SUCRA_j = \frac{\sum_{b=1}^{a-1} cum_{j,b}}{a-1}$$

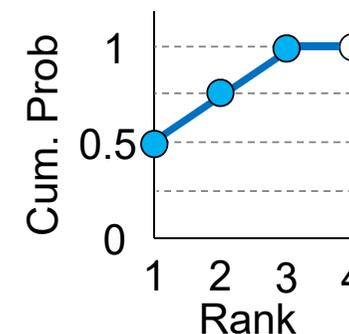
a: 薬剤数
 b: 順位
 j: 治療
 cum_{j,b}: 累積確率

Rank <i>b</i>	Treatment <i>j</i>			
	A	B	C	D
1	0.25 (0.25)	0.50 (0.50)	0.25 (0.25)	0 (0)
2	0.25 (0.50)	0.25 (0.75)	0.50 (0.75)	0 (0)
3	0.25 (0.75)	0.25 (1)	0.25 (1)	0.25 (0.25)
4	0.25 (1)	0 (1)	0 (1)	0.75 (1)
<i>SUCRA</i>	50%	75%	67%	8%

Treatment A



Treatment B



SUCRAの解釈

- 100% …… どの薬剤と比較しても完全に優れている治療
- 0% …… どの薬剤と比較しても完全に劣っている治療
- X% …… どの薬剤にも完全に優れている治療の有効性のX%を達成することを意味する

チェックリスト 「考察」

考察 考察には以下が含まれているか。

- 主な結果について記載され、要約されているか。
- 解析に関する内的妥当性 - 外的妥当性
- 対象者に対する結果の意味合い

• チェックの方針

- 内的妥当性及び外的妥当性の批判的な評価
- 類似性と一致性を仮定が妥当か評価
- 結果について
 - 当該NMAの結果が、他のメタアナリシスを含む他のエビデンスに基づく結果に沿ったものか
 - 生物学的、臨床的な観点から、比較した治療間の観察された相違を説明
- 実臨床での意思決定に関する結論の妥当性
- 利益相反について、ISPOR TFのレポート(2014)より
 - すべての利益相反が明示され、投稿した論文は査読を受けている必要
 - 各著者の役割を開示するために、貢献した内容が明確に記載されている必要

2014年のISPORタスクフォースによるチェックリスト

Relevance	
	Is the population relevant?
	Are any relevant interventions missing?
	Are any relevant outcomes missing?
	Is the context (settings and circumstances) applicable?
Credibility	
Evidence Base Used for the Indirect Comparison or Network Meta-Analysis	Attempt to include all relevant RCTs?
	1 network?
	No poor quality RCTs?
	No differences in effect modifiers between direct comparisons?
Analysis	Naiive comparisons avoided?
	Consistency assessed?
	With consistency, was direct & indirect evidence included?
	Account for inconsistency/Minimize bias?
	Valid rationale for FE/RE model?
	Rationale for heterogeneity assumptions in RE model discussed?
	Subgroup or meta-regression analysis?
Reporting Quality and Transparency	Network & source data presented?
	Direct & indirect results reported?
	Are all contrasts presented with uncertainty?
	Ranking of treatments presented?
	Results by subgroup or levels of effect-modifiers presented?
Interpretation	Conclusions fair & balanced?
Conflict of Interest	Conflict of interest? If yes, steps taken to address these?

4. 解釈と意思決定

4.1. 解釈

4.2. 意思決定

• 解釈

- 類似性に問題があるとき
 - 例： NMAに用いる試験数が少なくメタ回帰をするには限界がある
 - 諦めず，感度分析
 - 例： 間接比較におけるバイアスの方向について仮説を立て，バイアスのない間接比較とどれだけ異なり得るのか，バイアスにより誤った結論や意思決定を導き出してしまうのか検討
- “最良”な治療は有効性評価項目だけでは特定できない
 - 安全性，利便性といった有効性以外の特徴も踏まえてるべき
 - 総合的な解釈を得るため複数の評価項目に対するNMAの結果のまとめ方を検討

• 意思決定

- 意思決定のためのエビデンスとして何が適切か？
 - RCT, RCTに基づくNMA, 観察研究, (観察研究を含むNMA)
- 意思決定の質は向上させるためには・・・, 上記のどのエビデンスを用いるにしても
 - どんなタイプのエビデンスが意思決定に利用されたのかを明確にし, 透明性を担保する
 - 意思決定の限界と影響を評価する