# A PRACTICAL APPROACH TO CONCEPTS AND METHODS USED TO ASSESS HETEROGENEITY AND INCONSISTENCY IN NETWORK META-ANALYSES

Dipen Patel
*Pharmerit International*

Varun Ektare
*Pharmerit International*

Berhanu Alemayehu
*Merck Inc*

Sonya J. Snedecor
*Pharmerit International*

Workshop W24, Wednesday, May 20, 2015
ISPOR 20th International Meeting, Philadelphia, PA, USA

# INTRODUCTION TO NETWORK META-ANALYSES

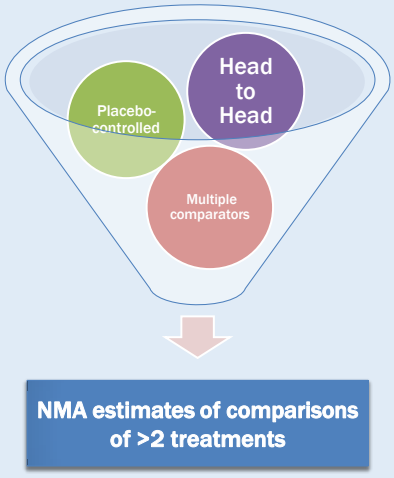
Sonya J. Snedecor
*Pharmerit International*

2

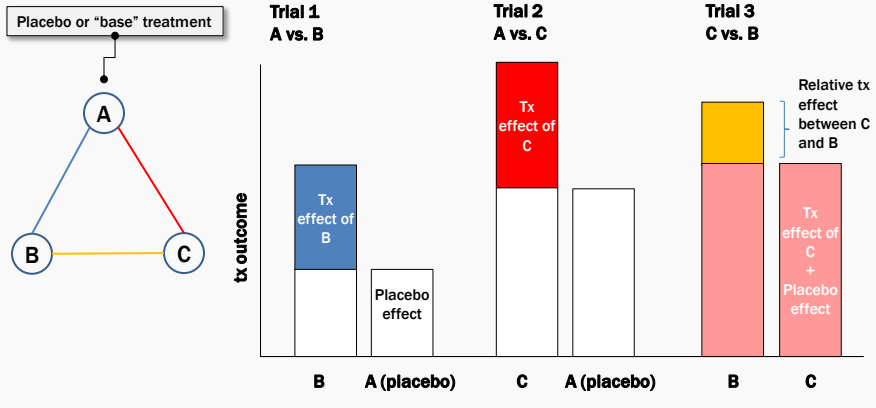
# COMPREHENSIVE DECISION-MAKING REQUIRES COMPARISONS OF ALL RELEVANT COMPETING INTERVENTIONS

## Network Meta-Analysis (NMA):

The statistical combination of outcomes from several studies across multiple treatments

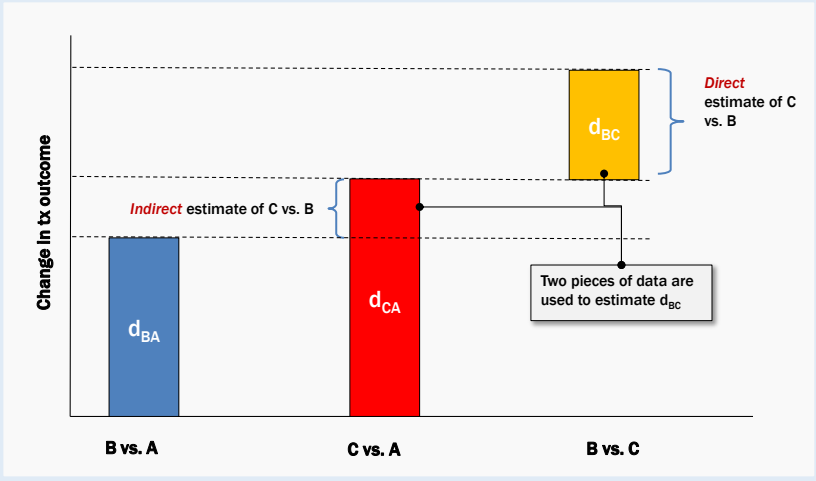



# NMA ESTIMATION OF TREATMENT EFFECTS



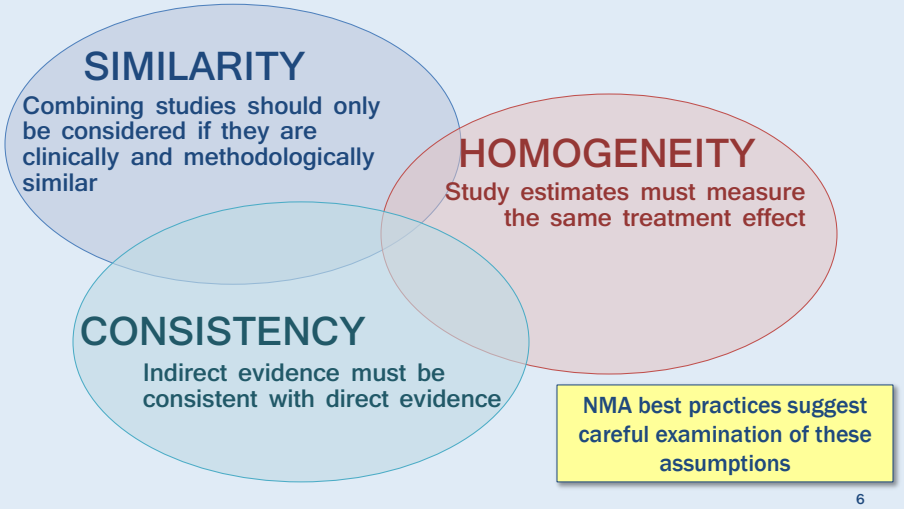

*Source: ISPOR Task Force Report: ITC & NMA Study Questionnaire, DRAFT version Sep 24, 2013*

