

Multiple Treatment Meta-Analysis

II

Sofia Dias
University of Bristol
s.dias@bristol.ac.uk

SMG Training Course, March 2010, Cardiff
With thanks to: Georgia Salanti, Nicky Welton, Tony Ades, Debbi Caldwell, Alex Sutton

Overview

- Bayesian pairwise meta-analysis
- Extension to multiple treatments
 - Consistency assumptions
- Measures of model fit and model comparison
- Inconsistency models
 - How many inconsistencies?
 - how direct and indirect evidence combine
 - graphical/statistical outputs (p-values)
- Further reading and possible extensions

2

Model

- Likelihood: will depend on type of outcome
 - Normal for log-OR, log-RR, Risk-diff, mean, mean change from baseline, mean-diff, log-HR
 - Binomial for no. events/total
 - Poisson for no. events given person years at risk
- Scale for model: will depend on likelihood
 - Normal likelihood, pooled effect on natural scale
 - Binomial likelihood, pooled effect on logit scale (logistic regression)
 - Poisson likelihood, pooled effect on log scale (log-linear model)
- Arm-based summaries will estimate a baseline effect plus a relative effect
 - E.g. $\log\text{-odds} = \text{baseline} + \text{relative effect}$

3

Computation

- Using Markov Chain Monte Carlo
- Straightforward in WinBUGS 1.4.3
- Some Statistical knowledge recommended
 - Probably true for all Meta-analyses anyway!

4

A quick overview of

BAYESIAN PAIRWISE META-ANALYSIS

Generic Fixed Effect Model

$$Y_i \sim Normal(\mu, V_i)$$

- Y_i is the observed effect in study i with variance V_i
- All studies assumed to be estimating the same underlying effect size μ
 - Statistical Homogeneity
- For a Bayesian analysis, a prior distribution must be specified for μ , for example on $\ln(\text{OR})$ scale,

$$\mu \sim Normal(0, 10^5)$$

6

Appendix 1: Choice of Prior for μ

- Often amount of information in studies would overwhelm any reasonable prior - therefore choice not critical
- A priori we would be 95% certain that true value of μ is between $(0 - 1.96 \times 316 \text{ and } 0 + 1.96 \times 316)^*$
- On an odds ratio scale that is equivalent to $(10^{-269} \text{ to } 10^{269})$
i.e. very vague and essentially flat over the realistic range of interest
- Could, of course, include informative priors...

*Note: $316 = \sqrt{10^5}$

7

Generic Random Effects Model

$$Y_i \sim Normal(\theta_i, V_i)$$

- Y_i is the observed effect in study i with variance V_i
- Across studies $\theta_i \sim Normal(\mu, \tau^2)$
- prior distribution for μ , as before

$$\mu \sim Normal(0, 10^5)$$

- Prior distribution for τ : Uniform(0,10) or half-normal
 - Requires care when evidence sparse (Lambert et al, SiM 2005)

8

Some Advantages of Bayesian MA

- Can cope with zero cells
- Incorporates uncertainty in the heterogeneity parameter
- Easily extended to incorporate covariates
- Predictive distributions straightforward
- Can include informative prior distributions for eg heterogeneity parameter, when evidence sparse
- Normality of true effects in a random-effects analysis
 - Can be easily relaxed in WinBUGS to eg t-distribution

9

An extension to

MULTIPLE TREATMENT META-ANALYSIS

(MIXED TREATMENT COMPARISONS, NETWORK META-ANALYSIS)

Assumptions

- *Appropriate modelling of data (as before)*
 - Likelihood and link function
- *Comparability of studies*
 - exchangeability in all aspects other than particular treatment comparison being made
- *Equal heterogeneity (RE variance) in each comparison*
 - not strictly necessary (Lu and Ades, Biostatistics 2009)

11

Bayesian Multiple Treatment Meta-analysis

1. Four treatments A, B, C, D
2. Take treatment A as the reference treatment
 - makes no difference to relative effects, but aids interpretation if eg. placebo is chosen
3. Then the treatment effects (eg. log odds ratios) of B, C, D relative to A are the **basic** parameters
4. Given them priors:

$$\mu_{AB}, \mu_{AC}, \mu_{AD} \sim N(0, 10^5)$$

12

Functional parameters in MTC

The remaining contrasts are **functional** parameters

$$\left. \begin{aligned} \mu_{BC} &= \mu_{AC} - \mu_{AB} \\ \mu_{BD} &= \mu_{AD} - \mu_{AB} \\ \mu_{CD} &= \mu_{AD} - \mu_{AC} \end{aligned} \right\} \text{CONSISTENCY assumption}$$



- Any information on functional parameters tells us indirectly about basic parameters
 - There is a degree of redundancy in the network
- Either FE or RE model satisfying these conditions

