Quantifying indirect evidence in network meta-analysis

野間 久史
統計数理研究所
2017年5月25日
横浜市立大学 京浜統計セミナー
e-mail: noma@ism.ac.jp
URL: http://normanh.skr.jp

本日のお話

Research Article
Statistics in Medicine

Received: 28 February 2016, Accepted: 9 November 2016 Published online 4 December 2016 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/sim.7187

Quantifying indirect evidence in network meta-analysis

Hisashi Noma,[a*†] Shiro Tanaka,[b] Shigeyuki Matsui,[c] Andrea Cipriani[d] and Toshi A. Furukawa[c]

Network meta-analysis enables comprehensive synthesis of evidence concerning multiple treatments and their simultaneous comparisons based on both direct and indirect evidence. A fundamental pre-requisite of network meta-analysis is the consistency of evidence that is obtained from different sources, particularly whether direct and indirect evidence are in accordance with each other or not, and how they may influence the overall estimates. We have developed an efficient method to quantify indirect evidence, as well as a testing procedure to evaluate their inconsistency using Lindsay's composite likelihood method. We also show that this estimator has complete information for the indirect evidence. Using this method, we can assess the degree of consistency between direct and indirect evidence and their contribution rates to the overall estimate. Sensitivity analyses can be also conducted with this method to assess the influences of potentially inconsistent treatment contrasts on the overall results. These methods can provide useful information for overall comparative results that might be biased from specific inconsistent treatment contrasts. We also provide some fundamental requirements for valid inference on these methods concerning consistency restrictions on multi-arm trials. In addition, the efficiency of the developed method is demonstrated based on simulation studies. Applications to a network meta-analysis of 12 new-generation antidepressants are presented. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: network meta-analysis; indirect evidence; likelihood factorization; composite likelihood methods; inconsistency; sensitivity analysis

- ネットワークメタアナリシスにおける間接エビデンスの定量的な評価方法を開発した
- Lindsayの複合尤度法を用いることによって、一致性／漸近有効性を有する間接エビデンスの要約統計量を構成
- ネットワーク上の直接・間接エビデンスのInconsistencyを評価する有効な検定手法を導出

Statist. Med. 2017, 36: 917-927. 2

メタアナリシス　Meta-Analysis

- 過去に行われた臨床試験の結果を統合し、総合的な治療効果の評価を行うための方法
- 個々の試験において、サンプルサイズが十分に確保できず、十分な精度・検出力での評価ができなくても、試験間の情報を併せることにより、より精度の高い評価が可能となる

3

抗うつ薬のメタアナリシス

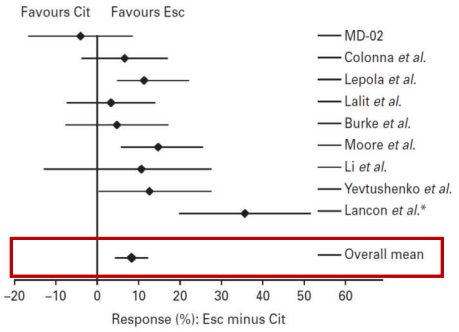


Fig. 2. Estimated difference in response rates at week 8 (or last assessment if <8 wk) (Esc−Cit) with 95% confidence intervals. Esc, Escitalopram; Cit, citalopram. Response is defined as the percentage of patients with ⩾50% improvement from baseline (MADRS or $HAMD_{17}$). * The naturalistic study of Lançon *et al.* (2006) was not included in the calculation of the overall mean.

- EscitaopramとCitaopramの有効性を比較したメタアナリシス
- ESC vs. CITの比較を行った8つのランダム化臨床試験を系統的に収集し、その結果を統合
  - N=2009（ESC 995人, CIT 1014人）
- 8つの試験の結果によって推定された総合的な治療効果は、反応率の差 8.3% (95%CI: 4.4-12.3) で、ESCのほうが有意に反応率が高いという結果に

Montgomery et al. (2011)　4

蓄積されるメタアナリシスのエビデンス

"…statistically significant differences in terms of efficacy …. between **Fluoxetine and Venlafaxine**, but the clinical meaning of these differences is uncertain…"

"…**meta-analysis** highlighted a trend in favour of **Sertraline** over other **Fluoxetine**"

"Although **Mirtazapine** is likely to have a faster onset of action than **Sertraline and Paroxetine** no significant differences were observed…"

Which interventions work? In whom?

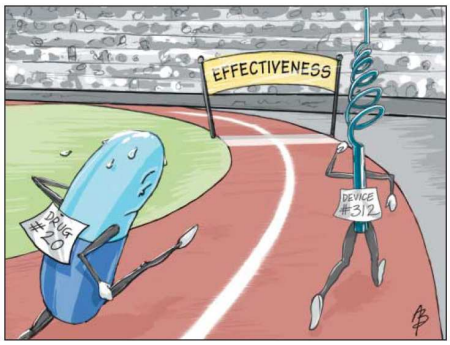
5

従来のメタアナリシスの限界と難点

- ２つの治療の対比較についての情報しか得られない？？
- 実際には、うつ病の治療薬として使用されている薬剤には、多くの種類がある
  - 個々の薬剤の有効性・安全性は異なるかもしれない
  - 薬価もさまざま（Fluoxetine: 28€, Venlafaxine:111€, Sertaline: 76€, etc）
- 利用可能なすべての薬剤について、有効性・安全性を系統的に比較・評価した、信頼のできるエビデンスを得るには？？

6

Comparative Effectiveness Research (CER)



Comparative effectiveness research allows investigators to determine the superiority, inferiority, or equivalence of various interventions when pitted against each other.

Mitka, M. (2010). US Government Kicks Off Program for Comparative Effectiveness Researches. JAMA 304: 2230-1.

- The Institute of Medicine committee has defined CER as “the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels.”