# NMA ESTIMATION OF TREATMENT EFFECTS




*Source: ISPOR Task Force Report: ITC & NMA Study Questionnaire, DRAFT version Sep 24, 2013*

# 3 BASIC ASSUMPTIONS OF NMA



# NMA GOOD PRACTICES

*Hoaglin et al. VIH 2011 "Conducting ITC and NMA Studies: Report of the ISPOR Task Force on ITC Good Research Practices—Part 2"*

"Evaluation of homogeneity and consistency (if the network supports both direct and indirect comparisons) should be specified as components of the analysis…"

7

# QUALITY REPORTING ASSESSMENT ITEMS

*Donegan et al. PLoS One 2010 "ITCs: A Review of Reporting and Methodological Quality"*

- Is a reasonable approach used to assess the assumption of similarity?
- Is the method used to determine the presence of statistical heterogeneity adequate?
- Is the homogeneity assumption satisfied or is statistical heterogeneity accounted for if present?
- Is consistency of effects assessed?

8

# ISPOR-AMCP-NPC NMA CREDIBILITY ITEMS

*Jansen et al. VIH 2014 "ITC/NMA Study Questionnaire to Assess Relevance and Credibility to Inform Health Care Decision Making: An ISPOR-AMCP-NPC Good Practice Task Force Report"*

- Are systematic differences in treatment effect modifiers across the … comparisons identified before comparing individual study results?
- If both direct and indirect comparisons are available for pairwise contrasts (i.e., closed [network] loops), was agreement in treatment effects (i.e., consistency) evaluated or discussed?
- With inconsistency … across the different types of comparisons in the network of trials, did the researchers attempt to minimize this bias with the analysis?
- If there are indications of heterogeneity, were subgroup analyses or meta-regression analysis with prespecified covariates performed?

9

# LEARNING OBJECTIVES

- Key concepts of similarity, homogeneity, and consistency
- Methods to detect heterogeneity and inconsistency
- Investigate possible methodological solutions when assumptions are violated
- Importance of appropriate conduct of NMAs for use in HTAs

10

# SIMILARITY

Dipen Patel
*Pharmerit International*

## WHAT FACTORS DETERMINE SIMILARITY?

- Included trials should be "comparable" in terms of key factors that could affect the outcome of treatment
- If differences in patient or study characteristics would not be expected to influence treatment effect, the assumption of similarity is not violated
- **There are no statistical methods to test for similarity**
- Must use clinical knowledge and best judgement to assess appropriate comparability

# IDENTIFY KNOWN TREATMENT EFFECT MODIFIERS

- Effect modifiers include more than patient characteristics
- Trial and outcome properties are also important
- The PICOS framework can be used to identify these variables

*Would the treatment be expected to work equally in all patients included into the meta-analysis?*

| | Description | Sample Variab |
|---|---|---|
| P | Patient Population | Demographics, baseline clinical characteristics, disease severity |
| I | Intervention | Dose, mode of admin, duration |
| C | Comparator | Active treatment, placebo, concomitant meds |
| O | Outcomes | Definitions, thresholds, ITT vs. PP, EOS vs. EOT, LOCF vs. NC=F, |
| S | Setting | Study design, study duration, location/country, method of outcome assessment |

# IDENTIFY KNOWN TREATMENT EFFECT MODIFIERS

- Effect modifiers include more than patient characteristics
- Trial and outcome properties are also important
- The PICOS framework can be used to identify these variables

*Dosing and duration may or may not be important to treatment outcome.*

| | Description | Sample Variab |
|---|---|---|
| P | Patient Population | Demographics, baseline clinical characteristics, disease severity |
| I | Intervention | Dose, mode of admin, duration |
| C | Comparator | Active treatment, placebo, concomitant meds |
| O | Outcomes | Definitions, thresholds, ITT vs. PP, EOS vs. EOT, LOCF vs. NC=F, |
| S | Setting | Study design, study duration, location/country, method of outcome assessment |

# IDENTIFY KNOWN TREATMENT EFFECT MODIFIERS

- Effect modifiers include more than patient characteristics
- Trial and outcome properties are also important
- The PICOS framework can be used to identify these variables

| | Description | Sample Variables |
|---|---|---|
| P | Patient Population | Demographics, baseline clinical characteristics, disease severity |
| I | Intervention | Dose, mode of admin, duration |
| C | Comparator | Active treatment, placebo, concomitant meds |
| O | Outcomes | Definitions, thresholds, ITT vs. PP, EOS vs. EOT, LOCF vs. NC=F, |
| S | Setting | Study design, study duration, location/country, method of outcome assessment |

*In pair-wise meta-analyses the comparator must be the same for each trial. In NMA, the comparators need not be equal, but it must fit within the network.*

# IDENTIFY KNOWN TREATMENT EFFECT MODIFIERS

- Effect modifiers include more than patient characteristics
- Trial and outcome properties are also important
- The PICOS framework can be used to identify these variables

| | Description | Sample Variables |
|---|---|---|
| P | Patient Population | Demographics, baseline clinical characteristics, disease severity |
| I | Intervention | Dose, mode of admin, duration |
| C | Comparator | Active treatment, placebo, concomitant meds |
| O | Outcomes | Definitions, thresholds, ITT vs. PP, EOS vs. EOT, LOCF vs. NC=F, |
| S | Setting | Study design, study duration, location/country, method of outcome assessment |