Consistency

- We assume that the treatment effect μ_{BC} estimated by BC trials, **would be the same** as the treatment effect estimated by the AC and AB trials if they had included B and C arms
- Assume that trial arms are missing at random
 - reason they are missing is not related to treatment effect

Generic random effects Model

$$Y_i \sim \text{Normal}(\theta_i, V_i)$$

$$\theta_i \sim \text{Normal}(\mu_{bk}, \tau^2)$$

Generic random effects Model

$$Y_i \sim \text{Normal}(\theta_i, V_i)$$

$$\theta_i \sim \text{Normal}(\underbrace{\mu_{Ak} - \mu_{Ab}}_{\text{Consistency assumptions}}, \tau^2)$$

- So trial of BvsC will have $k=C$ and $b=B$
- For a Bayesian analysis, prior distributions are required for τ^2 and all basic parameters μ_{Aj} .
- Models which do not assume common heterogeneity are available (Lu and Ades *Biostatistics*, 2009)

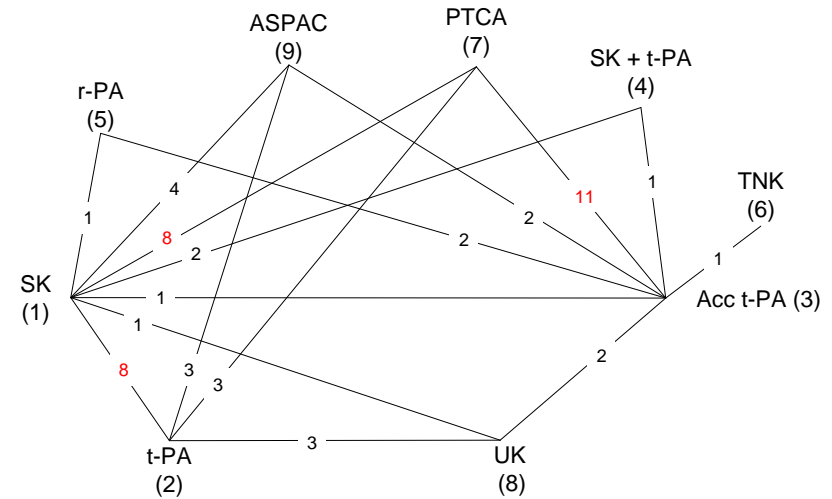
Example: Treatment for acute myocardial infarction*

- 8 thrombolytic drugs and surgery
- 9 treatments, 50 trials
- Two very large 3-arm trials
- 16 direct comparisons (out of 36)

*see eg Dias *et al* SiM in press, for details

17

Thrombo: Treatment Network



18

Question:

- In a network with 9 different treatments how many basic parameters?

19

FE Model

- $i=1, \dots, 50$ trials; $k=1, 2, 3$ arm number

- Likelihood

$$r_{ik} \sim \text{Binomial}(\pi_{ik}, n_{ik})$$

- Link function (scale)

$$\text{logit}(\pi_{ik}) = \eta_i + (\underbrace{\mu_{1t_k} - \mu_{1t_1}}_{\text{Consistency assumptions}}) I_{k \neq 1}$$

- priors

$$\eta_i \sim \text{Normal}(0, 10^5)$$

$$\mu_{1j} \sim \text{Normal}(0, 10^5), j = 2, \dots, 9 \text{ treatments}$$

Consistency assumptions

Nuisance parameters

20

Thrombo: log-odds ratios (FE model)

Treat		No of studies	Pairwise MA		MTC	
X	Y		$\hat{\mu}_{XY}^{Pair}$	var	$\hat{\mu}_{XY}^{MTC}$	var
1	2	8	-0.004	0.001	0.002	0.001
1	3	1	-0.159	0.002	-0.178	0.002
1	5	1	-0.060	0.008	-0.124	0.004
1	7	8	-0.665	0.034	-0.476	0.010
1	8	1	-0.369	0.269	-0.202	0.049
1	9	4	-0.006	0.002	0.016	0.001
2	7	3	-0.543	0.174	-0.478	0.011
2	8	3	-0.294	0.120	-0.205	0.049
2	9	3	0.017	0.002	0.014	0.001
3	4	1	0.113	0.003	0.128	0.003
3	5	2	0.019	0.004	0.054	0.003
3	7	11	-0.215	0.014	-0.298	0.010
3	8	2	0.144	0.127	-0.025	0.049
3	9	2	1.407	0.173	0.194	0.003

21

RE Model

- $i=1, \dots, 50$ trials; $k=1, 2, 3$ arm number

- Likelihood $r_{ik} \sim \text{Binomial}(\pi_{ik}, n_{ik})$

- Link function

$$\text{logit}(\pi_{ik}) = \eta_i + \theta_{ik} I_{k \neq 1}$$

- RE distribution $\theta_{ik} \sim \text{Normal}(\underbrace{\mu_{1t_k} - \mu_{1t_1}}_{\text{Consistency assumptions}}, \tau^2)$

- priors

$$\eta_i \sim \text{Normal}(0, 10^5)$$

$$\mu_{1j} \sim \text{Normal}(0, 10^5), j = 2, \dots, 9 \text{ treatments}$$

$$\tau \sim \text{Unif}(0, 10)$$

22

FE or RE model?