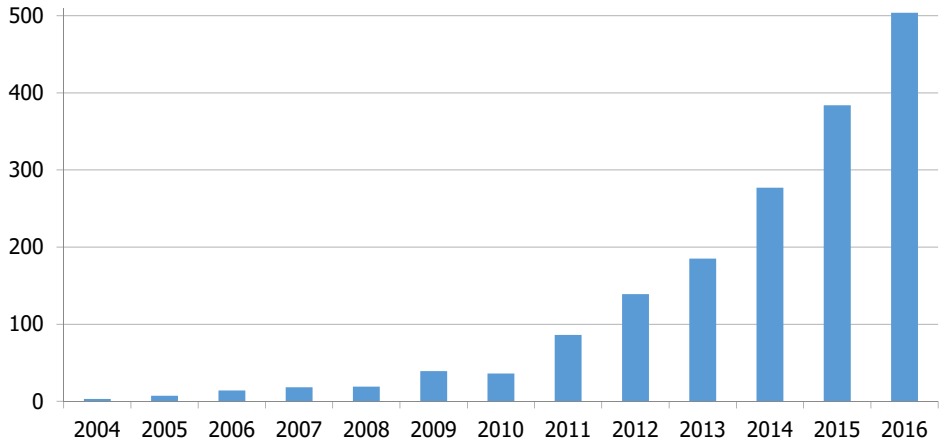
http://www.nap.edu/read/12648/chapter/4 7

Network Meta-Analysis (NMA)

- 有効性についてのエビデンスが確立された複数の治療法のComparative Effectivenessを評価するために新たに開発されたエビデンス統合のための方法論
- 2種類の治療の対比較に限らず、利用可能な複数（3種類以上）の治療法を、過去の臨床試験から得られたエビデンスを統合し、系統的に比較するためのメタアナリシスの方法論
- Multiple treatment comparison meta-analysis, Mixed-treatment comparisonなどともいわれる

Caldwell et al. (2005), Salanti (2012) 8

Number of Publications

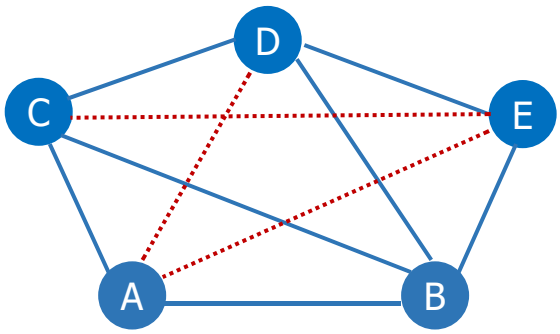


Pubmed search: ("network meta-analysis" OR "multiple treatment comparison meta-analysis“ OR "mixed comparison meta-analysis")

9

‘Network’ Meta-Analysis

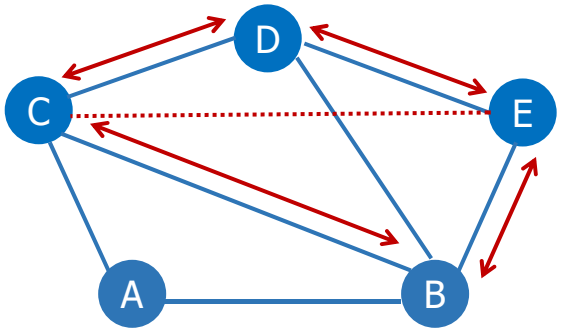
- 直接比較のエビデンスがあるパスを通して、治療効果を比較することができる
- 直接比較のエビデンスがないパスでの比較を行うことはできる？？



10

Synthesizing *Indirect* Evidence

- 直接比較のエビデンスだけでは、すべての治療のComparative Effectivenessを評価することができない？？
- ネットワーク上の間接的な連結のあるパスの情報を活用する！！



11

NMA for 12 new-generation antidepressants

Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis

Andrea Cipriani, Toshiaki A Furukawa, Georgia Salanti, John R Geddes, Julian P T Higgins, Rachel Churchill, Norio Watanabe, Atsuo Nakagawa, Ichiro M Omori, Hugh McGuire, Michele Tansella, Corrado Barbui

Summary

Background Conventional meta-analyses have shown inconsistent results for efficacy of second-generation antidepressants. We therefore did a multiple-treatments meta-analysis, which accounts for both direct and indirect comparisons, to assess the effects of 12 new-generation antidepressants on major depression.

Methods We systematically reviewed 117 randomised controlled trials (25 928 participants) from 1991 up to Nov 30, 2007, which compared any of the following antidepressants at therapeutic dose range for the acute treatment of unipolar major depression in adults: bupropion, citalopram, duloxetine, escitalopram, fluoxetine, fluvoxamine, milnacipran, mirtazapine, paroxetine, reboxetine, sertraline, and venlafaxine. The main outcomes were the proportion of patients who responded to or dropped out of the allocated treatment. Analysis was done on an intention-to-treat basis.

Findings Mirtazapine, escitalopram, venlafaxine, and sertraline were significantly more efficacious than duloxetine (odds ratios [OR] 1·39, 1·33, 1·30 and 1·27, respectively), fluoxetine (1·37, 1·32, 1·28, and 1·25, respectively), fluvoxamine (1·41, 1·35, 1·30, and 1·27, respectively), paroxetine (1·35, 1·30, 1·27, and 1·22, respectively), and reboxetine (2·03, 1·95, 1·89, and 1·85, respectively). Reboxetine was significantly less efficacious than all the other antidepressants tested. Escitalopram and sertraline showed the best profile of acceptability, leading to significantly fewer discontinuations than did duloxetine, fluvoxamine, paroxetine, reboxetine, and venlafaxine.

Interpretation Clinically important differences exist between commonly prescribed antidepressants for both efficacy and acceptability in favour of escitalopram and sertraline. Sertraline might be the best choice when starting treatment for moderate to severe major depression in adults because it has the most favourable balance between benefits, acceptability, and acquisition cost.

Funding None.

Published Online January 29, 2009 DOI:10.1016/S0140-6736(09)60046-5

See Online/Comment DOI:10.1016/S0140-6736(09)60047-7

Department of Medicine and Public Health, Section of Psychiatry and Clinical Psychology, University of Verona, Italy (A Cipriani PhD, C Barbui MD, Prof M Tansella MD); Department of Psychiatry and Cognitive-Behavioral Medicine, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan (Prof T A Furukawa MD, N Watanabe PhD, I M Omori PhD); Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Greece (G Salanti PhD); Department of Psychiatry, University of Oxford, UK (A Cipriani, Prof J R Geddes MD); MRC Biostatistics Unit Institute of Public Health, University of Cambridge, UK

12

12種類の新世代抗うつ薬のNMA

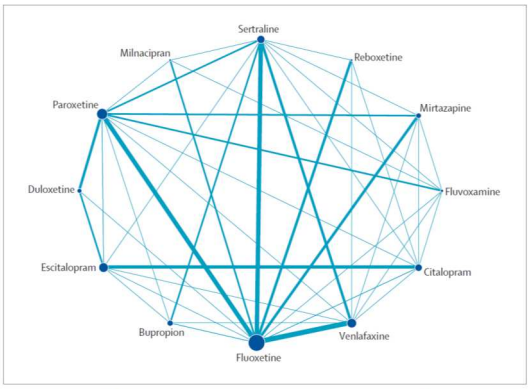


Figure 2: **Network of eligible comparisons for the multiple-treatment meta-analysis for efficacy (response rate)**
The width of the lines is proportional to the number of trials comparing each pair of treatments, and the size of each node is proportional to the number of randomised participants (sample size). The network of eligible comparisons for acceptability (dropout rate) analysis is similar.