*How outcomes are calculated can influence observed treatment effect.*

# IDENTIFY KNOWN TREATMENT EFFECT MODIFIERS

- Effect modifiers include more than patient characteristics
- Trial and outcome properties are also important
- The PICOS framework can be used to identify these variables

| | Description | Sample Variables |
|---|---|---|
| P | Patient Population | Demographics, baseline clinical characteristics, disease severity |
| I | Intervention | Dose, mode of admin, duration |
| C | Comparator | Active treatment, placebo, concomitant meds |
| O | Outcomes | Definitions, thresholds, ITT vs. PP, EOS vs. EOT, LOCF vs. NC=F, |
| S | Setting | Study design, study duration, location/country, method of outcome assessment |

*Some general study characteristics can be important. Eg, timing of assessments, study locations with different standards of care, patient vs. physician-reported outcomes.*

# QUALITATIVE ASSESSMENT OF SIMILARITY

- Study quality scales and checklists can help assess similarity
  - JADAD
  - Quality checklist templates for clinical studies and reviews
  - NICE quality assessment tool
- Low-quality studies should be investigated for appropriateness
- Potential patient-level effect modifiers should be qualitatively assessed to ensure comparability across studies
- Summary tables, scatter plots or charts document relevant patient characteristics and outcome measures across studies
- These help identify outliers and other dissimilarity

*National Collaborating Centre for Mental Health . "Self-Harm: The NICE Guideline on Longer-Term Management" http://www.ncbi.nlm.nih.gov/books/n/nicecg133/pdf/*
*Rodgers et al., Health Technol Assess, 2011. 15(10): Appendix 2*
*Halpern S, Douglas , M. (eds). Appendix: JADAD Scale for Reporting Randomized Controlled Trials, in Evidence-based Obstetric Anesthesia.*

# SUMMARY TABLE

*Relevant study characteristics are listed and examined for important differences*

*Key inclusion criteria are same across studies, but some differences in exclusion criteria. Clinical assessment needed to ascertain whether these differences can be potential effect modifiers and require any adjustment*

| Study (IDs) | Inclusion Criteria | Exclusion Criteria | 3rd Agent | Tx Duration / Primary Time-point | Primary Outcome Definition | Baseline | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Age | CD4 | VL (log) |
| **Study A** | Age >=18 yrs, ART-naïve (<1 week exposure), | Active AIDS illness, Clinically sig. disease, Creatinine clearance < 70ml/min, … | EFV, DRV/r | 48 weeks | TLOVR-50 | 35.5 | 228 | 4.86 |
| | | | | | | 35.3 | 218 | 4.84 |
| **Study B** | Age >=18 yrs, ART-naïve (<1 week exposure), | Active alcohol or substance use Creatinine clearance < 70ml/min… | EFV, ATV/r | 48 weeks | CVR-50 (NC=F) | 36 | 205 | 5.01 |
| | | | | | | 37 | 204 | 4.96 |
| **Study C** | Age >=18 yrs, ART-naïve (<1 week exposure), | Pregnant or breastfeeding, Decreased hepatic function, … | EFV, LPV/r | 48 weeks | TLOVR-50 | 38 | 205 | 5 |
| | | | | | | 40 | 216 | 5 |
| **Study D** | Age >=18 yrs, ART-naïve (<1 week exposure), | Active AIDS, active alcohol or substance use, Pregnant or breastfeeding, …. | EFV, RAL | 48 weeks | Snapshot-50 | 38 | 382 | 4.78 |
| | | | | | | 38 | 391 | 4.73 |


# SUMMARY TABLE

*Relevant study characteristics are listed and examined for important differences*


*Time-point for primary outcome is same across studies, and does not require any further consideration*


| Study (IDs) | Inclusion Criteria | Exclusion Criteria | 3rd Agent | Tx Duration / Primary Time-point | Primary Outcome Definition | Baseline | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Age | CD4 | VL (log) |
| **Study A** | Age >=18 yrs, ART-naïve (<1 week exposure), | Active AIDS illness, Clinically sig. disease, Creatinine clearance < 70ml/min, … | EFV, DRV/r | 48 weeks | TLOVR-50 | 35.5 | 228 | 4.86 |
| | | | | | | 35.3 | 218 | 4.84 |
| **Study B** | Age >=18 yrs, ART-naïve (<1 week exposure), | Active alcohol or substance use Creatinine clearance < 70ml/min… | EFV, ATV/r | 48 weeks | CVR-50 (NC=F) | 36 | 205 | 5.01 |
| | | | | | | 37 | 204 | 4.96 |
| **Study C** | Age >=18 yrs, ART-naïve (<1 week exposure), | Pregnant or breastfeeding, Decreased hepatic function, … | EFV, LPV/r | 48 weeks | TLOVR-50 | 38 | 205 | 5 |
| | | | | | | 40 | 216 | 5 |
| **Study D** | Age >=18 yrs, ART-naïve (<1 week exposure), | Active AIDS, active alcohol or substance use, Pregnant or breastfeeding, …. | EFV, RAL | 48 weeks | Snapshot-50 | 38 | 382 | 4.78 |
| | | | | | | 38 | 391 | 4.73 |

# SUMMARY TABLE

*Relevant study characteristics are listed and examined for important differences*