- Is heterogeneity always present?
- For a well defined population and decision problem, there may be little heterogeneity
 - Or this may be explained by covariates
 - Is this ever the case in “lumped” Cochrane Reviews?
- Outputs from RE model harder to interpret
- Problems with estimation of variance of RE distribution when data sparse
- FE model preferable if it can be justified...

23

FE or RE model?

- Choose between two models
- To assess model fit calculate residual deviance
 - Compare to number of unconstrained data points
- For model comparison use DIC
 - Penalises a better fit by the effective number of parameters, pD

24

Residual Deviance

- The best fit we can get is where the model predictions equal the observed data
 - Saturated model
- Residual deviance is the deviance for the **current model**, minus the deviance for a **saturated model**

$$D_{res} = -2(\text{loglik}_{model} - \text{loglik}_{sat})$$

- Calculated at each iteration of MCMC algorithm
- Summarised by posterior mean \bar{D}_{res}
- If the model is an adequate fit, we expect \bar{D}_{res} to be roughly equal to the number of unconstrained data points

25

Appendix 2: Calculating \bar{D}_{res}

- At each iteration, the residual deviance, D_{res} , is calculated as the sum of the deviances for each data point, eg for Binomial

$$D_{res} = \sum_i 2 \left(r_i \log \left(\frac{r_i}{\hat{r}_i} \right) + (n_i - r_i) \log \left(\frac{n_i - r_i}{n_i - \hat{r}_i} \right) \right)$$

$$= \sum_i \text{dev}_i$$

- r_i = observed no. events
- $\hat{r}_i = p_i n_i$ = expected no. events from current model
- dev_i is the deviance residual
- Summarised by the posterior mean \bar{D}_{res} (over M iterations)

26

Model Comparison

- Deviance Information Criteria (DIC)
 - Take deviance for current model (= $-2 \times \text{loglik}$ for current model)
 - Penalise by effective no of parameters

$$DIC = \bar{D}_{model} + p_D$$

- Extension of Akaike's Information Criterion
- Trade-off between fit and complexity
- Differences of 5 (?3) are important
- Can also use posterior mean of residual deviance \bar{D}_{res} (differs only by a constant – does not matter for comparisons)

Effective No. Parameters, p_D

- Fixed Effects Model
 - p_D = no. parameters
- Random Effects Model
 - p_D depends on between study variance, τ^2
 - For τ^2 close to 0, $\theta_i = \mu$; 1 parameter (as in fixed effects model)
 - For very large τ^2 , $\theta_i = \theta_i$; one parameter for each study

Appendix 3: Calculating p_D

- At each iteration, calculate

$$D_{res} = \sum_i dev_i$$

- For each data point, posterior mean of $dev_i = \bar{D}_i$ (mean taken over M iterations)
- Calculate posterior mean of fitted values, eg in Binomial $\hat{\alpha}_i$ is the posterior mean of α_i .
- Calculate deviance at the posterior mean of the fitted values \tilde{D}_i (replace \hat{r}_i with \tilde{r}_i in formula for residual deviance)

29

Appendix 3 (cont): Calculating p_D

- The effective number of parameters p_D is calculated as the sum, over all data points, of the leverages, i.e. the sum of the posterior mean of the residual deviances, minus the deviances at the posterior mean of the fitted values

$$pD = \sum_i leverage_i = \sum_i [\bar{D}_i - \tilde{D}_i]$$

30

How many parameters in this example?

Fixed effects model ($\tau^2 = 0$)

η_i , 50 studies

μ_{1k} , fixed treatment effects for 8 basic parameters

Random effects model ($\tau^2 > 0$)

η_i , 50 studies

$\theta_{ik} (k \neq 1)$, from common distribution

Independent effects model ($\tau^2 \rightarrow \infty$)

η_i , 50 studies

$\theta_{ik} (k \neq 1)$, no shrinkage in treatment effects for 52 arms

58 parameters

$p_D = 61.6$

Up to 102 parameters

31

Fixed v Random Effects Models

Model	Residual Deviance* (posterior mean)	p_D	DIC	Heterogeneity (posterior median)
Random Effects	102.7	61.6	164.3	0.079
Fixed Effects	106.0	57.7	163.7	-

*Compare to 102 data points

- RE model appears to fit best
- but no advantage given more parameters...
- ... unless believe heterogeneity...