- 1991年から2007年に実施された、左記の12種類の抗うつ薬のいずれかを比較したランダム化臨床試験（117試験；25,298 人）を系統的に収集
- 有効性に関するアウトカムは、割りつけられた治療への反応の有無（=0,1）
- 効果の指標には、Odds Ratio (OR) を採用
- ネットワークメタアナリシスによって、12種類の抗うつ薬の有効性・許容性を系統的に評価

Cipriani, Furukawa, Salanti et al. (2009)

13

NMAの利点①：推定精度・検出力の向上

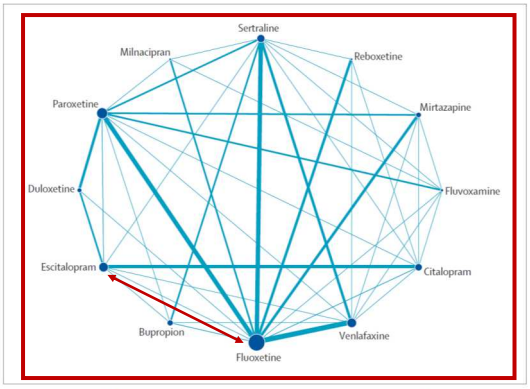


Figure 2: **Network of eligible comparisons for the multiple-treatment meta-analysis for efficacy (response rate)**
The width of the lines is proportional to the number of trials comparing each pair of treatments, and the size of each node is proportional to the number of randomised participants (sample size). The network of eligible comparisons for acceptability (dropout rate) analysis is similar.

- 間接比較のエビデンスを加えることにより、より精度の高い評価が可能になる
- Escitalopram vs Fluoxetine
  - 直接比較　2試験
    OR 1.23, 95%CI: 0.87, 1.75
  - 間接比較のエビデンスを加えると…
    OR 1.30, 95%CI: 1.12, 1.50
- 統計的な情報量が大きくなるため、精度が高まり、有意な結果が得られる

Cipriani, Furukawa, Salanti et al. (2009)

14

NMAの利点②：直接比較のエビデンスがなくても

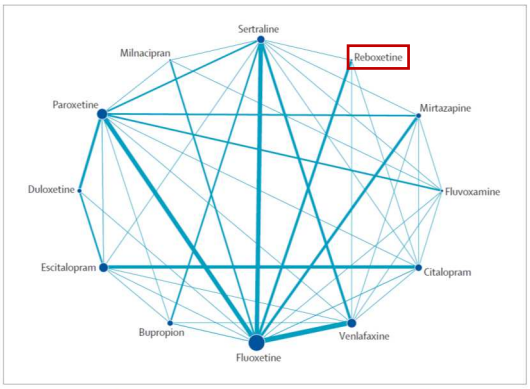


Figure 2: **Network of eligible comparisons for the multiple-treatment meta-analysis for efficacy (response rate)**
The width of the lines is proportional to the number of trials comparing each pair of treatments, and the size of each node is proportional to the number of randomised participants (sample size). The network of eligible comparisons for acceptability (dropout rate) analysis is similar.

- 直接比較をした試験が１つもないパスでも、間接比較のエビデンスによって、ネットワーク上のその他のすべての治療と有効性を比較することができる
- Reboxetineは、ネットワーク上で、直接比較のパスは、Sertraline, Fluoxetine, Venlafaxine, Citalopramの4薬剤しかないが…

Cipriani, Furukawa, Salanti et al. (2009)

15

Reboxetine vs.

	OR (95%CI)
Bupropion	0.63 (0.46, 0.83)
Citalopram	0.61 (0.47, 0.80)
Duloxetine	0.68 (0.50, 0.95)
Escitalopram	0.51 (0.39, 0.68)
Fluoxetine	0.68 (0.53, 0.86)
Fluvoxamine	0.69 (0.50, 0.97)
Milnacipran	0.67 (0.46, 0.97)
Mirtazapine	0.49 (0.36, 0.66)
Paroxetine	0.67 (0.50, 0.86)
Sertraline	0.54 (0.41, 0.71)

他の11薬剤すべてに有意に劣るという結果に

16

NMAの利点③：Comparative Effectivenessの評価

- ネットワーク上のすべての治療を比較した治療効果の差の推定値，信頼区間，P値を得ることができる（オッズ比，ハザード比など）
- 治療の順位付け（ランキング）を行い、対象となった治療間でのそれぞれの順位を推定することができる

17

対比較オッズ比の推定値，信頼区間

Efficacy (response rate) (95% CI)　Comparison　Acceptability (dropout rate) (95% CI)