***Outcome definitions are different, and requires clinical discussion (understanding of each definition) to decide whether they can be considered similar. For HIV, FDA has been updating the required outcome definition in clinical trials, resulting in these differences***

| Study (IDs) | Inclusion Criteria | Exclusion Criteria | 3rd Agent | Tx Duration / Primary Time-point | Primary Outcome Definition | Baseline | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Age | CD4 | VL (log) |
| **Study A** | Age >=18 yrs, ART-naïve (<1 week exposure), | Active AIDS illness, Clinically sig. disease, Creatinine clearance < 70ml/min, … | EFV, DRV/r | 48 weeks | TLOVR-50 | 35.5<br>35.3 | 228<br>218 | 4.86<br>4.84 |
| **Study B** | Age >=18 yrs, ART-naïve (<1 week exposure), | Active alcohol or substance use Creatinine clearance < 70ml/min… | EFV, ATV/r | 48 weeks | CVR-50 (NC=F) | 36<br>37 | 205<br>204 | 5.01<br>4.96 |
| **Study C** | Age >=18 yrs, ART-naïve (<1 week exposure), | Pregnant or breastfeeding, Decreased hepatic function, … | EFV, LPV/r | 48 weeks | TLOVR-50 | 38<br>40 | 205<br>216 | 5<br>5 |
| **Study D** | Age >=18 yrs, ART-naïve (<1 week exposure), | Active AIDS, active alcohol or substance use, Pregnant or breastfeeding, …. | EFV, RAL | 48 weeks | Snapshot-50 | 38<br>38 | 382<br>391 | 4.78<br>4.73 |


# SUMMARY TABLE

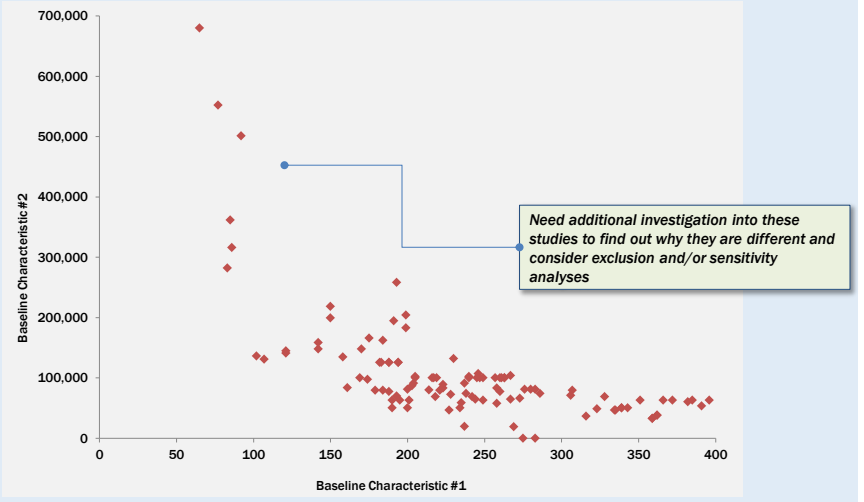
*Relevant study characteristics are listed and examined for important differences*


***Baseline age looks similar; Study D has higher baseline CD4 and requires further investigation; Baseline Viral load look comparable on log scale, but has variation on absolute scale***


| Study (IDs) | Inclusion Criteria | Exclusion Criteria | 3rd Agent | Tx Duration / Primary Time-point | Primary Outcome Definition | Baseline | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Age | CD4 | VL (log) |
| **Study A** | Age >=18 yrs, ART-naïve (<1 week exposure), | Active AIDS illness, Clinically sig. disease, Creatinine clearance < 70ml/min, … | EFV, DRV/r | 48 weeks | TLOVR-50 | 35.5<br>35.3 | 228<br>218 | 4.86<br>4.84 |
| **Study B** | Age >=18 yrs, ART-naïve (<1 week exposure), | Active alcohol or substance use Creatinine clearance < 70ml/min… | EFV, ATV/r | 48 weeks | CVR-50 (NC=F) | 36<br>37 | 205<br>204 | 5.01<br>4.96 |
| **Study C** | Age >=18 yrs, ART-naïve (<1 week exposure), | Pregnant or breastfeeding, Decreased hepatic function, … | EFV, LPV/r | 48 weeks | TLOVR-50 | 38<br>40 | 205<br>216 | 5<br>5 |
| **Study D** | Age >=18 yrs, ART-naïve (<1 week exposure), | Active AIDS, active alcohol or substance use, Pregnant or breastfeeding, …. | EFV, RAL | 48 weeks | Snapshot-50 | 38<br>38 | 382<br>391 | 4.78<br>4.73 |