32

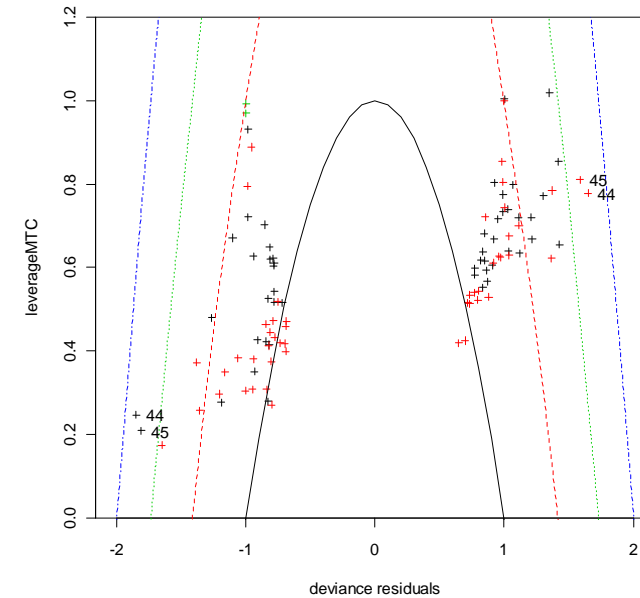
Diagnostic Plots

- Plot:
 - individual data points' contributions to the DIC (with sign given by difference between fitted and observed values)
 - against leverages (i.e. individual data points' contributions to pD)
- Highlight poorly fitting or highly influential data points:
 - Add parabolas of the form $x^2+y=c$
 - These represent contributions of c to the DIC
 - Points outside parabola with $c=3$, say, are highlighted

Spiegelhalter *et al* JRSS B, 2002

33

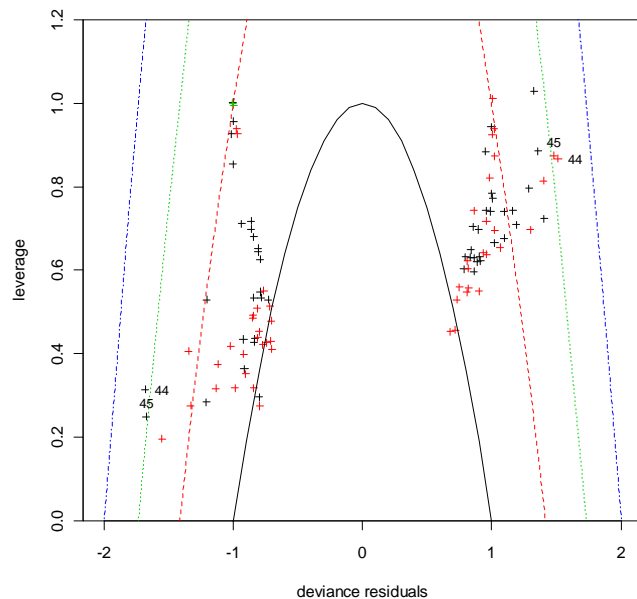
Leverage plot for FE MTC



Trials 44,45 compare treatments 3 and 9

34

Leverage plot for RE MTC



Trials 44,45 compare treatments 3 and 9

35

Checking for

INCONSISTENCY

What about inconsistency?

- The **true** treatment effects **must be** consistent
- But there may be inconsistencies in the **EVIDENCE**
- How to check for this?

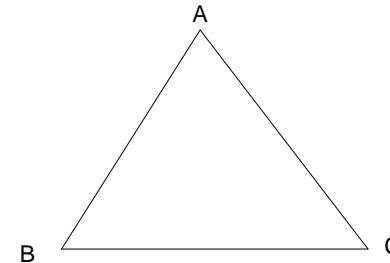
37

Question:

- How many “inconsistencies” could there be ?

Treatments A,B,C.

Trials or sets of trials AB, AC, BC



38

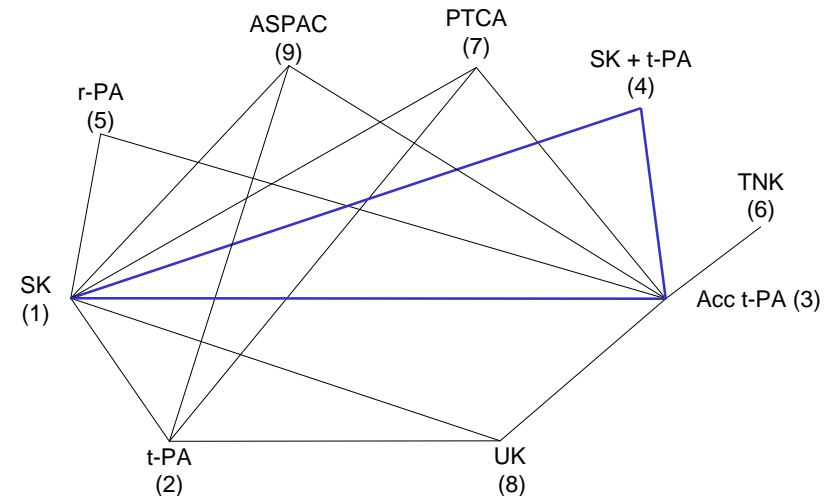
How many inconsistencies?

