

Basic assumptions of NMA and key concepts about statistical analysis

Hisashi Noma
 The Institute of Statistical Mathematics
 January 21, 2017
 Network Meta-Analysis Workshop in Kyoto
 e-mail: noma@ism.ac.jp
 URL: <http://normanh.skr.jp>

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

2

抗うつ薬のメタアナリシス

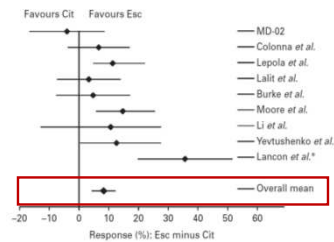


Fig. 2. Estimated difference in response rates at week 8 (or last assessment if <8 wk) (Esc - Cit) with 95% confidence intervals. Esc, Escitalopram; Cit, citalopram. Response is defined as the percentage of patients with $\geq 50\%$ improvement from baseline (MADRS or HAM-D). * The naturalistic study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

- ▶ EscitalopramとCitalopram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) 3

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?

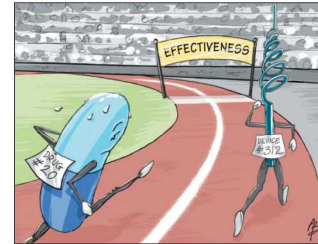
4

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が得られない??
- ▶ 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??**

5

Comparative Effectiveness Research (CER)



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.”

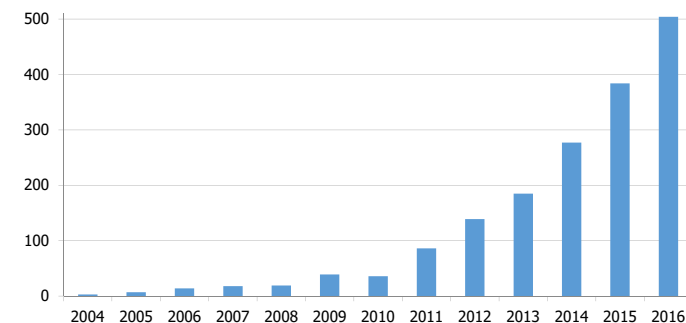
<http://www.nap.edu/read/12648/chapter/4> 6

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) 7

Number of Publications

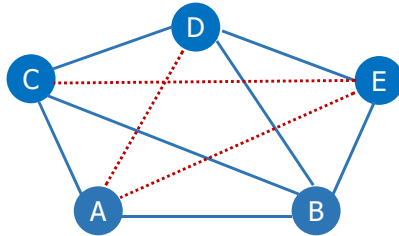


Pubmed search: ("network meta-analysis" OR "multiple treatment comparison meta-analysis" OR "mixed comparison meta-analysis")

8

'Network' Meta-Analysis

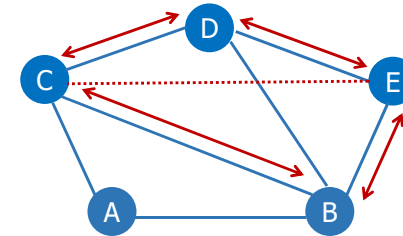
- ▶ ネットワーク上の直接比較のエビデンスがあるパスを通して、対象となった治療間の治療効果を比較することができる
- ▶ 直接比較のエビデンスがないパスでの比較を行うことはできる??



9

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!!



10

NMA for 12 new-generation antidepressants

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

Andreas Cipriani, Toshiki A. Furukawa, Giorgio Salanti, John R. Geddes, Julian P. F. Higgins, Rachel Churchill, Naveen Wilton, et al.

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, mirtazapine, nortriptyline, 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.

11

12種類の新世代抗うつ薬のNMA

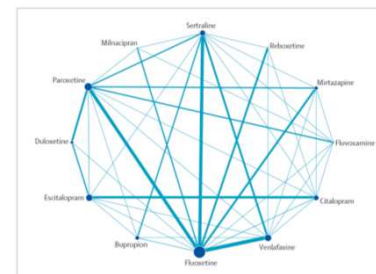


Figure 2: Network of eligible comparisons for the multiple-treatment meta-analysis for efficacy (response rate). The width of the lines is proportional to the number of trials comparing each pair of treatments, and the size of each node is proportional to the number of randomised participants (sample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

12

NMAの利点①：推定精度・検出力の向上

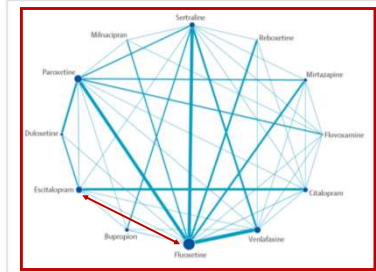


Figure 1: Network of eligible comparisons for the multiple-treatment meta-analysis for efficacy (response rate). The width of the lines is proportional to the number of trials comparing each pair of treatments, and the size of each node is proportional to the number of randomized participants (sample 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) 13

NMAの利点②：直接比較のエビデンスがなくても

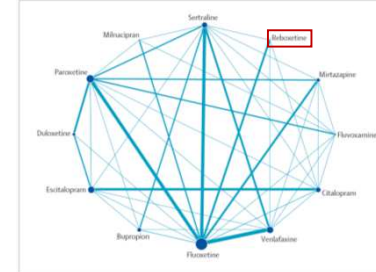


Figure 2: Network of eligible comparisons for the multiple-treatment meta-analysis for efficacy (response rate). The width of the lines is proportional to the number of trials comparing each pair of treatments, and the size of each node is proportional to the number of randomized participants (sample 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) 14

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

15

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

16

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	0.75	1.06	0.89	0.73
0.98	(0.78-1.28)	(0.59-1.01)	(0.74-1.08)	(0.73-1.06)	(0.58-1.14)
CIT	0.75	1.07	0.73	0.87	0.87
(0.59-1.01)	(0.73-1.49)	(0.54-1.61)	(0.60-1.14)	(0.65-1.10)	(0.62-1.22)
DUL	1.43	1.19	0.98	1.16	1.16
(1.09-1.89)	(0.91-1.57)	(0.67-1.41)	(0.77-1.73)	(0.83-1.61)	(0.84-1.69)
ESC	0.84	0.75	0.84	0.69	0.81
(0.63-1.01)	(0.70-1.01)	(0.70-1.01)	(0.50-1.24)	(0.51-1.15)	(0.43-1.07)
FLU	1.10	1.01	1.02	0.97	0.97
(0.90-1.29)	(0.91-1.11)	(0.82-1.22)	(0.69-1.37)	(0.72-1.23)	(0.79-1.35)
FX	1.10	1.13	1.01	1.18	1.18
(0.92-1.47)	(0.89-1.47)	(0.81-1.39)	(0.76-1.72)	(0.67-1.89)	(0.64-1.47)
MIL	1.07	0.97	1.39	0.97	0.94
(0.77-1.48)	(0.78-1.39)	(0.66-1.88)	(0.95-1.78)	(0.74-1.31)	(0.68-1.31)
MIR	0.79	0.80	0.73	0.74	0.74
(0.72-1.06)	(0.63-1.01)	(0.54-0.94)	(0.50-1.09)	(0.53-1.01)	(0.53-1.01)
PAR	1.06	1.08	1.20	1.00	1.35
(0.87-1.29)	(0.80-1.38)	(0.78-1.59)	(0.85-1.52)	(0.76-1.25)	(0.74-1.82)
REB	1.60	1.63	1.46	1.95	1.48
(1.20-2.06)	(1.29-2.08)	(1.09-2.03)	(1.02-2.08)	(1.14-2.86)	(1.09-2.01)
SER	0.87	0.88	0.79	1.06	0.81
(0.71-1.05)	(0.72-1.07)	(0.61-1.01)	(0.69-1.17)	(0.61-1.01)	(0.60-1.11)
VEN	0.85	0.86	0.77	1.03	0.78
(0.70-1.01)	(0.71-1.05)	(0.60-0.99)	(0.58-1.24)	(0.68-0.96)	(0.59-0.99)

Cipriani, Furukawa, Salanti et al. (2009) 17

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)

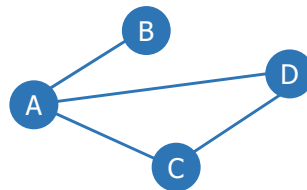
OR=odds ratio. CI=credibility interval. *p<0.05. For efficacy, OR higher than 1 favours fluoxetine. For acceptability, OR lower than 1 favours fluoxetine.