BUP	1·00 (0·78-1·28)	0·75 (0·55-1·01)	1·06 (0·86-1·32)	0·89 (0·74-1·08)	0·73 (0·53-1·00)	0·87 (0·58-1·24)	0·87 (0·66-1·14)	0·81 (0·65-1·00)	**0·62** (0·45-0·86)	1·01 (0·82-1·27)	0·84 (0·68-1·02)
0·98 (0·78-1·23)	CIT	0·75 (0·55-1·02)	1·07 (0·86-1·31)	0·90 (0·73-1·09)	**0·73** (0·54-0·99)	0·87 (0·60-1·24)	0·87 (0·66-1·15)	0·81 (0·65-1·01)	**0·62** (0·45-0·84)	1·02 (0·81-1·28)	0·84 (0·67-1·06)
1·09 (0·83-1·43)	1·12 (0·87-1·44)	DUL	**1·43** (1·09-1·85)	1·19 (0·91-1·57)	0·98 (0·67-1·41)	1·16 (0·77-1·73)	1·16 (0·83-1·61)	1·08 (0·84-1·40)	0·83 (0·57-1·22)	**1·36** (1·01-1·83)	1·12 (0·84-1·50)
0·82 (0·67-1·01)	0·84 (0·70-1·01)	**0·75** (0·60-0·93)	ESC	0·84 (0·70-1·01)	**0·69** (0·50-0·94)	0·81 (0·55-1·15)	0·81 (0·62-1·07)	**0·76** (0·62-0·93)	**0·58** (0·43-0·81)	0·95 (0·77-1·19)	**0·78** (0·64-0·97)
1·08 (0·90-1·29)	1·10 (0·93-1·31)	0·99 (0·79-1·24)	**1·32** (1·12-1·55)	FLU	0·82 (0·62-1·07)	0·97 (0·69-1·32)	0·97 (0·77-1·21)	0·91 (0·79-1·05)	**0·70** (0·53-0·92)	1·14 (0·96-1·36)	0·94 (0·81-1·09)
1·10 (0·83-1·47)	1·13 (0·86-1·47)	1·01 (0·74-1·38)	**1·35** (1·02-1·76)	1·02 (0·81-1·30)	FVX	1·18 (0·76-1·75)	1·18 (0·87-1·61)	1·10 (0·84-1·47)	0·85 (0·57-1·26)	**1·38** (1·03-1·89)	1·14 (0·86-1·54)
1·07 (0·77-1·48)	1·09 (0·78-1·50)	0·97 (0·69-1·38)	1·30 (0·95-1·78)	0·99 (0·74-1·31)	0·97 (0·68-1·37)	MIL	0·99 (0·69-1·53)	0·94 (0·68-1·31)	0·72 (0·48-1·10)	1·17 (0·84-1·72)	0·97 (0·69-1·40)
0·79 (0·72-1·00)	0·80 (0·63-1·01)	**0·72** (0·54-0·94)	0·96 (0·76-1·19)	**0·73** (0·60-0·88)	**0·71** (0·55-0·92)	0·74 (0·53-1·01)	MIR	0·93 (0·75-1·17)	0·72 (0·51-1·03)	1·17 (0·91-1·51)	0·97 (0·76-1·23)
1·06 (0·87-1·30)	1·08 (0·90-1·30)	0·97 (0·78-1·20)	**1·30** (1·10-1·53)	0·98 (0·86-1·12)	0·96 (0·76-1·23)	1·00 (0·74-1·33)	**1·35** (1·11-1·64)	PAR	0·77 (0·56-1·05)	**1·25** (1·04-1·52)	1·03 (0·86-1·24)
1·60 (1·20-2·16)	**1·63** (1·25-2·14)	**1·46** (1·05-2·02)	**1·95** (1·47-2·59)	**1·48** (1·16-1·90)	**1·45** (1·03-2·02)	**1·50** (1·03-2·18)	**2·03** (1·52-2·78)	**1·50** (1·16-1·98)	REB	**1·63** (1·19-2·24)	1·34 (0·99-1·83)
0·87 (0·72-1·05)	0·88 (0·72-1·07)	0·79 (0·62-1·01)	1·06 (0·88-1·27)	**0·80** (0·69-0·93)	0·79 (0·61-1·01)	0·81 (0·60-1·11)	1·10 (0·90-1·36)	**0·82** (0·69-0·96)	**0·54** (0·41-0·71)	SER	0·82 (0·67-1·00)
0·85 (0·70-1·01)	0·86 (0·71-1·05)	**0·77** (0·60-0·99)	1·03 (0·86-1·24)	**0·78** (0·68-0·90)	**0·77** (0·59-0·99)	0·79 (0·58-1·08)	1·08 (0·87-1·33)	**0·79** (0·67-0·94)	**0·53** (0·40-0·69)	0·98 (0·82-1·16)	VEN

Cipriani, Furukawa, Salanti et al. (2009)　18

Fluoxetineを基準とした結果

	Efficacy (response rate) OR (95% CI)	Acceptability (dropout rate) OR (95% CI)
Bupropion	0·93 (0·77–1·11)	1·12 (0·92–1·36)
Citalopram	0·91 (0·76–1·08)	1·11 (0·91–1·37)
Duloxetine	1·01 (0·81–1·27)	0·84 (0·64–1·10)
Escitalopram	0·76 (0·65–0·89)*	1·19 (0·99–1·44)
Fluvoxamine	1·02 (0·81–1·30)	0·82 (0·62–1·07)
Milnacipran	0·99 (0·74–1·31)	0·97 (0·69–1·32)
Mirtazapine	0·73 (0·60–0·88)*	0·97 (0·77–1·21)
Paroxetine	0·98 (0·86–1·12)	0·91 (0·79–1·05)
Reboxetine	1·48 (1·16–1·90)*	0·70 (0·53–0·92)*
Sertraline	0·80 (0·69–0·93)*	1·14 (0·96–1·36)
Venlafaxine	0·78 (0·68–0·90)*	0·94 (0·81–1·09)

OR=odds ratio. CI=credibilty interval. *p<0·05. For efficacy, OR higher than 1 favours fluoxetine. For acceptability, OR lower than 1 favours fluoxetine.

Table 4: **Efficacy and acceptability using fluoxetine as reference compound**

19

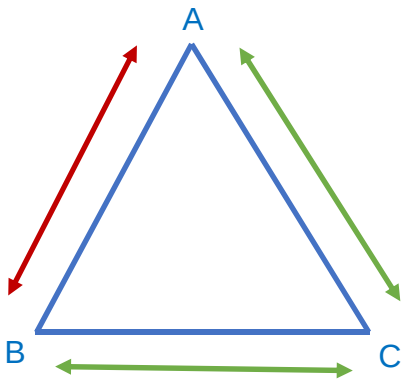
三竦み（さんすくみ）

- 蝲蛆食蛇、蛇食蛙、蛙食蝲蛆、互相食也

《「関尹子」三極から》蛇はなめくじをおそれ、なめくじは蛙（かえる）をおそれ、蛙は蛇をおそれること。転じて、三者が互いに牽制し合って、それぞれが自由に動けない状態。

20

ネットワーク上の比較の妥当性



- NMAにおいて、直接比較・間接比較のエビデンスを統合して、治療間の妥当な比較を行うためには、直接比較・間接比較のパスにおける治療間の差が一致していなくてはならない
  - 直接比較：A vs B
  - 間接比較：B vs C, C vs A
- 2つのパスにおけるA-B間の差が一致しなくては、NMAにおける治療の比較の妥当性は失われてしまう！！

Salanti (2012), Dias et al. (2013)

21

Inconsistency

- Consistency: ネットワーク上の直接・間接エビデンスによる治療間のEffect Sizeの差が一致（consistent）すること
  - ネットワーク上の任意の治療のペアにおいて、直接・間接エビデンスによるEffect Sizeの差が互いに一致して整合すること
  - より概念的には、Transitivityと言われることも
- Inconsistency: Lack of consistency
  - ネットワーク上の治療の比較の妥当性が、前提として成り立たないということに！！