- Inconsistencies are properties of loops
- Inconsistency degrees of freedom (ICDF) is the maximum number of possible inconsistencies*
- Informally described as the number of **independent** 3-way loops in the evidence structure
- In this example the ICDF is seven
 - Count independent 3-way loops
 - Discount any loops formed *only* by 3-arm trials
 - One such loop (1,3,4), in this example.
- Multiple testing?

39

* Lu and Ades, JASA 2006

Thrombo: Treatment Network



40

Inconsistency & heterogeneity

- “heterogeneity” in treatment effects is the variation in treatment effects between trials
 - **WITHIN** pair-wise contrasts, eg within AB trials
- “inconsistency” is variation in treatment effects
 - **BETWEEN** pair-wise contrasts, eg AB, AC results inconsistent with BC.

Both due to ‘missing’ covariates: factors that interact with the treatment effect but vary between trials

To measure heterogeneity, must look at **trials**.

To measure inconsistency, can focus on the **pooled summaries** of evidence on pair-wise contrasts...

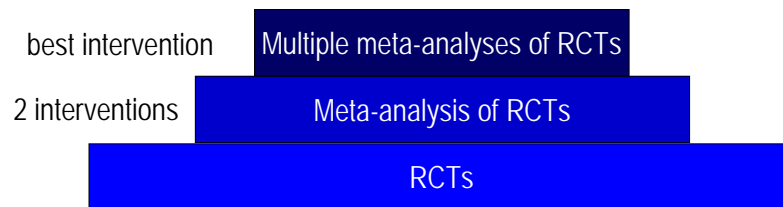
41

Inconsistency & heterogeneity

- We can have inconsistency when no heterogeneity is present (i.e. in a FE model)
- But will a RE model disguise true inconsistencies?
 - Possibly, depends on the evidence network
 - Not the case in the Thrombolytics example

42

Inconsistency - Heterogeneity



- Recall
 - heterogeneity relates to variability of distribution of random treatment effects
 - Inconsistency relates to validity of consistency assumptions, ie across comparisons

43

Main ideas for checking consistency

1. Compare posterior distributions obtained from direct and indirect evidence for each comparison
2. Model fit/comparison problem
 - Fit models with and without consistency assumptions
 - Compare model fit (residual deviance, DIC)
3. A mixture of both

44

Comparison of direct and indirect estimates

- Method for triangles (Bucher JCE, 1997)
 - Separate Pairwise meta-analyses on all contrasts
 - Calculate indirect estimate (using consistency equations)
 - Ignores network
- Evaluation of concordance within closed loops
 - previous session
- Can be extended to whole networks
 - Dias et al. Statistics in Medicine (in press, 2010)
 - Problems when three-arm trials included or when random effects models used.

45

Model comparison

- In a complex treatment network, what is direct evidence for one comparison is indirect for another...
- ... and there are multiple ways in which to form an 'indirect' comparison
- Better to think as model criticism

Is my consistency model reasonably supported by the evidence?

46

Inconsistency models*

Consistency model

9 treatments,
8 basic parameters

$$\mu_{1,2}, \mu_{1,3}, \mu_{1,4}, \dots \sim N(0, 10^5)$$

$$\mu_{2,3} = \mu_{1,3} - \mu_{1,2}$$

...

$$\mu_{8,9} = \mu_{1,9} - \mu_{1,8}$$

Inconsistency model

Add 7 parameters
(8+7=15 parameters)
Compare model fit

$$\mu_{2,7} = \mu_{1,7} - \mu_{1,2} + \omega_{1,2,7}$$

$$\mu_{2,8} = \mu_{1,8} - \mu_{1,2} + \omega_{1,2,8}$$

...

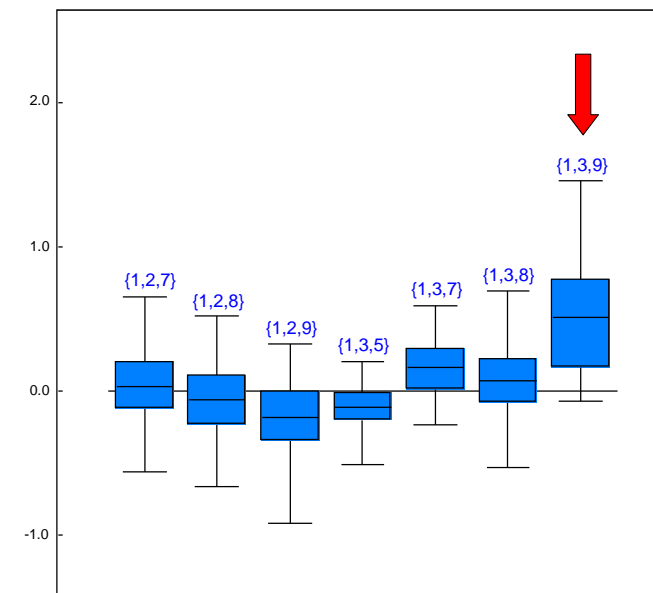
$$\omega_{1,x,y} \sim N(0, \sigma_{Inconsistency}^2) \quad \mu_{3,9} = \mu_{1,9} - \mu_{1,3} + \omega_{1,3,9}$$