Table 4: Efficacy and acceptability using fluoxetine as reference compound

18

NMAの基本的な統計モデル

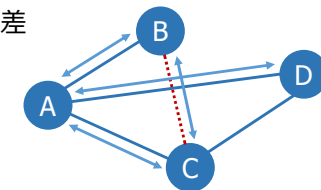
- ▶ T 種類の治療のネットワークにおいて、どのような統計モデルから治療効果の差や治療間の順位を推定する？
- ▶ 重要な概念
 - ▶ 多変量メタアナリシス
 - ▶ モデルの自由度
 - ▶ 3アーム以上の試験の取り扱い
 - ▶ Heterogeneity
 - ▶ Inconsistency



19

ネットワークのモデルを組み立てよう

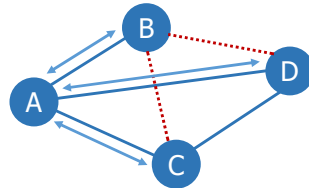
- ▶ まずは、reference となる治療を決める
- ▶ Aを基準としたペアワイズな比較の差
 - ▶ A vs. B δ_{AB}
 - ▶ A vs. C δ_{AC}
 - ▶ A vs. D δ_{AD}
- ▶ このとき、例えば、B vs. Cは...??
 - ▶ $\delta_{BC} = \delta_{AC} - \delta_{AB}$
 - ▶ A vs. B, A vs. Cの差が定めれば、自動的に決まってしまう？
- ▶ ネットワーク上のエビデンスに不整合がないなら、3つのパラメータですべての対比較の差が規定できる...??



20

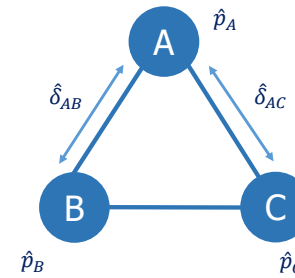
ネットワークのモデルの自由度

- ▶ T 種類の治療を比較する場合、referenceの治療を対照とした、 $T-1$ のパラメータによって、ネットワーク上のすべての対比較についての差は、Fixされる
- ▶ Aを対照とした3つのパラメータによってモデルの自由度分の制約を与えてやれば (A vs. B, C, D)、他のすべての対比較の差は、自動的にFixされる



21

Multi-Arm Trialsの関連の問題



- ▶ 3 arm trial (A vs. B vs. C) において、Odds Ratioを推定してやると...
- ▶ A vs. B
 - ▶ $\hat{\delta}_{AB} = \log(\hat{p}_B(1 - \hat{p}_A) / \{\hat{p}_A(1 - \hat{p}_B)\})$
- ▶ A vs. C
 - ▶ $\hat{\delta}_{AC} = \log(\hat{p}_C(1 - \hat{p}_A) / \{\hat{p}_A(1 - \hat{p}_C)\})$
- ▶ \hat{p}_A が共通しているため、アウトカム指標の $\hat{\delta}_{AB}$, $\hat{\delta}_{AC}$ には相関が生じることになる！！

22

関連のモデル化

- ▶ 3群以上の比較試験は、相対的には少数であるが、多くのNMAにおいて対象として含まれることが多い
- ▶ 直接・間接エビデンスを統合するためのモデルには、この相関を適切にモデル化しなくては、最終的な推定値にもバイアスが生じることになる
- ▶ Univariateなメタアナリシスのモデルではなく、相関をモデル化するためのMultivariateなメタアナリシスのモデルを用いる必要がある！！

23

メタアナリシスにおける変量効果モデル

- ▶ $Y_i \sim N(\theta_i, s_i^2)$
- ▶ $\theta_i \sim N(\mu, \tau^2)$
 - ▶ Y_i : i 番目の試験の効果の指標 (対数オッズ比など) の推定値
 - ▶ θ_i : i 番目の試験の効果の指標 (対数オッズ比など) の真値
 - ▶ s_i^2 : Y_i の分散
 - ▶ μ, τ^2 : θ_i 'sの平均 (平均治療効果) と分散 (異質性の程度を表す分散)
- ▶ θ_i を変量効果によって、試験間で異なるものとしてモデル化することによって、試験間での治療効果の異質性 (heterogeneity) を考慮した推測を行えるようにしている

DerSimonian and Laird (1986), Whitehead and Whitehead (1991) 24

Example: 変量効果モデルによる解析結果

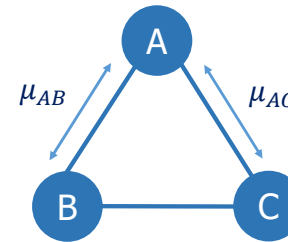
Reference	CS		HS		Weight	Risk Ratio [95% Confidence Interval]	Random-Effects Model
	Events	Total	Events	Total			
WSHAG 1991							
Craig SR 1996	13	46	13	49	15.9%	1.07 [0.55-2.05]	
Valverde A 1996	7	53	8	63	10.3%	1.04 [0.40-2.68]	
Law S. 1997	20	50	5	55	10.9%	4.40 [1.79-10.84]	
Laterza E. 1999	3	18	2	20	4.2%	1.67 [0.31-8.87]	
Walther B 2003	12	42	8	41	13.0%	1.46 [0.67-3.21]	
Hsu HH 2004	5	28	4	28	7.2%	1.25 [0.37-4.17]	
Okuyama M 2007	2	14	0	18	1.5%	6.33 [0.33-122.21]	
Luechakiettaisak P 2008	19	52	10	52	15.7%	1.90 [0.98-3.68]	
Aquino JL 2009	3	15	5	15	6.9%	0.60 [0.17-2.07]	
Zhang YS 2010	13	261	2	236	5.2%	5.88 [1.34-25.77]	
Ma RD 2010	9	47	5	52	9.2%	1.99 [0.72-5.52]	
Total	106	626	62	629	100.0%	1.67 [1.16-2.41]	

Heterogeneity: $\tau^2 = 0.11$; $\chi^2 = 14.30$, $df = 10$ ($P = 0.16$); $I^2 = 30\%$
 Test for overall effect: $Z = 2.73$ ($P = 0.006$)

Honda et al. (2013)

25

多変量メタアナリシスのモデル



- ▶ A, B, Cの3治療のネットワーク
- ▶ μ_{AB}, μ_{AC} を決めれば、 μ_{BC} も自動的に定まる
- ▶ Y_{i1}, Y_{i2} を i 番目の試験における A vs. B, A vs Cの比較の対数オッズ比とすると、 μ_{AB}, μ_{AC} の推定には、 Y_{i1}, Y_{i2} の分散と共分散（相関）を考慮したモデルを用いる必要がある