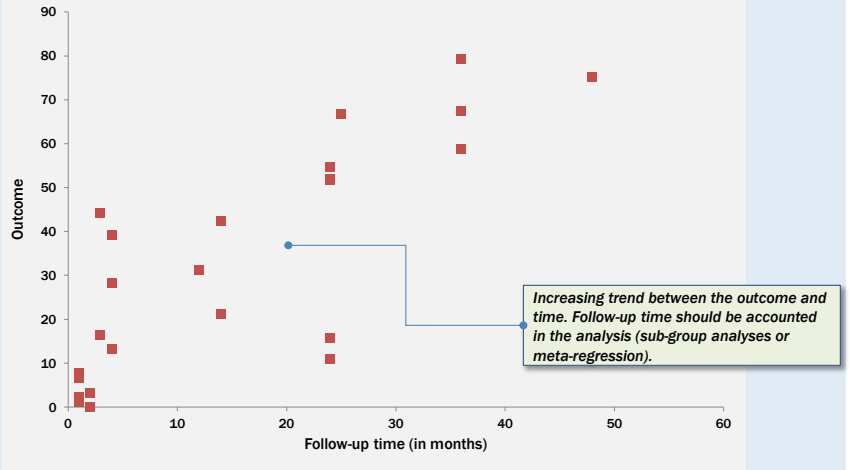
# SCATTER PLOT

*Visual plot of baseline values can help identify outliers*



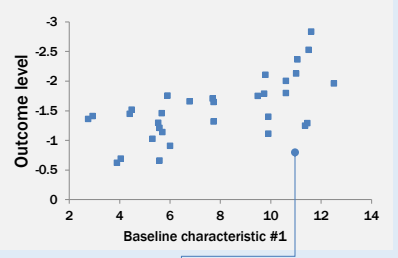

# SCATTER PLOT

*Outcome vs. Follow-Up Time*

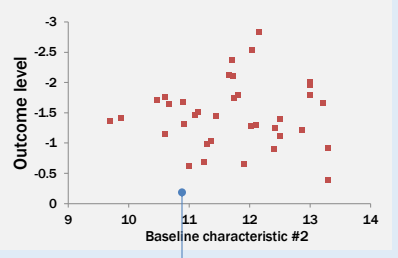


# SCATTER PLOT

*Outcome vs. Baseline severity*



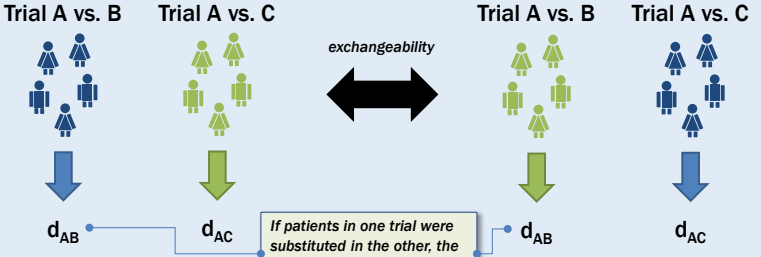

*Increasing trend between the outcome level of characteristic #1. This variable should be accounted in the analysis (sub-group analyses or meta-regression).*




*No association between outcome and baseline characteristic #2.*

# SIMILARITY vs. EXCHANGEABILITY

- The concept of similarity is similar to *exchangeability*




*If patients in one trial were substituted in the other, the observed treatment effect is expected to be the same*

- Exchangeability is an implicit assumption of NMA methodology
- Exchangeability cannot be explicitly tested, but similarity can be considered a practical proxy

*Alemayehu 2011, Comparative Effectiveness Research 2011:1 51–55*

# A FINAL WORD ON SIMILARITY

- Similarity is very important! Dissimilarity between trials will likely affect the other two assumptions
- No two independent RCTs will ever be completely similar
  - Small differences in trials are one of the strengths of meta-analyses: greater generalizability to the general population and any single RCT
- In case of dissimilarity between trials, sub-group analysis and/or meta-regression should be considered
  - These methods are explained under homogeneity assumption
- Common Criticism of Meta-Analysis: "Studies may have different patient populations that may influence outcomes!"
  - Similarity assessment aims to ensure that the analysis contains a patient population that is reasonably "combinable"

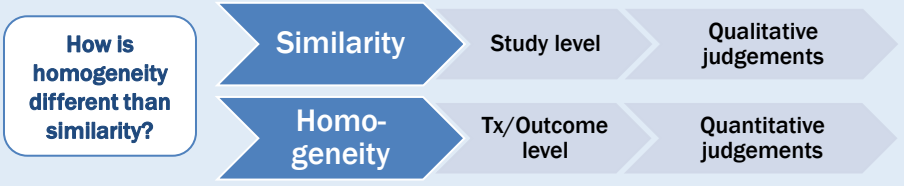
**Of course, not 100% of the time*

# HOMOGENEITY

Dipen Patel
*Pharmerit International*

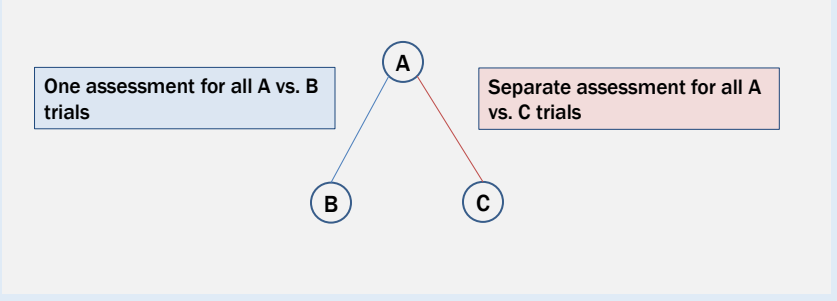
# WHAT IS HOMOGENEITY?