WARNING: This model requires careful parameterisation

* Lu and Ades, JASA 2006

47

Box plot of inconsistency factors ω (RE model)



48

Independent mean effects model

Consistency model $\mu_{1,2}, \mu_{1,3}, \mu_{1,4}, \dots, \mu_{1,9} \sim N(0, 10^5)$
 9 treatments, $\mu_{2,3} = \mu_{1,3} - \mu_{1,2}$
 8 basic parameters ...
 $\mu_{8,9} = \mu_{1,9} - \mu_{1,8}$

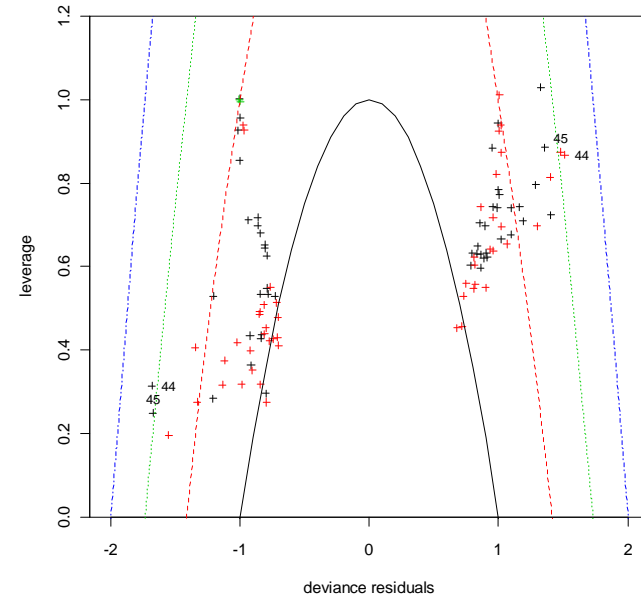
Independent mean effects model
 15 parameters (one for each pairwise contrast)

NOTE: Same number of parameters as inconsistency model!

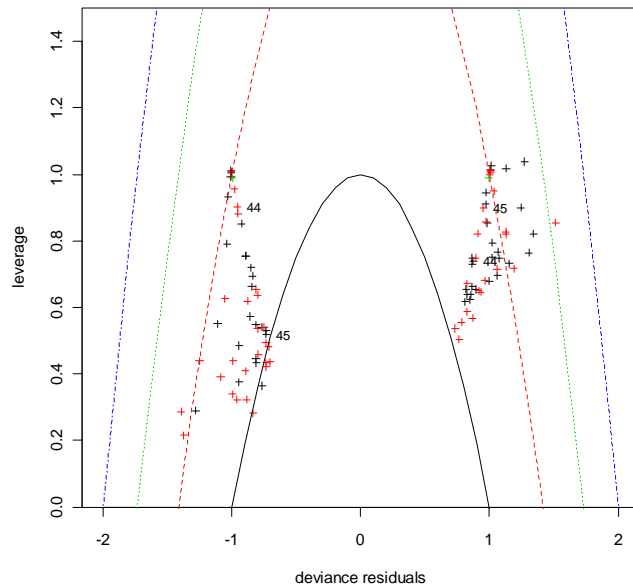
$\mu_{1,2}, \mu_{1,3}, \mu_{1,4}, \dots, \mu_{3,7}, \mu_{3,8}, \mu_{3,9} \sim N(0, 10^5)$