26

多変量メタアナリシスの変量効果モデル

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix}, \begin{pmatrix} s_{i11}^2 & s_{i12}^2 \\ s_{i21}^2 & s_{i22}^2 \end{pmatrix} \right)$$

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{i11}^2 & \sigma_{i12}^2 \\ \sigma_{i21}^2 & \sigma_{i22}^2 \end{pmatrix} \right)$$

- ▶ s_{i11}^2, s_{i22}^2 : i 番目の試験における Y_{i1}, Y_{i2} の分散
- ▶ s_{i12}^2, s_{i21}^2 : i 番目の試験における Y_{i1}, Y_{i2} の共分散 ($s_{i12}^2 = s_{i21}^2$)
- ▶ $s_{i11}^2, s_{i22}^2, s_{i12}^2, s_{i21}^2$ を上記のように配置した行列を、分散共分散行列という ($\sigma_{i11}^2, \sigma_{i22}^2, \sigma_{i12}^2, \sigma_{i21}^2$ についても同様)
- ▶ 共分散 s_{i12}^2 の推定値は、Trikalinos and Olkin (2008) の推定値などを用いることができる

27

ベクトル・行列を使った表記法

$$Y_i \sim N(\theta_i, S_i)$$

$$\theta_i \sim N(\mu, \Sigma)$$

$$Y_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix}, \theta_i = \begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix}, S_i = \begin{pmatrix} s_{i11}^2 & s_{i12}^2 \\ s_{i21}^2 & s_{i22}^2 \end{pmatrix}, \mu = \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{i11}^2 & \sigma_{i12}^2 \\ \sigma_{i21}^2 & \sigma_{i22}^2 \end{pmatrix}$$

- ▶ 多変量モデルを扱うときには、上記のように、ベクトル・行列の表記法を用いると便利
- ▶ 比較する治療の数 T が、3, 4, ... と増えていく場合も、コンポーネントの数が増えていくだけで、同じ表記法で表現することができる

Jackson et al. (2011), Mavridis and Salanti (2013) 28

推定されるパラメータ

- ▶ μ のコンポーネント
 - ▶ 先述の μ_{AB}, μ_{AC} に対応するパラメータ
 - ▶ 変量効果によるばらつきの中での平均治療効果を表すパラメータ（ネットワーク上のその他の対比較の平均治療効果も、すべてこれらのパラメータで規定される）
- ▶ Σ のコンポーネント
 - ▶ 試験間の治療効果の異質性を表すパラメータ
 - ▶ 対比較のメタアナリシスとは異なり、相関（共分散）のパラメータも推定しなくてはならない点に注意！！

29

推定方法①：最尤法

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

→ $Max_{(\hat{\mu}, \hat{\Sigma})}$ where $W_i = (\Sigma + S_i)^{-1}$

- ▶ 尤度関数を最大にするパラメータの値を推定値とする方法
 - ▶ ロジスティック回帰やCox回帰と同じ推定方法で、医学研究の統計解析では、最もポピュラーな方法
 - ▶ N が十分に大きいときは、最も精度の良い理想的な推定量となるが、変量効果モデルのもとでは、 N が十分に大きくないときは、一定のバイアスが生じることが知られている
- White (2011, 2015) 30

推定方法②：制限付き最尤法（REML法）

$$\ell_{RL}(\mu, \Sigma) = \ell(\mu_1, \mu_c, \eta) - \frac{1}{2} \log \left| \sum_{i=1}^N W_i \right| + \frac{1}{2} p \log 2\pi$$

→ $Max_{(\hat{\mu}, \hat{\Sigma})}$

- ▶ 最尤法によるバイアスを補正するために、パラメータ空間に制限をかけたもとの最尤法（restricted maximum likelihood; REML）
- ▶ N がmoderateである条件下では、最尤法よりも正確で良好な性質を持つ推定量であることが知られている（Verbeke and Molenberghs, 2000）
- ▶ 現状のNMAの解析では、REML法を用いるのが一般的！！

White (2011, 2015) 31

Dropしたアームの取り扱いについて

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix}, \begin{pmatrix} s_{i11}^2 & s_{i12}^2 \\ s_{i21}^2 & s_{i22}^2 \end{pmatrix} \right)$$

- ▶ A vs. B vs. Cの3群比較試験では、 Y_{i1}, Y_{i2} の両方を定義することができる
- ▶ A vs. Bの2群試験では、 Y_{i2} （A vs. C）が定義できない...??
 - ▶ $\ell_{RL}(\mu, \Sigma)$ には、 Y_{i2} をDropした、 Y_{i1} の尤度のみを組み込む
- ▶ 同様に、A vs. Cの2群試験では、 $\ell_{RL}(\mu, \Sigma)$ には、 Y_{i1} をDropした、 Y_{i2} の尤度のみを組み込む
- ▶ それでは、Aが含まれない、B vs. Cの試験のデータは、どのようにモデルに組み込めばよい？

32

Data Augmentation法

- ▶ Referenceとなるアームが含まれない試験については、フォーマルに、微小なデータを足して (augmentして)、オッズ比などの効果の指標の推定値を構成する
- ▶ 例えば、Binary outcomeのNMAであれば、A群は 0.01人のアームであり、そのうちの 0.001人がイベントを起こしたものととして、オッズ比を計算する
- ▶ B vs. Cの2群試験では、A vs. B, A vs. Cのオッズ比を組み込むことになるが、ネットワークの自由度の制約により、B vs. Cのオッズ比は自動的にFixされることになる

White et al. (2012) 33

Illustrative Example

Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis

William J Elliott, Peter M Meyer

Summary

Background The effect of different classes of antihypertensive drugs on incident diabetes mellitus is controversial because traditional meta-analyses are hindered by heterogeneity across trials and the absence of trials comparing angiotensin-converting-enzyme (ACE) inhibitors with angiotensin receptor blockers (ARB). We therefore undertook a network meta-analysis, which accounts for both direct and indirect comparisons to assess the effects of antihypertensive agents on incident diabetes.

Methods We undertook a systematic review up to Sept 15, 2006, and identified 48 randomised groups of 22 clinical trials with 143 153 participants who did not have diabetes at randomisation and so were eligible for inclusion in our analysis. 17 trials enrolled patients with hypertension, three enrolled high-risk patients, and one enrolled those with heart failure. The main outcome was the proportion of patients who developed diabetes.

Findings Initial drug therapy used in the trials (and the number of patients with diabetes of the total number at risk) included: an ARB (1189 of 14 185, or 8.38%), ACE inhibitor (1618 of 22 941, or 7.05%), calcium-channel blocker (CCB, 2791 of 38 607, or 7.23%), placebo (1686 of 24 767, or 6.81%), β blocker (2705 of 35 745, or 7.57%), or diuretic (998 of 18 699, or 5.34%). With an initial diuretic as the standard of comparison (eight groups), the degree of incoherence (a measure of how closely the entire network fits together) was small ($I^2=0.000017$, eight degrees of freedom). The odds ratios were: ARB (five groups) 0.57 (95% CI 0.46-0.72, $p<0.0001$); ACE inhibitor (eight groups) 0.67 (0.56-0.80, $p<0.0001$); CCB (nine groups) 0.75 (0.62-0.90, $p=0.002$); placebo (nine groups) 0.77 (0.63-0.94, $p=0.009$); β blocker (nine groups) 0.90 (0.75-1.09, $p=0.30$). These estimates changed little in many sensitivity analyses.

Interpretation The association of antihypertensive drugs with incident diabetes is therefore lowest for ARB and ACE inhibitors followed by CCB and placebo, β blockers and diuretics in rank order.