- The variation observed among the data is within the realm of what would be expected by random variation alone
- If not, trials may not be estimating the ***same*** treatment effect
- Homogeneity is necessarily ***a property of data from studies of the same treatment comparisons***
- This assumption is applicable to network meta-analyses as it is in pairwise meta-analyses




# HOW TO TEST FOR HETEROGENEITY?

Assessed for each collection of identical treatment comparisons within the network if ≥2 data sources are available



# HOW TO TEST FOR HETEROGENEITY?

- Cochran Q-statistic
  - Statistical test to assess whether observed differences in outcomes are due to chance alone
  - It has low power to detect heterogeneity when the number of studies is small or studies themselves have small sample size
- $I^2$ statistic
  - Describes the percentage of observed variation in the Q statistic due to heterogeneity rather than sampling error (chance), does not inherently depend upon the number of studies
- In a Bayesian framework, NICE recommends comparing model diagnostics between fixed-effect and random-effects models
  - A better model fit for RE would suggest heterogeneity*

***Caution, the choice between RE and FE models should be made on a number of factors***

*Dias S, Med Decis Making. 2013 Jul;33(5):618-40; NICE DSU TECHNICAL SUPPORT DOCUMENT 3: HETEROGENEITY: SUBGROUPS, META-REGRESSION, BIAS AND BIAS-ADJUSTMENT Cochrane Handbook; http://handbook.cochrane.org/chapter_9/9_5_2_identifying_and_measuring_heterogeneity.htm*

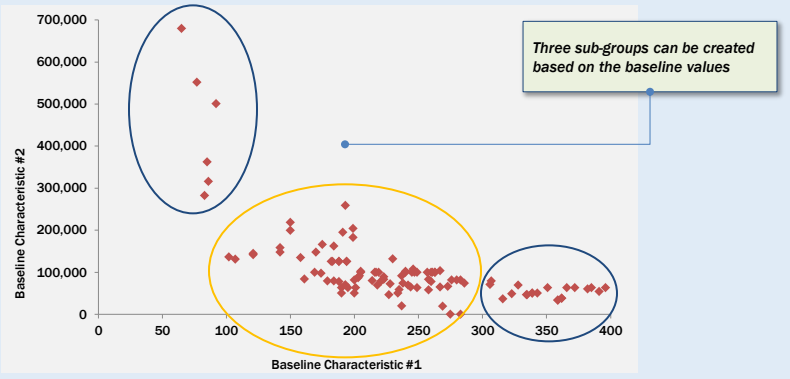
# WHAT TO DO IN PRESENCE OF HETEROGENEITY?

- Heterogeneity is an indication of dissimilarity in some effect-modifying factor
- In presence of heterogeneity, the first task should be to thoroughly explore and compare patient and trial characteristics across the studies
  - This should be already done as part of similarity testing
- If systematic differences are detected, then following methods could be considered:
  1. Sub-group analyses
  2. Random-effects modeling
  3. Meta-regression (depends on data availability)

*Dias S et al. Med Decis Making 2013;33:618–640*

# #1 SUB-GROUP ANALYSES

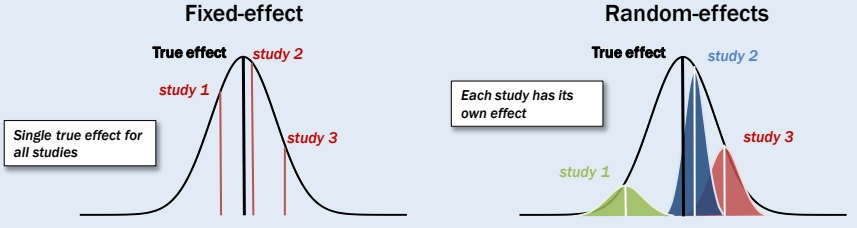
- Conduct separate analyses with data subsets stratified by the treatment modifying variables
  - Treatment effect is estimated in each stratum




*Petitti D, Meta analysis, decision analysis and cost-effectiveness analysis. 2nd edition*
*Sutton A, Heterogeneity and meta-regression, University of Leicester, Indirect and mixed treatment comparison course*

# #2 RANDOM-EFFECTS MODEL

- The basic assumption of the RE model is that there is small variation in the treatment effects among the studies

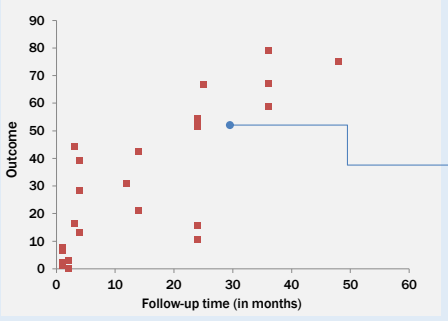



Caution: RE models should not be used without investigation of the sources of heterogeneity and consideration of the amount of data available to estimate the parameters

*Petitti D, Meta analysis, decision analysis and cost-effectiveness analysis. 2nd edition*

# #3 META-REGRESSION

- Essentially like a traditional regression
  - Effect size as a dependent variable and study-level characteristics as independent / explanatory variables
- Major challenge: Regression on study-level characteristics is often not fruitful due to lack of power
  - Availability of individual patient level data would be ideal

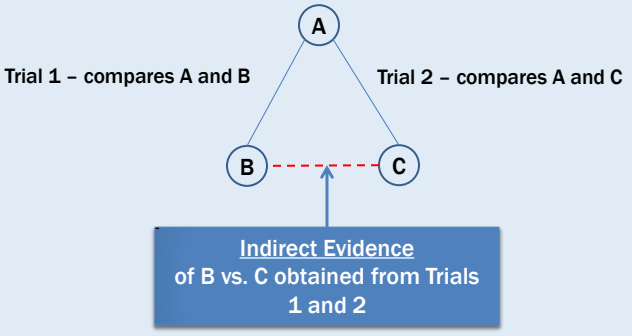



*Increasing trend between the outcome and time. Follow-up time should be included as a covariate in a meta-regression.*