22

Local inconsistency on the network

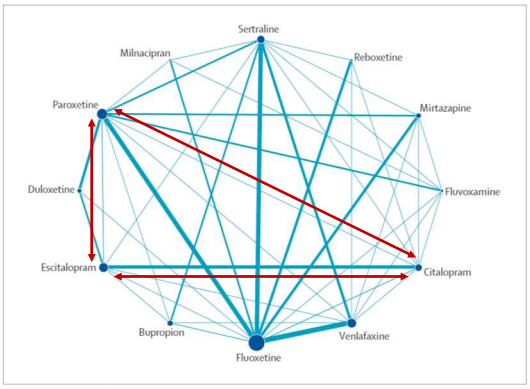
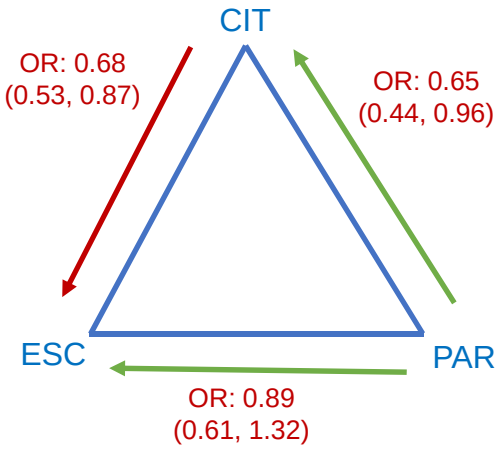


Figure 2: **Network of eligible comparisons for the multiple-treatment meta-analysis for efficacy (response rate)** The width of the lines is proportional to the number of trials comparing each pair of treatments, and the size of each node is proportional to the number of randomised participants (sample size). The network of eligible comparisons for acceptability (dropout rate) analysis is similar.

- ネットワーク上の任意のTriangleに対して、三辺のパスにおけるEffect SizeはConsistentか？
- NMAの解析モデルは、ネットワーク上のすべてのパスにConsistencyを仮定しているので、1つでもそれが崩れると、妥当性が成り立たない
- Triangle上のInconsistencyは、簡単な手順で検定をすることができる
- 例えば、CIT-ESC-PARのTriangleでは…

Cipriani, Furukawa, Salanti et al. (2009) 23

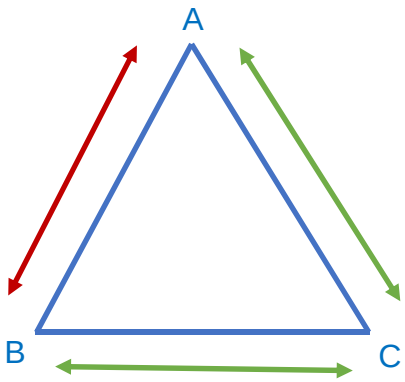
CIT-ESC-PARのTriangle Loop



- CIT vs. ESCの直接比較では、ESCのほうが有意に反応率が高い
- CIT vs. PARでは、CITのほうが有意に反応率が高い
- ESCは、PARよりも有意に反応率が高い…
- わけではなく、ESCとPARの直接比較では、有意差が出ず、ORも 0.89 ほどしかない
- いわゆる「三竦」に近い状態！！
- ESC-PAR-CIT間のパス上のEffect Sizeは一致しない？？

Cipriani, Furukawa, Salanti et al. (2009) 24

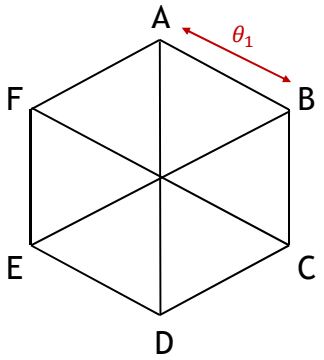
Inconsistencyの評価



- Inconsistency: Whether direct and indirect evidence accords or not, which have been obtained from different sources.
- Consistencyの仮定が崩れるようであれば、ネットワーク上の治療間の比較の妥当性が成り立たないことに
- Direct comparison: A vs B
- Indirect comparison: B vs C, C vs A
  - How the direct and indirect comparisons are consistent?

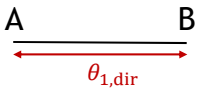
Salanti (2012), Dias et al. (2013) 25

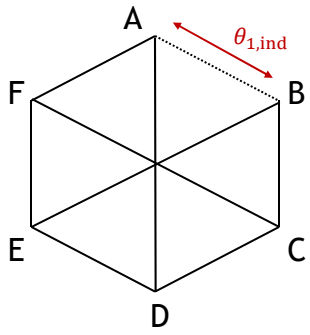
Inconsistency on the Entire Network



Entire Network

Direct comparison for A vs B





Indirect comparison network except for direct comparison of A vs B

研究の目的

- ネットワーク上の特定のパス（A vs. B）において、間接比較のエビデンスを要約するための有効な方法を開発する
- これによって、直接比較・間接比較のエビデンスのInconsistencyを検定する有効な検定手法を与える
- 加えて、最終的な推定値に対する直接比較・間接比較のエビデンスの寄与率を評価する方法を開発
- ネットワーク上の間接エビデンスの直感的・定量的な解釈を可能にするとともに、最終的な結果に影響し得る、潜在的なバイアスを探索するためにも有用な手法となる

27

Multivariate Random-Effects Model

- Suppose Y_{ir} is the estimated treatment effect for the rth outcome in the ith trial ($i = 1, 2, \dots, N$; $r = 1,2, \dots, p$), which is defined as a contrast to a common baseline treatment (e.g., placebo).
  - $Y_{ir} \sim N(\mu_{ir}, s_{ir}^2)$, $\mu_{ir} \sim N(\theta_r, \tau_r^2)$
- s_{ir}^2 is within-study variance (ordinarily, fixed to a valid estimate).
- θ_r is corresponding to the average treatment effect, and τ_r^2 is across-studies variance.