Lancet 2007; 369: 201-07
Department of Preventive Medicine, Rush Medical College of Rush University at Rush University Medical Center, Chicago, IL 60612, USA (P.M.) (Elliott PM)

Correspondence to: Prof William J Elliott will@rml.rush.edu

Elliott and Meyer (2007)

34

降圧薬と糖尿病の発症リスクに関するNMA

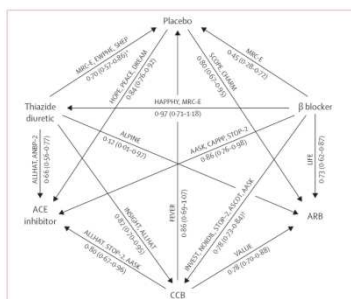


Figure 2: Network of clinical trials of antihypertensive drugs in which incident diabetes was reported. See panel for full trial names. For each pair-wise comparison, arrowhead points to class of antihypertensive drug with lower risk of incident diabetes in traditional meta-analyses. Summary odds ratio and 95% CI for comparison are shown below arrow. I^2 value for heterogeneity=0.008. I^2 value for homogeneity=0.05.

- ▶ 2006年までに実施された、左記の5種類の降圧薬 (+ プラセボ) のいずれかを比較したランダム化臨床試験のうち、糖尿病患者の発症者数を報告した試験 (22試験; 143,153人) を系統的にレビュー
- ▶ アウトカムは、糖尿病の発症の有無 (= 0,1; Binary Outcome)
- ▶ ネットワークメタアナリシスによって、糖尿病の発症リスクを系統的に比較

Elliott and Meyer (2007) 35

Summary of the 22 trials

Elliott and Meyer (2007)

Year	Duration (years)	Drug 1	New cases of diabetes/total	Drug 2	New cases of diabetes/total	Drug 3	New cases of diabetes/total	
AASK*	2006	3.8	ACE inhibitor	45/410	β blocker	70/405	CCB	32/202
ALLHAT*	2002	4.0	ACE inhibitor	119/4096	CCB	154/3954	Diuretic	302/7666
ALPINE*	2003	1.0	ARB	3/196	Diuretic	8/196	-	-
ANBP-2*	2005	4.1	ACE inhibitor	138/2800	Diuretic	200/2826	-	-
ASCOT*	2005	5.5	β blocker	799/7040	CCB	567/7072	-	-
CAPP*	1999	6.1	ACE inhibitor	33/5181	β blocker	380/5230	-	-
CHARM*	2003	-3.1	ARB	166/2725	Placebo	202/2721	-	-
DISAMP*	2006	-3.0	ACE inhibitor	449/2623	Placebo	479/1646	-	-
DWPK*	1991	4.7	Diuretic	20/416	Placebo	20/424	-	-
FEVER*	2005	3.3	CCB	177/4841	Placebo	154/4870	-	-
HAPPY*	1987	3.8	β blocker	86/3297	Diuretic	75/3272	-	-
HOPE*	2001	4.5	ACE inhibitor	102/2837	Placebo	155/4881	-	-
INSIGHT*	2000	3.0	CCB	136/2568	Diuretic	176/2511	-	-
INVEST*	2003	4.0	β blocker	665/8078	CCB	569/8098	-	-
LIFE*	2002	4.8	ARB	242/4020	β blocker	320/3979	-	-
MIRAC*	1992	5.8	β blocker	37/1102	Diuretic	43/1081	Placebo	34/2213
NORDAL*	2000	4.5	β blocker or diuretic	253/5059	CCB	316/5095	-	-
PEACE*	2004	4.8	ACE inhibitor	335/3432	Placebo	399/3472	-	-
SCORE*	2003	3.7	ARB	93/2167	Placebo	115/2125	-	-
SHEP*	1998	3.0	Diuretic	140/1611	Placebo	118/1578	-	-
STOP-2*	1999	4.0	ACE inhibitor	93/1970	β blocker or diuretic	97/1960	CCB	95/1965
VALIE*	2004	4.2	ARB	690/5087	CCB	845/5074	-	-

Table 1: Summary of clinical trials of antihypertensive drugs that reported new cases of diabetes

36

Stata: network

- ▶ 英国 MRC Biostatistics UnitのIan White博士によって開発された、ネットワークメタアナリシスの解析のための Stata の計算モジュール
- ▶ ここまで述べてきたようなネットワークメタアナリシスの解析から、後ほど述べる Inconsistencyの評価までを、簡単なコマンドで実行することができる
- ▶ 以降で使用する事例プログラム&データセットは、<http://normanh.skr.jp/materials.html> からダウンロードすることができます

White (2015) 37

データセットのフォーマット (Arm-based)

Study ID	治療の種類	アームごとのイベント数	アームごとの対象者数
1	Diuretic	49	1501
2	Placebo	34	2273
3	Beta blocker	37	1102
4	Diuretic	29	416
5	Placebo	20	424
6	Diuretic	140	1631
7	Placebo	110	1578
8	Diuretic	75	3232
9	Beta blocker	66	3297
10	Diuretic	302	6766
11	CCB	154	3664
12	ACE i/HHR	119	4096
13	Diuretic	170	2511
14	CCB	136	2928
15	Diuretic	200	2920
16	Diuretic	138	2900
17	ACE i/HHR	0	199
18	ARB	11	199
19	Diuretic	154	4670
20	CCB	177	4641
21	Placebo	489	2645
22	ACE i/HHR	448	2623
23	Diuretic	156	2583
24	Placebo	102	2637
25	ACE i/HHR		

<http://normanh.skr.jp/diabetes.csv> のデータセット

38

Stata: Setup

```
network setup d n, studyvar(study) trtvar(trt) ref(Placebo) rr
```

Treatments used

A:	ACE inhibitor
B:	ARB
C:	Beta blocker
D:	CCB
E:	Diuretic
F (reference):	Placebo

Measure Log risk ratio

Studies

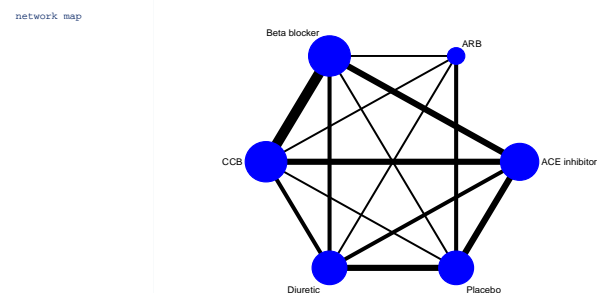
ID variable:	study
Number used:	22
IDs with zero cells:	[none]
IDs with augmented reference arm:	4 5 6 7 8 15 16 17 18 19 20 21 22
- observations added:	0.001
- mean in augmented observations:	study-specific mean

Network information

Components:	1 (connected)
D.f. for inconsistency:	13
D.f. for heterogeneity:	8

39

Stata: Network plot



ネットワーク上のノードの大きさが、治療ごとの合計のサンプルサイズ、線の太さが、直接比較の人数の合計を反映している

40

Stata: 変量効果モデルによる解析

```
network meta consistency, eform

Command is: mmeta _y_S , eform bcovariance(exch 0.5) longparm suppress(uv mm) vars(_y_A _y_B _y_C _y_D _y_E)
Note: using method reml
Note: using variables _y_A _y_B _y_C _y_D _y_E
Note: 22 observations on 5 variables
Note: variance-covariance matrix is proportional to .5*I(5)+.5*J(5,5,1)

initial:      log likelihood = -99.591901
rescale:     log likelihood = -83.136815
rescale eq:  log likelihood = -75.720162
Iteration 0:  log likelihood = -75.720162
Iteration 1:  log likelihood = -75.271504 (not concave)
Iteration 2:  log likelihood = -74.969571 (not concave)
Iteration 3:  log likelihood = -74.952587
Iteration 4:  log likelihood = -74.926246
Iteration 5:  log likelihood = -74.926188
Iteration 6:  log likelihood = -74.926188
```