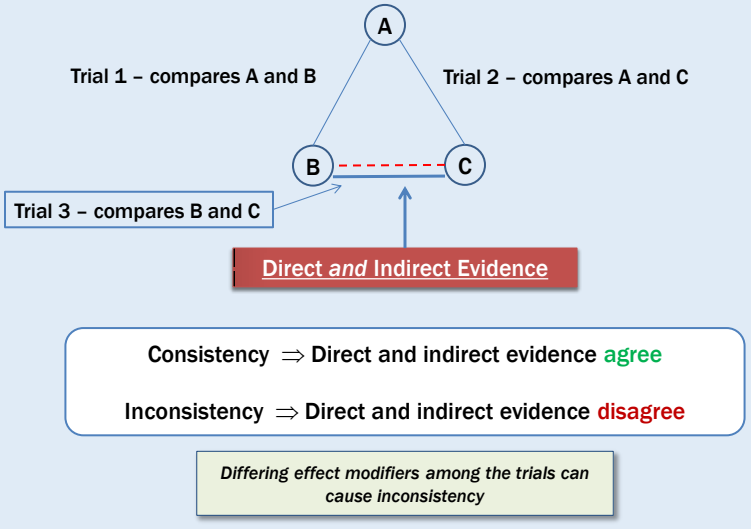
# CONSISTENCY

Varun Ektare
*Pharmerit International*

# THERE ARE 2 TYPES OF TRIAL EVIDENCE



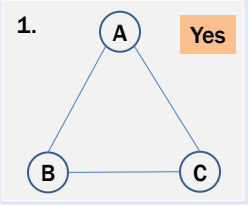

# THERE ARE 2 TYPES OF TRIAL EVIDENCE



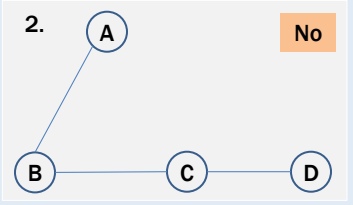
# IS INCONSISTENCY POSSIBLE IN THE FOLLOWING NETWORKS?

*Recall, there must be direct and indirect evidence to have potential inconsistency*

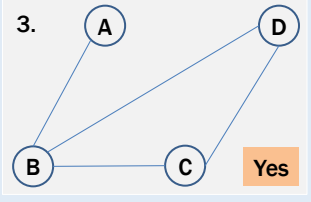
1.



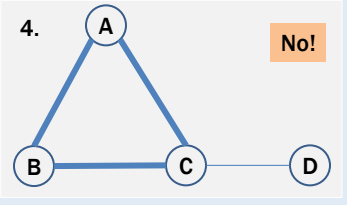

2.




3.




4.

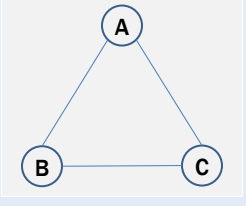



Bold connections are from a single, multi-arm trial

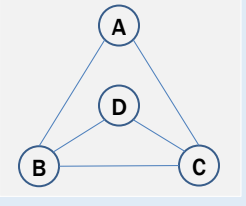
# HOW MANY INCONSISTENCIES CAN THERE BE?

Number of potential inconsistencies in any evidence network* is equal to

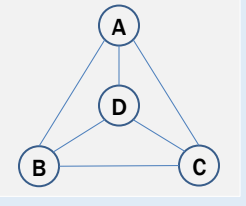
**No. of pairwise contrasts – No. of treatments + 1**




3 – 3 + 1 = 1




5 – 4 + 1 = 2

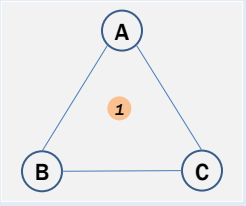


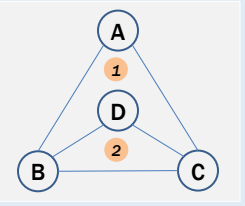

6 – 4 + 1 = 3

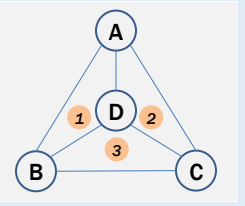
*Also known as inconsistency degrees of freedom (ICDF)

# EACH LOOP OF EVIDENCE CAN CONTRIBUTE ONE INCONSISTENCY

The number of potential inconsistencies is equal to the number of independent loops in the network




# METHODS TO TEST FOR INCONSISTENCY

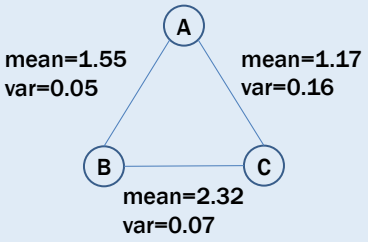
1. Bucher method
   - Can be used on triangle structures where three direct estimates are available
   - All such "triangles" should be evaluated one by one
2. Node-splitting
   - Direct and indirect studies are separated and a difference in estimates is calculated
   - Repeated for all treatment comparisons where inconsistency is possible
3. Inconsistency model
   - Could be considered "independence" model because all treatment comparisons are estimated independently
   - Treatment effects are not estimated relative to a reference treatment

# #1 BUCHER METHOD

- Compare direct and indirect evidence within a single loop
- Measure of inconsistency is the difference between direct and indirect estimates
- Statistically test whether inconsistency is present
- Must repeat these steps for each independent loop in the evidence network
- Each test is simple, but may have to perform many if network is complex

*Bucher HC, Guyatt GH, Griffith LE, 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.*