White (2011), White et al. (2012) 28

Matrix Notation

- $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{ip}),\ (i = 1,2, \ldots, N)$
- $Y_i \sim N(\mu_i, S_i),\ \mu_i \sim N(\theta, \Sigma)$
- $\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \ldots \\ \theta_p \end{pmatrix}, S_i = \begin{pmatrix} s_{i1}^2 & \rho_{i12}s_{i1}s_{i2} & \cdots & \rho_{i1p}s_{i1}s_{ip} \\ \rho_{i12}s_{i1}s_{i2} & s_{i2}^2 & \ldots & \rho_{i2p}s_{i2}s_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{i1p}s_{i1}s_{ip} & \rho_{i2p}s_{i2}s_{ip} & \cdots & s_{ip}^2 \end{pmatrix}$
- $\Sigma = \begin{pmatrix} \tau_1^2 & \kappa_{12}\tau_1\tau_2 & \cdots & \kappa_{1p}\tau_1\tau_p \\ \kappa_{12}\tau_1\tau_2 & \tau_2^2 & \ldots & \kappa_{2p}\tau_2\tau_p \\ \vdots & \vdots & \ddots & \vdots \\ \kappa_{1p}\tau_1\tau_p & \kappa_{2p}\tau_2\tau_p & \cdots & \tau_p^2 \end{pmatrix}$

White (2011), White et al. (2012)

対数尤度関数

- Parameter vector：$\eta = (\theta_1, \ldots, \theta_p, \tau_1, \ldots, \tau_p, \kappa_{12}, \ldots, \kappa_{(p-1)p})$

$$\ell(\eta) = \sum_{i=1}^{N} \log p(y_{i1}, \ldots, y_{ip}|\eta)$$

$$= -\frac{1}{2}\sum_{i=1}^{N}\{\log|\Sigma + S_i| + (y_i - \theta)W_i(y_i - \theta)^T + p_i \log 2\pi\}$$

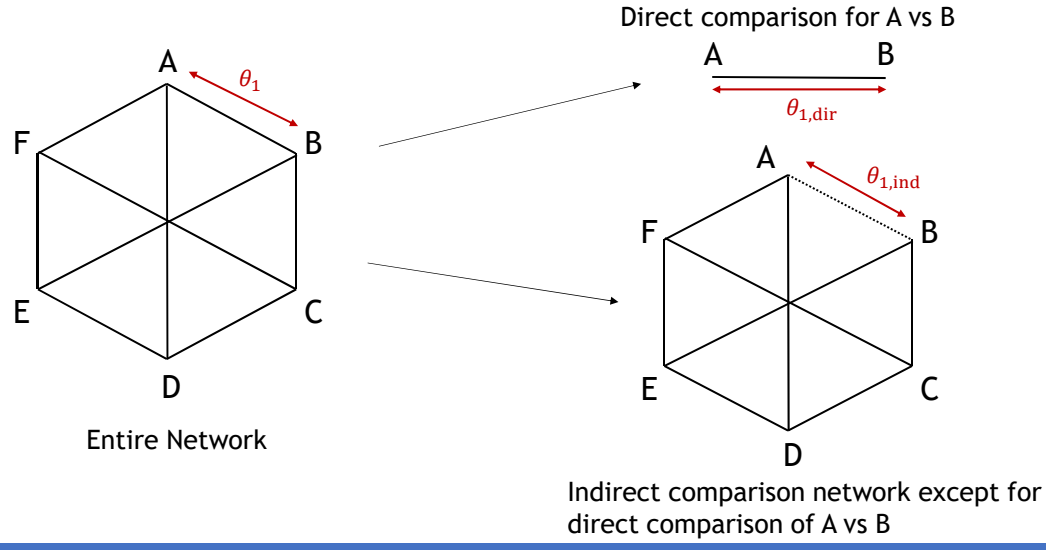
- where $W_i = (\Sigma + S_i)^{-1}$

Setting and Objective

- 一般性を失うことなく、ここでは、関心のある対比較が、Y_is $(Y_{i1}, \ldots, Y_{iN})$ のうちの1番目の成分であるとする（便宜上、この比較をA vs B の比較と呼ぶことにする）
- θ_1 の最尤推定値は、対数尤度関数 $\ell(\eta)$ を最大化することによって得ることができる
- 直接比較のエビデンス：A vs. Bの直接比較のエビデンスによる θ_1の最尤推定値
- 間接比較のエビデンス：全体のネットワーク上から、直接比較の情報を除いた θ_1 の推定値
- θ_1の間接エビデンスの情報はどのように要約すればよいのだろうか？？

31

Inconsistency on the Entire Network



直接比較のエビデンス

- 直接比較 A vs B に対するDerSimonian-Lairdの変量効果モデル
  - $Y_{i1} \sim N(\mu_{i1}, s_{i1}^2)$, $\mu_{i1} \sim N(\theta_1, \tau_1^2)$
- 対数尤度関数

$$\ell_{\mathrm{dir}}(\eta) = \sum_{i=1}^{N} \log p(y_{i1}|\theta_1, \tau_1^2) = -\frac{1}{2}\sum_{i=1}^{N}\left\{\log(\tau_1^2 + s_{i1}^2) + \frac{(y_{i1}-\theta_1)^2}{\tau_1^2 + s_{i1}^2} + \log 2\pi\right\}$$

- 直感的には、間接比較の情報は、ネットワーク全体の尤度 $\ell(\eta)$ から $\ell_{\mathrm{dir}}(\eta)$ を差し引いたものから得られると考えられる

DerSimonian and Laird (1982)

尤度の分解 (1)

- 1. 2 arm trials that compares A vs B

$$\ell_i(\eta) = \log p(y_{i1}|\theta_1, \tau_1^2)$$

  - Only involves direct comparison information
- 2. More than 2 arm trials that include A vs B

$$\ell_i(\eta) = \log p(y_{i1}|\theta_1, \tau_1^2) + \log p(y_{i2}, \ldots, y_{ip}|y_{i1}, \eta)$$

  - that can be decomposed to a marginal likelihood of Y_{i1} and conditional likelihood of $(Y_{i2}, \ldots, Y_{ip})$ conditioned on $Y_{i1} = y_{i1}$.

尤度の分解 (2)

- 3. The other trials that do not include A vs B

$$\ell_i(\eta) = \log p(y_{i2}, \ldots, y_{ip}|\eta)$$

  - that do not involve direct comparison information of A vs B.