デフォルトでは、REML法での解析が行われる。log likelihoodが、制限付き対数尤度の値で、収束するまで、徐々に大きくなっていく反復計算が行われている (Newton-Raphson法)

41

Stata: 変量効果モデルによる解析

```
Multivariate meta-analysis
Variance-covariance matrix = proportional .5*I(5)+.5*J(5,5,1)
Method = reml                      Number of dimensions = 5
Restricted log likelihood = -74.926188  Number of observations = 22
```

	exp(Coef)	Std. Err.	z	P> z	[95% Conf. Interval]
_y_A					
_cons	.8922365	.0550374	-1.85	0.065	.7906309 1.0069
_y_B					
_cons	.841197	.0674252	-2.16	0.031	.7189034 .984294
_y_C					
_cons	1.214073	.0902986	2.61	0.009	1.049386 1.404605
_y_D					
_cons	1.035141	.0749977	0.48	0.634	.8981084 1.193083
_y_E					
_cons	1.299829	.1010935	3.37	0.001	1.116052 1.513868

Placebo (F) をreferenceとした場合のそれぞれのリスク比のREML推定値、および、95%信頼区間・P値が出力されている

42

Stata: 変量効果モデルによる解析

```
Estimated between-studies SDs and correlation matrix:
      SD      _y_A      _y_B      _y_C      _y_D      _y_E
_y_A .10517721      1      .      .      .      .
_y_B .10517721      .5      1      .      .      .
_y_C .10517721      .5      .5      1      .      .
_y_D .10517721      .5      .5      .5      1      .
_y_E .10517721      .5      .5      .5      .5      1
mmmeta command stored as F9
```

試験間分散共分散行列の推定値。すべての相関係数は、0.50にFixされている。これは、Stata networkのデフォルト設定である。この箇所が、試験間分散の平方根 (標準偏差) を表している。± 0.10 の範囲に、2/3のデータが含まれるとすると、試験ごとの治療効果の大きさは、一定の異質性があるものと考えられる。

43

Stata: League Tableの作成

```
netleague, lab(PLA ACE ARB BET CCB DIU) sort(ACE ARB BET CCB DIU PLA) eform

The league table has been stored at the end of the dataset
```

	ARB	_BET_	_CCB_	_DIU_	_PLA_
1	ACE	0.94 (0.80,1.12)	1.36 (1.19,1.55)	1.16 (1.01,1.33)	1.46 (1.26,1.69)
2	ARB	1.44 (1.23,1.69)	1.23 (1.06,1.43)	1.55 (1.28,1.86)	1.19 (1.02,1.39)
3	BET	0.85 (0.77,0.95)	1.07 (0.92,1.24)	0.82 (0.71,0.95)	
4	ARB	0.81 (0.70,0.95)	1.17 (1.05,1.30)	0.88 (0.75,1.03)	1.26 (1.09,1.45)
5	BET	0.85 (0.77,0.95)	1.07 (0.92,1.24)	0.82 (0.71,0.95)	
6	ARB	0.81 (0.70,0.95)	1.17 (1.05,1.30)	0.88 (0.75,1.03)	1.26 (1.09,1.45)
7	BET	0.85 (0.77,0.95)	1.07 (0.92,1.24)	0.82 (0.71,0.95)	
8	ARB	0.81 (0.70,0.95)	1.17 (1.05,1.30)	0.88 (0.75,1.03)	1.26 (1.09,1.45)
9	BET	0.85 (0.77,0.95)	1.07 (0.92,1.24)	0.82 (0.71,0.95)	
10	ARB	0.81 (0.70,0.95)	1.17 (1.05,1.30)	0.88 (0.75,1.03)	1.26 (1.09,1.45)
11	BET	0.85 (0.77,0.95)	1.07 (0.92,1.24)	0.82 (0.71,0.95)	
12	ARB	0.81 (0.70,0.95)	1.17 (1.05,1.30)	0.88 (0.75,1.03)	1.26 (1.09,1.45)

44

Stata: 順位の確率の推定

```
network rank min, line meanrank xlabel(1/4) seed(37195) reps(5000)
```

Command is: `mvmeta, noest pbest(min in 1, zero id(study) line meanrank xlabel(1/4) seed(37195) reps(5000) striprefix(_y_) zeroname(F) rename(A = ACE inhibitor, B = ARB, C = Beta blocker, D = CCB, E = Diuretic > ic, F = Placebo))`

Option line specified -> option all assumed

Estimated probabilities (%) of each treatment being the best (and other ranks)
 - assuming the minimum parameter is the best
 - using 5000 draws
 - allowing for parameter uncertainty

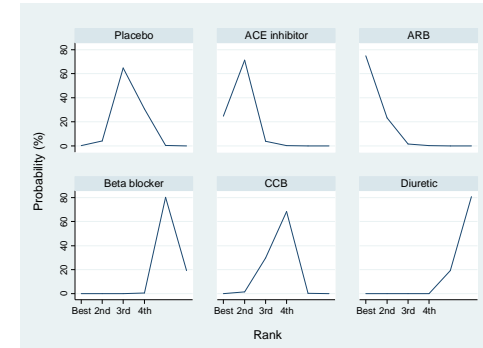
それぞれの治療が「何位」になるかの「確率」を推定することができる

study and Rank		Treatment					
		Placebo	ACE inhibitor	ARB	Beta blocker	CCB	Diuretic
1	Best	0.2	24.7	75.0	0.0	0.0	0.0
	2nd	4.0	71.2	23.3	0.0	1.5	0.0
	3rd	64.7	3.9	1.6	0.0	29.8	0.0
	4th	30.8	0.2	0.1	0.6	68.2	0.1
	5th	0.3	0.0	0.0	80.1	0.4	19.2
	Worst	0.0	0.0	0.0	19.3	0.0	80.7
	MEAN RANK	3.3	1.8	1.3	5.2	3.7	5.8
	SUCRA	0.5	0.8	0.9	0.2	0.5	0.0

mvmeta command is stored in F9

45

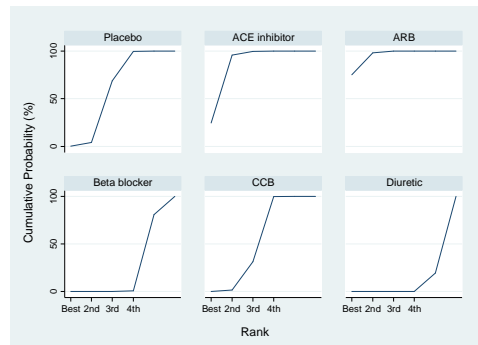
順位の確率の推定値



46

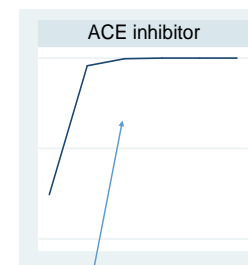
累積確率曲線とSUCRA

```
network rank min, line cumulative meanrank xlabel(1/4) seed(37195) reps(5000)
```