No consistency assumptions

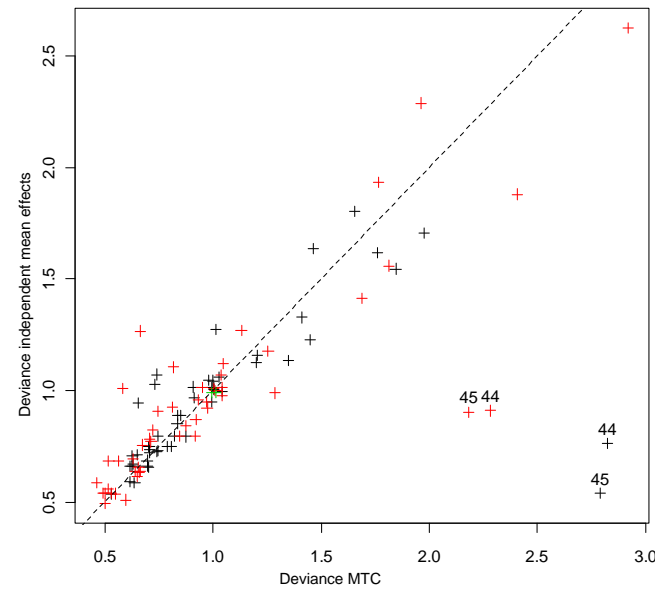
Leverage plot for RE MTC



Leverage plot indep. mean effects



Compare residual deviance for each data point



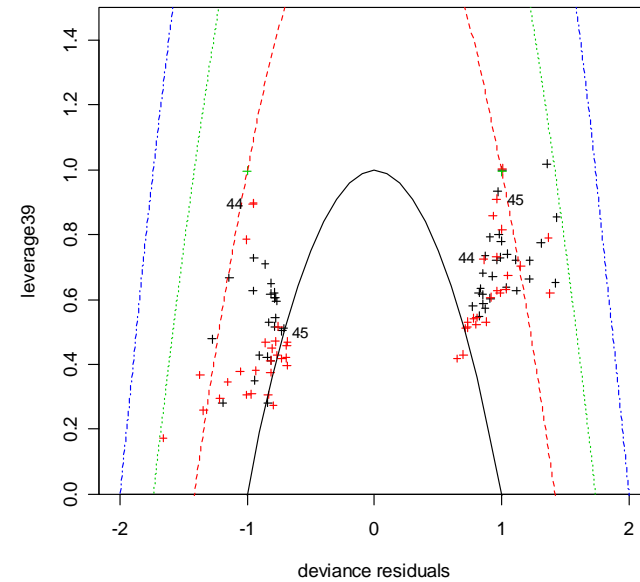
Trials 44,45
 compare
 treatments 3
 and 9

Node-splitting*

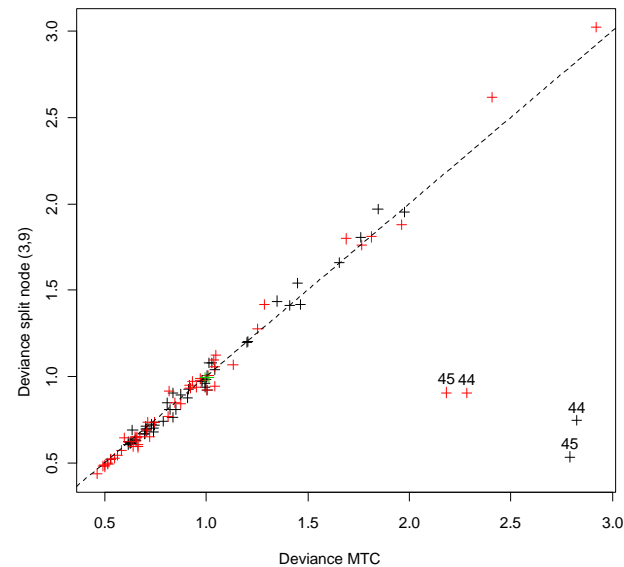
- Splits on each contrast, μ (node, eg. 2vs3)
 - Studies which compare 2 and 3 directly inform *direct* estimate
 - Rest of data with arms 2 and 3 removed inform *indirect* estimate
- Relaxes consistency assumption for one contrast at a time
- Compare model fit
 - Check between-trial heterogeneity parameters
 - residual deviance, DIC statistics
- Draw plots of posterior distributions based on direct and indirect evidence
 - Bayesian p-value to check for consistency
- Computationally intensive
 - Needs to be done for every node

* Marshall & Spiegelhalter, Bayesian Analysis (2007)
Dias et al. Statistics in Medicine (in press, 2010)

Leverage plot when node(3,9) split



Compare residual deviance for each data point



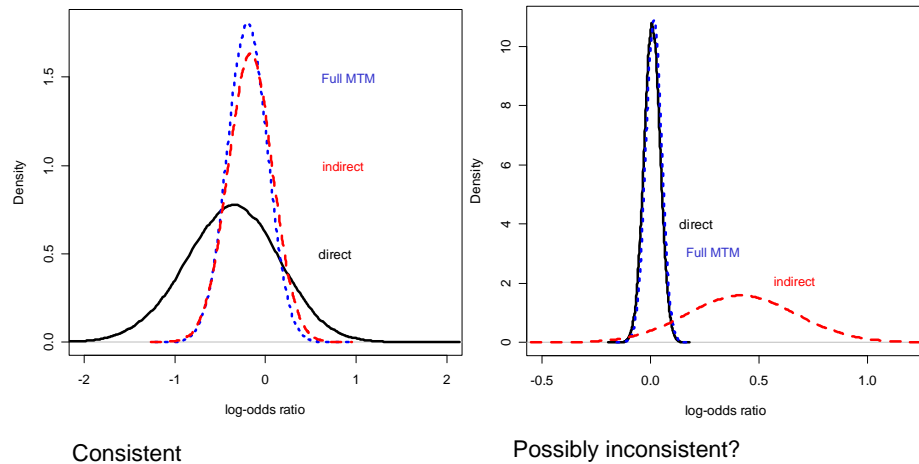
Trials 44,45 compare treatments 3 and 9

Compare model fit (RE model)