35

尤度の分解

$$\ell(\eta) = \sum_{i\in\Xi_{\mathrm{I}}} \log p(y_{i1}|\theta_1, \tau_1^2) + \sum_{i\in\Xi_{\mathrm{II}}} \log p(y_{i1}|\theta_1, \tau_1^2) + \sum_{i\in\Xi_{\mathrm{II}}} \log p(y_{i2}, \ldots, y_{ip}|y_{i1}, \eta) + \sum_{i\in\Xi_{\mathrm{III}}} \log p(y_{i2}, \ldots, y_{ip}|\eta)$$

- 対数尤度関数は、直接比較のエビデンスの尤度の成分とそれ以外の成分に、加法的に分解することができる
- 後者の残余成分を、全体のネットワークにおける、間接エビデンスの情報を持つコンポネントと定義する

36

間接エビデンスの複合尤度関数

$$\ell_{\mathrm{ind}}(\eta) = \sum_{i \in \Xi_{\mathrm{II}}} \log p(y_{i2}, \ldots, y_{ip} | y_{i1}, \eta) + \sum_{i \in \Xi_{\mathrm{III}}} \log p(y_{i2}, \ldots, y_{ip} | \eta)$$

- 対応する対数尤度の成分は、それぞれの試験の周辺尤度，条件付尤度によって構成される
- つまり、これを擬似尤度と見なせば、対応する擬似スコア関数の不偏性（推定量の一致性）は必ず保証される！
- 間接エビデンスの尤度の情報を十分に利用した推定法に！
- Lindsay (1988) によって提案された複合尤度法（composite likelihood method）と解釈することができる！

37

Theorem 1: 漸近的性質

- Let the maximum composite likelihood estimator (MCLE) of $\ell_{\mathrm{ind}}(\eta)$ as $\hat{\eta}_{\mathrm{ind}}$.
- $\hat{\eta}_{\mathrm{ind}}$ is a consistent estimator of η.
- The asymptotic distribution of $\hat{\eta}_{\mathrm{ind}}$ is

$$\sqrt{N_{\mathrm{ind}}}(\hat{\eta}_{\mathrm{ind}} - \eta) \to N(0, I_{\mathrm{ind}}^{-1}(\eta))$$

  - where $I_{\mathrm{ind}}(\eta) = -E[\partial^2 \ell_{\mathrm{ind}}(\eta) / \partial\eta \partial\eta^T]$, N_{ind} is the number of trials that contribute to $\ell_{\mathrm{ind}}(\eta)$.

38

情報の十分性

- Because of the additivity of the log likelihood

$$\ell(\eta) = \ell_{\mathrm{dir}}(\eta) + \ell_{\mathrm{ind}}(\eta)$$

- Additivity is fulfilled for the information matrices:

$$I(\eta) = I_{\mathrm{dir}}(\eta) + I_{\mathrm{ind}}(\eta)$$

- $\hat{\eta}_{\mathrm{ind}}$ has the sufficient information on the entire network except for the direct comparison component.

39

Theorem 2: 情報量の分解

- The overall estimator $\hat{\theta}$ can be decomposed to the direct and indirect components as

$$\hat{\theta}_1 = w_{\mathrm{dir}}(\eta)\hat{\theta}_{1,\mathrm{dir}} + w_{\mathrm{ind}}(\eta)\hat{\theta}_{1,\mathrm{ind}} + o_p(1)$$
$$\hat{\theta}_j = \hat{\theta}_{j,\mathrm{ind}} + o_p(1),\ (j = 2,\ldots,p)$$

- $w_{\mathrm{dir}}(\eta) = I_{11,\mathrm{dir}}(\eta)/I_{11}(\eta)\,, w_{\mathrm{ind}}(\eta) = 1 - w_{\mathrm{dir}}(\eta)$
- $I_{11,\mathrm{dir}}(\eta) = -E[\partial^2 \ell_{\mathrm{dir}}(\eta)/\partial\theta_1^2],\ I_{11}(\eta) = -E[\partial^2 \ell(\eta)/\partial\theta_1^2]$
- The fraction of the Fisher information accords to the contribution rate of each component!!

40

Inconsistencyの検定

- For a specific pairwise comparison A vs B (θ_1), a formal test can be constructed for evaluating consistency between direct and indirect evidence.
- Test for H_0: $\theta_{1,\mathrm{dir}} = \theta_{1,\mathrm{ind}}$
  - $\hat{\theta}_{1,\mathrm{dir}}$: MLE for the conventional direct comparison
  - $\hat{\theta}_{1,\mathrm{ind}}$: MCLE for the indirect evidence network assuming that the entire network consistency is fulfilled except for the excluded direct comparison component
- Through inversing the asymptotic distributions of $\hat{\theta}_{1,\mathrm{dir}}$ and $\hat{\theta}_{1,\mathrm{ind}}$, the pseudo-Wald test can be constructed.

41

NMA for 12 New Generation Antidepressants

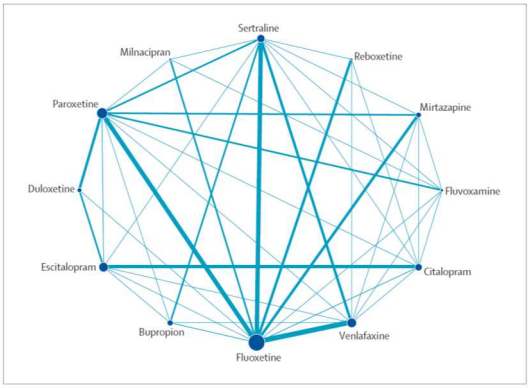


Figure 2: **Network of eligible comparisons for the multiple-treatment meta-analysis for efficacy (response rate)**
The width of the lines is proportional to the number of trials comparing each pair of treatments, and the size of each node is proportional to the number of randomised participants (sample size). The network of eligible comparisons for acceptability (dropout rate) analysis is similar.

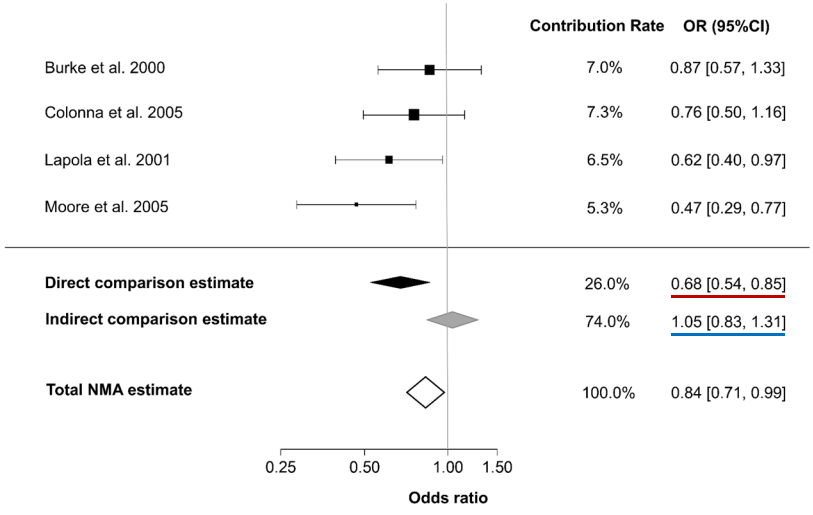
- Review for 117 RCTs (25,298 patients) from 1991 to 2007, which compared 12 antidepressants for the acute treatment of unipolar major depression.
- The outcome was response to the allocated treatment (binary).
- The comparative efficacy for the 12 antidepressants was evaluated by NMA.