47

SUCRA



累積確率曲線の下側領域の面積

- ▶ Surface Under the Cumulative Ranking
- ▶ 順位の確率の累積値の曲線の下側領域の面積をとった指標
- ▶ 「各順位の確率」は、直感的にはわかりやすい指標であるが、微小なデータの変化により、大きく値が変わる不安定な指標 (Misleadingな指標となってしまうことも)
- ▶ 治療ごとの相対的な位置づけを評価する上で、より不確実性を考慮した指標としては、SUCRAのほうが好ましいとも言われている

Salanti et al. (2011) 48

Stata: SUCRA

```
network rank min, line meanrank xlabel(1/4) seed(37195) reps(5000)

Command is: mvmeta, noest pbest(min in 1, zero id(study) line meanrank xlabel(1/4) seed(37195) reps(5000)
striprefix(_y_) zeroname(F) rename(A = ACE inhibitor, B = ARB, C = Beta blocker, D = CCB, E = Diuret
> ic, F = Placebo))
Option line specified -> option all assumed

Estimated probabilities (%) of each treatment being the best (and other ranks)
- assuming the minimum parameter is the best
- using 5000 draws
- allowing for parameter uncertainty
```

study and Rank		Treatment					
		Placebo	ACE inhibitor	ARB	Beta blocker	CCB	Diuretic
1	Best	0.2	24.7	75.0	0.0	0.0	0.0
	2nd	4.0	71.2	23.3	0.0	1.5	0.0
	3rd	64.7	3.9	1.6	0.0	29.8	0.0
	4th	30.8	0.2	0.1	0.6	68.2	0.1
	5th	0.3	0.0	0.0	80.1	0.4	19.2
	Worst	0.0	0.0	0.0	19.3	0.0	80.7
	MEAN RANK	3.3	1.8	1.3	5.2	3.7	5.8
	SUCRA	0.5	0.8	0.9	0.2	0.5	0.0

mvmeta command is stored in F9

49

補足：ベイズ統計の手法による解析



Thomas Bayes (1701-1761)

- ▶ ベイズ統計の手法を用いて、同様の解析を行うことができる
- ▶ 初期の頃の研究は、こちらのほうが多く使われていた (Cipriani et al. (2009), Elliot and Meyer (2007) など)
- ▶ Multi-arm trialsの相関のモデル化などは、複雑なマルチレベルモデルを扱やすいベイズ統計の手法が得意であったという背景も
- ▶ Rのgemtc packageなどを使えば、ベイズモデルによる解析も、JAGSを用いて比較的容易に実行することができる

50

三竦み (さんすくみ)

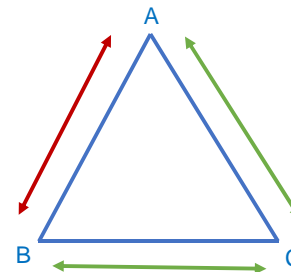


- ▶ 蝗虫食蛇、蛇食蛙、蛙食蝗虫、互相食也

《「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。

51

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)

52

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！！

53

Local inconsistency on the network

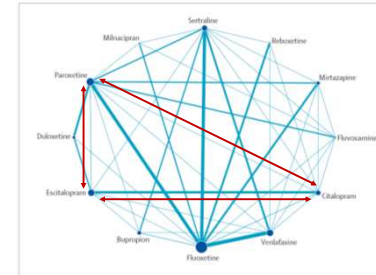
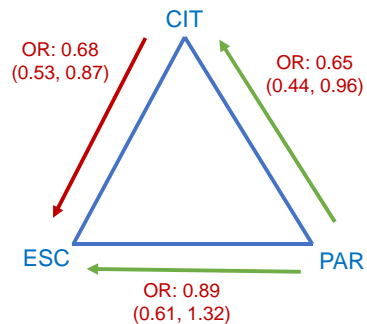


Figure 2: Network of eligible comparisons for the multiple-treatment meta-analysis for efficacy (response rate). The width of the lines is proportional to the number of trials comparing each pair of treatments, and the size of each node is proportional to the number of randomized participants (sample 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を仮定しているため、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) 54

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「三棘」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) 55

Inconsistencyの評価の方法

- ▶ Local Inconsistencyの検定
- ▶ Global Inconsistencyの検定
- ▶ Side-splitting法

56

Local Inconsistencyの検定

- ▶ ネットワーク上の部分的なTriangle loopsに対して、直接比較のパスと間接比較のパスのEffect Sizeが一致するかどうかを検定する

$$Z = \frac{\delta_{Direct} - \delta_{Indirect}}{\sqrt{V(\delta_{Direct}) + V(\delta_{Indirect})}} \sim N(0,1)$$

- ▶ 有意になったTriangle Loopは、Inconsistencyが疑われるので、バイアスを生じさせる要因がないかどうかを詳細に検討する必要がある

Bucher et al. (1999), Chaimani et al. (2013) 57

Stata: Local Inconsistency Tests

```
ifplot _y _stderr _t1 _t2 study, tau2(0.011)
```

* 16 triangular loops found

Loop	IF	seIF	z_value	p_value	CI_95
B-C-E	1.808	1.067	1.695	0.090	(0.00, 3.90)
B-D-E	1.682	1.059	1.587	0.112	(0.00, 3.76)
B-E-F	1.492	1.064	1.403	0.161	(0.00, 3.58)
A-C-F	0.801	0.248	3.234	0.001	(0.32, 1.29)
B-C-F	0.703	0.262	2.684	0.007	(0.19, 1.22)
C-E-F	0.571	0.290	1.970	0.049	(0.00, 1.14)
C-D-F	0.447	0.261	1.709	0.087	(0.00, 0.96)
A-C-D	0.277	0.113	2.447	0.014	(0.06, 0.50)
A-C-E	0.251	0.162	1.552	0.121	(0.00, 0.57)
A-E-F	0.162	0.134	1.205	0.228	(0.00, 0.42)
B-D-F	0.151	0.144	1.047	0.295	(0.00, 0.43)
A-D-F	0.116	0.145	0.798	0.425	(0.00, 0.40)
B-C-D	0.106	0.102	1.034	0.301	(0.00, 0.31)
A-D-E	0.058	0.159	0.365	0.715	(0.00, 0.37)
D-E-F	0.038	0.165	0.233	0.816	(0.00, 0.36)
C-D-E	0.020	0.152	0.134	0.893	(0.00, 0.32)

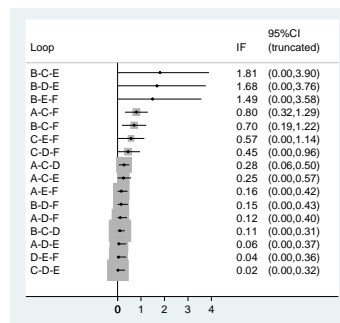
ネットワーク上のすべてのTriangle Loopsをすべてリストアップし、Local Inconsistency Testsを行う。

IF (Inconsistency Factor) は、ループ上の3辺の特定のパスにおける効果サイズの差 $\delta_{Direct} - \delta_{Indirect}$ をとったもの。(P値は、これが0かどうかの検定による)

4/16が有意 (P<0.05) になっている。有意になったループは要検討！！

58

Stata: Inconsistency plot



ifplot関数から出力される図。前頁の表におけるIFを視覚的にプロットしたもの。

59

Global Inconsistencyの検定

- ▶ Local Inconsistency Testは、直感的であり、わかりやすい評価方法であるが、ループごとに情報を限定してしまうため、検出力は相対的に低い
- ▶ また、メインの解析モデル全体におけるConsistencyの仮定が成り立っているかどうかのほうに関心の対象である
- ▶ ネットワーク全体の情報を用いることによって、より検出力が高いことも期待できる、包括的なGlobal Inconsistencyの検定を行うには？

60

Global Inconsistencyの検定を構成しよう

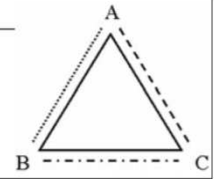
- ▶ 帰無仮説：ネットワーク上にInconsistentなパスは1つもない
- ▶ 対立仮説：ネットワーク上にInconsistentなパスが少なくとも1つはある
- ▶ 帰無仮説に対応するConsistencyを仮定したモデルと、ネットワーク上に想定されるすべてのInconsistentな要因をすべて考慮した「最大のInconsistencyモデル」を比較する適合度検定（尤度比検定による）によって、Global Inconsistencyを評価することができる

61

最大のInconsistencyモデルとは？

Table 1. Design-by-treatment interaction model for three treatments (all possible designs). Heterogeneity terms have been omitted.