Model	Residual deviance*	pD	DIC	Between-trial heterogeneity (posterior median)
MTC	102.7	61.6	164.3	0.08
Independent mean effects	97.4	67.8	165.2	0.11
Inconsistency (ω -factors)	98.4	64.6	163.0	0.07 inconsistency variance = 0.35
Node (3,9) split	96.9	58.7	155.6	0.05

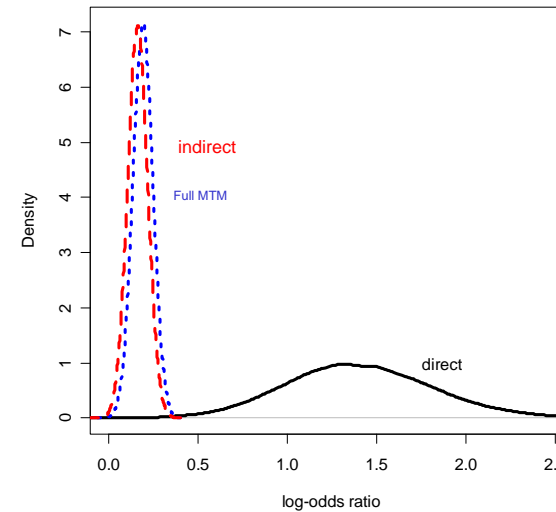
* Compare to 102 data points

Compare direct and indirect evidence



57

Inconsistent!



- Node (3,9) is split
- Direct evidence on (3,9) conflicts with indirect evidence
- Bayesian p-value < 0.005
- MTM dominated by indirect evidence from very large trials
- Only 2 small trials directly compare treats 3 and 9

58

Question

- Would you trust the direct head-to-head trials or the MTM results?
- Direct log-OR = 1.407, variance = **0.173**
 - Based on two small trials
- MTM log-OR = 0.194, variance = **0.003**
 - Based on 'borrowing strength' from evidence on all other trials (some very large)
 - And on the assumption of CONSISTENCY!

59

Why is there inconsistency?

- We have found evidence of inconsistency in node (3,9) and evidence loop (1,3,9)
 - The two direct trials comparing treat 3 vs 9 have less absolute mortality in control arm than other studies using treatment 3 as control...
 - Other baseline characteristics did not reveal other causes (same time period, no apparent difference in clinical factors)

60

Considerations on Inconsistency

- ALL these methods can ONLY detect inconsistency in a general sense.
- They cannot say which evidence is “wrong”.
- Inconsistency is a property of evidence “loops”, not of particular edges.
- Identifying which edge, or edges, are “wrong” is a task for clinical epidemiology, not statistics.
- Need to question if reasonable to combine trials in MTM *a priori*
- When no evidence of inconsistency, we can be reassured that the core MTM assumptions are met

61

Extensions and

FURTHER READING

Bias adjustment

- Given a mechanism for bias
 - e.g. lack of allocation concealment or blinding
- Estimate and adjust for bias within the network
 - Using degree of redundancy afforded by consistency assumption
- Requires a large network with multiple combinations of “biased” and “unbiased” evidence...

[Dias S, Welton NJ, Marinho V, Salanti G, Higgins JPT and Ades AE. Estimation and adjustment of bias in randomised evidence using mixed treatment comparison meta-analysis. In press, Journal of the Royal Statistical Society Series A.](#)

63

References

- Bucher HC, Guyatt GH, Griffith LE and Walter SD. The Results of Direct and Indirect Treatment Comparisons in Meta-Analysis of Randomized Controlled Trials. *Journal of Clinical Epidemiology* 1997; 50: 683-691.
- Dias S, Welton NJ, Caldwell DM and Ades AE. Checking consistency in mixed treatment comparison meta-analysis. In press, *Statistics in Medicine*.
- Lambert PC, Sutton AJ, Burton P, Abrams KR and Jones D. How vague is vague? Assessment of the use of vague prior distributions for variance components. *Statistics in Medicine* 2005; 24:2401-2428.
- Lu G and Ades AE. Assessing evidence consistency in mixed treatment comparisons. *Journal of the American Statistical Association* 2006; 101: 447-459.
- Lu G and Ades AE. Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics* 2009; 10: 792-805.
- Marshall EC and Spiegelhalter DJ. Identifying outliers in Bayesian hierarchical models: a simulation-based approach. *Bayesian Analysis* 2007; 2(2):409-444.
- Spiegelhalter DJ, Best NG, Carlin BP and van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society (B)* 2002; 64:583-616.

64

For further details on MTM, including courses
<http://bristol.ac.uk/cobm/research/mpes>