Cipriani, Furukawa, Salanti et al. (2009) 42

Selected Results (by the Order of P-values)

	Number of direct comparison	OR for the whole network	Contribution rate of direct comparison	OR for direct comparison	OR for indirect comparison	P-value
Citalopram vs Escitalopram	4	0.84 (0.71, 0.99)	26.0%	0.68 (0.54, 0.85)	1.05 (0.83, 1.32)	0.009
Mirtazapine vs Venlafaxine	2	1.07 (0.88, 1.29)	8.3%	1.53 (1.03, 2.27)	0.95 (0.77, 1.19)	0.041
Bupropion vs Fluoxetine	3	1.06 (0.91, 1.25)	8.0%	0.82 (0.61, 1.11)	1.19 (0.98, 1.44)	0.042
Citalopram vs Paroxetine	1	1.09 (0.92, 1.29)	6.3%	1.54 (1.03, 2.31)	1.02 (0.85, 1.23)	0.068
Bupropion vs Sertraline	3	0.87 (0.73, 1.03)	16.3%	1.07 (0.79, 1.46)	0.78 (0.63, 0.97)	0.103
Fluoxetine vs Sertraline	8	0.81 (0.71, 0.94)	29.3%	0.70 (0.56, 0.89)	0.89 (0.74, 1.06)	0.123
Fluvoxamine vs Milnacipran	1	0.98 (0.71, 1.36)	11.4%	0.57 (0.26, 1.24)	1.10 (0.77, 1.57)	0.131

Escitalopram vs. Citalopram



P=0.009

44

感度解析への応用

- ネットワーク上の特定のパスに不一致性が疑われた？
  - 全体のエビデンスは、それによって妥当性を失うのか？
- 今回の複合尤度法は、直接比較のエビデンスの不一致性が疑われる場合、その情報のみを取り除いた、それ以外のネットワーク全体のエビデンスの統合解析の結果を出力することができる
- 特定の不一致性が疑われる直接比較のエビデンスを除いた場合、全体のエビデンスにどの程度の影響を及ぼしているかの感度解析に応用することもできる！

45

	Entire Evidence		Indirect Evidence [CL]	
	OR (95%C.I.)	P-value	OR (95%C.I.)	P-value
vs. Fluoxetine				
Duloxetine	1.02 (0.83, 1.26)	0.866	1.07 (0.87, 1.33)	0.512
Paroxetine	1.00 (0.89, 1.13)	0.955	1.01 (0.90, 1.14)	0.827
Sertraline	0.81 (0.71, 0.94)	0.004	0.82 (0.71, 0.94)	0.005
Citalopram	0.92 (0.78, 1.07)	0.273	0.81 (0.67, 0.97)	0.024
Escitalopram	0.77 (0.67, 0.89)	0.001	0.85 (0.72, 1.00)	0.045
Fluoxamine	1.02 (0.82, 1.28)	0.834	1.01 (0.81, 1.26)	0.924
Milnacipran	1.01 (0.77, 1.31)	0.966	1.01 (0.77, 1.32)	0.955
Venlafaxine	0.79 (0.70, 0.90)	< 0.001	0.80 (0.70, 0.91)	0.001
Reboxetine	1.48 (1.18, 1.86)	0.001	1.42 (1.13, 1.80)	0.003
Bupropione	0.94 (0.80, 1.11)	0.457	0.96 (0.81, 1.13)	0.621
Mirtazapine	0.74 (0.62, 0.89)	0.001	0.74 (0.62, 0.88)	0.001

46

Concluding Remarks

- ネットワークメタアナリシスにおける、間接比較のエビデンスの要約方法と、それに基づく、直接・間接比較のエビデンスの不一致性の有効な検定手法を開発した
- 単純に、不一致性の検定のP値を評価するだけではなく、間接比較のエビデンスがどの程度の治療効果の差を示しているのか、直感的・定量的な評価が可能である
- 特定の不一致性が疑われるパスに対しての感度解析にも応用することができる
- 実践的にも有用なエビデンスの精査・解釈の方法として利用できることが期待できる

47

文献

- Caldwell, D. M., Ades, A. E., and Higgins, J. P. (2005). Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *British Medical Journal* **331**, 897-900.
- Cipriani, A., Furukawa, T. A., Salanti, G., *et al.* (2009). Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* **373**, 746-758.
- DerSimonian, R., and Larid, N. M. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177-188.
- Dias, S., Welton, N. J., Sutton, A. J., Caldwell, D. M., Lu, G., and Ades, A. E. (2013). Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Medical Decision Making* **33**, 641-656.
- Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics* **80**, 221-239.
- Mitka, M. (2010). US government kicks off program for comparative effectiveness research. *Journal of the American Medical Association* **304**, 2230-2231.

48

- Miura, T., Noma, H., Furukawa, T. A., Mitsuyasu, H., Tanaka, S., Stockton, S., Salanti, G., Motomura, K., Shimano-Katsuki, S., Leucht, S., Cipriani, A., Geddes, J. R. and Kanba, S. (2014). Comparative efficacy and tolerability of pharmacological treatments in the maintenance treatment of bipolar disorder: a network meta-analysis. *Lancet Psychiatry* **1**, 351-359.
- Montgomery, S., Hansen, T., and Kasper, S. (2011). Efficacy of escitalopram compared to citalopram: a meta-analysis. *International Journal of Neuropsychopharmacology* **14**, 261-268.
- Salanti, G. (2012). Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research Synthesis Methods* **3**, 80-97.
- White, I. R. (2011). Multivariate random-effects meta-regression: Updates to mvmeta. *The Stata Journal* **11**, 255-270.
- White, I. R., Barrett, J. K., Jackson, D., and Higgins, J. P. (2012). Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research Synthesis Methods* **3**, 111-125.

49