Design	A	B	C
ABC	Ref	δ^{AB}	δ^{AC}
AB	Ref	$\delta^{AB} + \omega_2^{AB}$	—
AC	Ref	—	$\delta^{AC} + \omega_3^{AC}$
BC	Ref	δ^{AB}	$\delta^{AC} + \omega_4^{AC}$



- ▶ ネットワーク上のInconsistencyは、ネットワーク上のデザインの違いによる共通のパス上の治療効果の差（交互作用）で、必要十分に表現することができる？

Higgins et al. (2012) 62

Jackson-Boddington-Whiteの定理

- ▶ ネットワークメタアナリシスにおける、すべての可能なInconsistency (all possible loop inconsistency) は、前頁のDesign-by-treatment interactionモデルによって表現することができる
Jackson, Boddington and White (2016)
- ▶ 多くの教科書では、「直接・間接エビデンスの不一致」と解説されているが、Inconsistencyの正体は、試験デザインの違いによる治療効果の系統的な差である（直接・間接エビデンスの不一致とは、Triangle loop上では、そのように説明ができる場合もあるという一例に過ぎなかった）
- ▶ Global Inconsistency検定では、すべてのDesign-by-treatment interactionをモデル化したMeta-regressionモデルを「最大のInconsistencyモデル」として適合度検定を行えばよい

63

Stata: Global Inconsistency Test

```
network meta inconsistency, eform

Command is: mvmeta _y_S , eform bscovariance(exch 0.5) longparm suppress(uv mm) eq(_y_B: des_BD des_BE des_BF, _y_C:
des_ACD des_CEF, _y_D: des_ADE des_CD des_DF, _y_E: des_AE des_CE des_CEF des_DE des_EF) vars(_y_A _
> y_B _y_C _y_D _y_E)
Note: using method reml
Note: regressing _y_A on (nothing)
Note: regressing _y_B on des_BD des_BE des_BF
Note: regressing _y_C on des_ACD des_CEF
Note: regressing _y_D on des_ADE des_CD des_DF
Note: regressing _y_E on des_AE des_CE des_CEF des_DE des_EF
Note: 22 observations on 5 variables
Note: variance-covariance matrix is proportional to .5*I(5)+.5*J(5,5,1)

initial:      log likelihood = -84.483086
rescale:      log likelihood = -80.796424
rescale eq:   log likelihood = -73.624709
Iteration 0:  log likelihood = -73.624709
Iteration 1:  log likelihood = -73.417611
Iteration 2:  log likelihood = -73.363835
Iteration 3:  log likelihood = -73.305553
Iteration 4:  log likelihood = -73.29981
Iteration 5:  log likelihood = -73.299449
Iteration 6:  log likelihood = -73.299448
```

Higgins et al. (2012) の要領で、ネットワーク上のすべての可能なDesign-by-treatment interactionを検出している

64

Stata: Global Inconsistency Test

```
Multivariate meta-analysis
Variance-covariance matrix = proportional .5*I(5)+.5*J(5,5,1)
Method = reml                      Number of dimensions = 5
Restricted log likelihood = -73.299448  Number of observations = 22
```

	exp(Coef)	Std. Err.	z	P> z	[95% Conf. Interval]
<hr/>					
._y_A					
_cons	.837139	.0618998	-2.40	0.016	.7241991 .967692
<hr/>					
._y_B					
des_BD	1.129723	.277225	0.50	0.619	.6983837 1.82747
des_BE	.2295892	.2487501	-1.36	0.174	.0274607 1.919516
des_BF	1.156663	.2498866	0.67	0.501	.7573772 1.766449
_cons	.7002361	.1316256	-1.90	0.058	.4844416 1.012156
<hr/>					
._y_C					
des_ACD	1.108445	.1955309	0.58	0.559	.7844429 1.566271
des_CEF	2.3361	.6756794	2.93	0.003	1.325242 4.118014
_cons	.9354728	.1310645	-0.48	0.634	.7108427 1.231087
<hr/>					
._y_D					
des_ADE	1.155471	.2382387	0.70	0.483	.7713576 1.730862
des_CD	.7657992	.1485723	-1.38	0.169	.5235687 1.120098
des_DF	1.190449	.2523924	0.82	0.411	.785677 1.803756
_cons	.9712599	.1513154	-0.19	0.852	.715689 1.318095

des_XXという変数が、デザインによる
交互作用の大きさを表すパラメータ
これの真値が、すべて0であれば、
Consistencyの仮定が正しいことに

65

Stata: Global Inconsistency Test

```
-----+-----
```

._y_E						
des_AE	.9346495	.1886806	-0.33	0.738	.6292368	1.3883
des_CE	.6391667	.1663452	-1.72	0.085	.3837834	1.064491
des_CEF	2.013074	.5903648	2.39	0.017	1.1331115	3.576394
des_DE	.9761228	.2403313	-0.10	0.922	.6024625	1.581535
des_EF	.9407282	.1990535	-0.29	0.773	.6213755	1.424211
_cons	1.286132	.2063327	1.57	0.117	.9391354	1.761339

```
-----+-----
```

Estimated between-studies SDs and correlation matrix:

	SD	._y_A	._y_B	._y_C	._y_D	._y_E
._y_A	.09459044	1
._y_B	.09459044	.5	1	.	.	.
._y_C	.09459044	.5	.5	1	.	.
._y_D	.09459044	.5	.5	.5	1	.
._y_E	.09459044	.5	.5	.5	.5	1

66

Stata: Global Inconsistency Test

```
Testing for inconsistency:
( 1) [_y_C]des_ACD = 0
( 2) [_y_D]des_ADE = 0
( 3) [_y_E]des_AE = 0
( 4) [_y_B]des_BD = 0
( 5) [_y_B]des_BE = 0
( 6) [_y_B]des_BF = 0
( 7) [_y_D]des_CD = 0
( 8) [_y_E]des_CE = 0
( 9) [_y_C]des_CEF = 0
(10) [_y_E]des_CEF = 0
(11) [_y_E]des_DE = 0
(12) [_y_D]des_DF = 0
(13) [_y_E]des_EF = 0

      chi2( 13) = 20.08
      Prob > chi2 = 0.0931
mmmeta command stored as P9; test command stored as P8
```

「すべてのdes_XXの変数が0」という帰無仮説に対する検定が、
Global Inconsistency検定に！！

67

Side-splitting法

- ▶ それでも、ネットワーク上の対比較ごとに、直接・間接エビデンスが一致するかどうかは関心の対象である
 - ▶ Local Inconsistency Testと同様、ネットワーク上のどのパスにInconsistencyの原因があるかを調べるためにも、重要
- ▶ ネットワーク上の対比較のパス (e.g., A vs. B) ごとに、治療効果の差の推定値を、直接・間接エビデンスに分離して、要約する方法が Side-splitting法である

Dias et al. (2010), White (2015) 68

Stata: Side-splitting法

```

network sidesplit all

Side Direct Indirect Difference
   Coef. Std. Err. Coef. Std. Err. Coef. Std. Err. P>|z|
A F    .1813452   .0780005   -.0064712   .103432   .1978164   .1302274   0.149
A C    .1513204   .0791164   .4577628   .0822366   -.3064425   .1129921   0.007
A D    .212802   .1055911   .1045301   .0884521   .1082719   .1377931   0.432
A E    .3879942   .1084953   .3693753   .1067656   .0186188   .1521513   0.903
B F    .2107558   .1149284   .1300889   .1205917   .0806669   .1665788   0.628
B C    .2896346   .1384078   .4112536   .1033414   -.121619   .1727315   0.481
B D    .2052038   .1242378   .2149158   .1101001   -.009712   .1660032   0.953
B E    2.079436   1.061005   .4220981   .0941393   1.657338   1.065173   0.120
C F    -.768696   .2534753   .1431706   .07303   -.6255254   .2637964   0.018
C D    -.1962854   .0605229   .0656189   .0991703   -.1306665   .1183602   0.270
C E    -.0081028   .1502878   .0962753   .0906343   -.1043781   .1752656   0.551
D F    -.1451698   .1496263   -.0026388   .0818946   -.142531   .1705718   0.403
D E    .1965023   .1076113   .256962   .1040179   -.0604597   .1493121   0.686
E F    -.3766222   .1274141   .1825192   .1000045   -.184103   .162119   0.256
    
```

↑ 直接エビデンスを要約した結果 ↑ 間接エビデンスを要約した結果 ↑ Inconsistency検定のP値

69

Cipriani et al. (2009) の事例では...

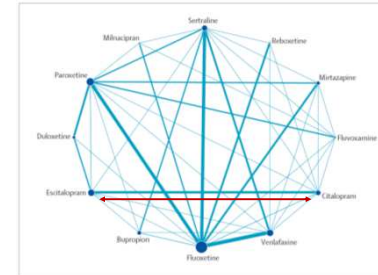


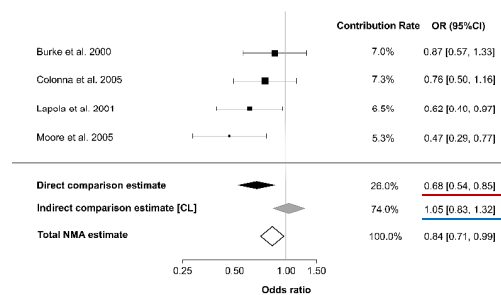
Figure 2. Network of eligible comparisons for the multiple-treatment meta-analysis for efficacy (response rate). The width of the lines is proportional to the number of trials comparing each pair of treatments, and the size of each node is proportional to the number of randomized participants (sample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

Cipriani, Furukawa, Salanti et al. (2009)

- ▶ Noma et al. (2016) では、Side-splitting式の間接エビデンスの要約を、複合尤度法という別なアプローチで行う方法を開発し、Inconsistencyを評価するための検定手法を与えた
- ▶ Cipriani et al. (2009) の事例において、ESC-CITのパスで評価をしてみると...

70

ESC-CITの直接・間接エビデンスの比較



P=0.009

※ ESCは、同じ製薬企業が開発した、CITのDerivative Drugである。

Noma et al. (2016) 71

まとめ

- ▶ 近年のComparative Effectiveness Researchの振興からも、ネットワークメタアナリシスは、システムティックレビューのスタンダードな方法のひとつとして、今後、ますます普及していくことが予想される
- ▶ 方法論そのものは魅力的であるが、強い理論的仮定に依存する方法であり、さまざまなバイアスが含まれるリスクがある
- ▶ PRISMA-NMA extended version (Hutton et al., 2015) も公表されており、NMA特有の問題に配慮して、結果の報告を行う必要がある

72

文献

- ▶ Bucher, H. C., Guyatt, G. H., Griffith, L. E., and Walter, S. D. (1997). The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology* 50, 683-691.
- ▶ 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.
- ▶ Chaimani, A., Higgins, J. P., Mavridis, D., Spyridonos, P., and Salanti, G. (2013). Graphical tools for network meta-analysis in STATA. *PLoS One* 8, e76654.
- ▶ 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 Laird, N. M. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* 7, 177-188.
- ▶ Dias, S., Welton, N. J., Caldwell, D. M., and Ades, A. E. (2010). Checking consistency in mixed treatment comparison meta-analysis. *Statistics in Medicine* 29, 932-944.
- ▶ 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.

73

- ▶ Elliott, W. J., and Meyer, P. M. (2007). Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. *Lancet* 369, 201-207.
- ▶ Higgins, J. P., Jackson, D., Barrett, J. K., Lu, G., Ades, A. E., and White, I. R. (2012). Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Research Synthesis Methods* 3, 98-110.
- ▶ Honda, M., Kuriyama, A., Noma, H., Nunobe, S., and Furukawa, T. A. (2013). Hand-sewn versus mechanical esophagogastric anastomosis after esophagectomy: a systematic review and meta-analysis. *Annals of Surgery* 257, 238-248.
- ▶ Hutton, B., Salanti, G., Caldwell, D. M., et al. (2015). The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Annals of Internal Medicine* 162, 777-784.
- ▶ Jackson, D., Boddington, P., and White, I. R. (2016). The design-by-treatment interaction model: a unifying framework for modelling loop inconsistency in network meta-analysis. *Research Synthesis Methods* 7, 329-332.

74

- ▶ Jackson, D., Riley, R., and White, I. R. (2011). Multivariate meta-analysis: potential and promise. *Statistics in Medicine* 30, 2481-2498.
- ▶ Mavridis, D., and Salanti, G. (2013). A practical introduction to multivariate meta-analysis. *Statistical Methods in Medical Research* 22, 133-158.
- ▶ Mitka, M. (2010). US government kicks off program for comparative effectiveness research. *Journal of the American Medical Association* 304, 2230-2231.
- ▶ Montgomery, S., Hansen, T., and Kasper, S. (2011). Efficacy of escitalopram compared to citalopram: a meta-analysis. *International Journal of Neuropsychopharmacology* 14, 261-268.
- ▶ Noma, H., Tanaka, S., Matsui, S., Cipriani, A., and Furukawa, T. A. (2016). Quantifying indirect evidence in network meta-analysis. *Statistics in Medicine*, DOI: 10.1002/sim.7187.
- ▶ 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.
- ▶ Salanti, G., Ades, A. E., and Ioannidis, J. P. (2011). Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal of Clinical Epidemiology* 64, 163-171.

75

- ▶ Trikalinos, T., and Olkin, I. (2008). A method for the meta-analysis of mutually exclusive binary outcomes. *Statistics in Medicine* 27, 4279-4300.
- ▶ Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer: New York.
- ▶ White, I. R. (2011). Multivariate random-effects meta-regression: Updates to mvmeta. *The Stata Journal* 11, 255-270.
- ▶ White, I. R. (2015). Network meta-analysis. *The Stata Journal* 15, 951-985.
- ▶ 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.
- ▶ Whitehead, A., and Whitehead, J. (1991). A general parametric approach to the meta-analysis of randomised clinical trials. *Statistics in Medicine* 10, 1665-1677.

76