# #1 BUCHER METHOD ILLUSTRATION

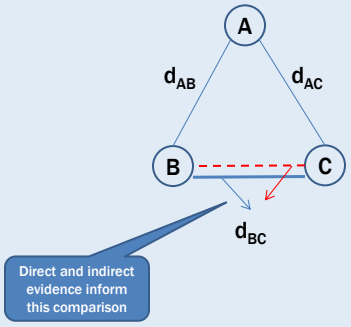



- Indirect estimate of B vs C:
  = 1.17 (A vs C) – 1.55 (A vs B) = **-0.38**
  variance = 0.16 + 0.05 = **0.21**
- Measure of inconsistency ($Z$):
  = 2.32 (Direct estimate) – (-0.38) (Indirect estimate) = **2.70**
  variance = 0.07 + 0.21 = **0.28**
- If $Z/\sqrt{Var(Z)}$ is rejected ($N(0,1)$) then the loop is inconsistent

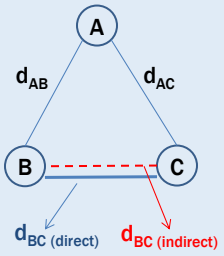
*In this case P<.000001, indicating inconsistency*

# #2 NODE-SPLITTING

## Full NMA estimates 3 parameters




## Node-splitting estimates separate parameters for direct and indirect evidence



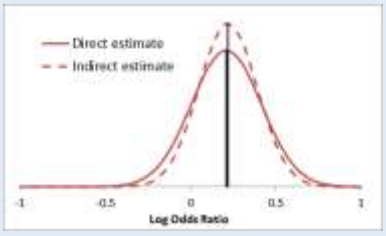

**Inconsistency is present if** $d_{BC\ (direct)} \neq d_{BC\ (indirect)}$

*Dias, S, Welton, N, Caldwell, D & Ades, A 2010, 'Checking consistency in mixed treatment comparison meta-analysis'. Statistics in Medicine, vol 29., pp. 932 - 944*

# #2 NODE-SPLITTING

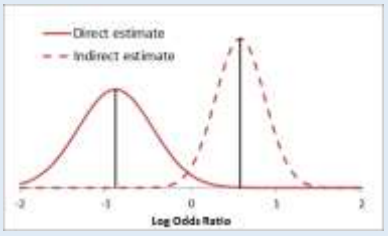
## Example of posterior distributions with direct and indirect evidence

**Consistent Evidence**




Posterior densities overlap indicating absence of inconsistency

**Inconsistent Evidence**

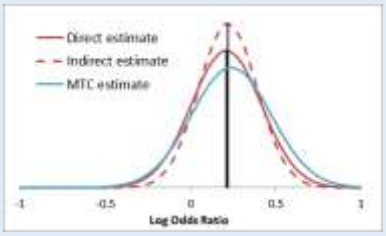



Posterior densities hardly overlap indicating presence of inconsistency

# #2 NODE-SPLITTING

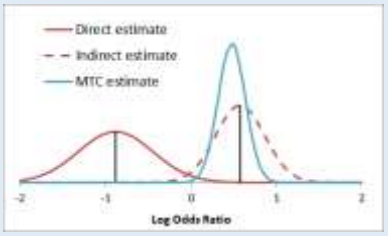
**What do we do with this information?**

Consistent Evidence




MTC estimate is similar to direct and indirect estimates

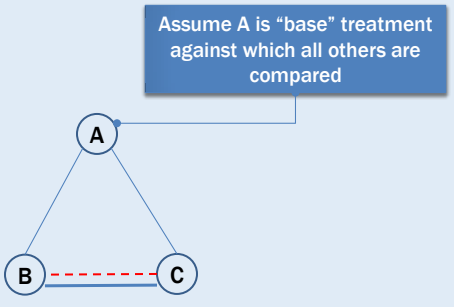
Inconsistent Evidence




MTC estimate is closer to indirect estimate, possibly because indirect trials are larger and more precise

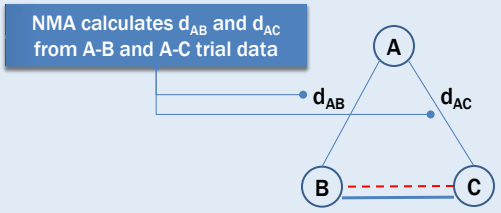
# #3 INCONSISTENCY MODEL

**Standard NMA calculations (estimated relative to "base" treatment)**



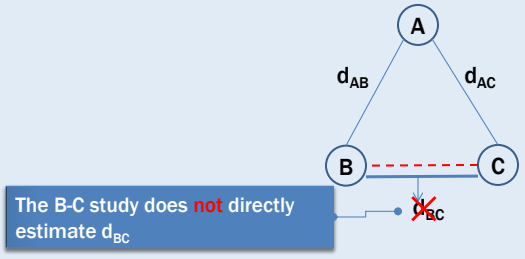
# #3 INCONSISTENCY MODEL

**Standard NMA calculations (estimated relative to "base" treatment)**



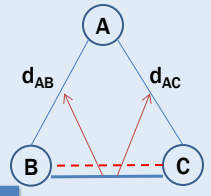

# #3 INCONSISTENCY MODEL

**Standard NMA calculations (estimated relative to "base" treatment)**



# #3 INCONSISTENCY MODEL

**Standard NMA calculations (estimated relative to "base" treatment)**




The B-C study does not directly estimate $d_{BC}$

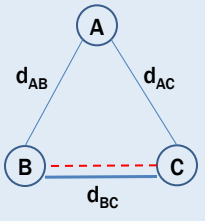
Rather, it is used to refine the estimates of $d_{AC}$ and $d_{AB}$

Then, at the end of the analysis, $d_{BC}$ is calculated by $d_{AC} - d_{AB}$

# #3 INCONSISTENCY MODEL

**With inconsistency model (all comparisons estimated independently)**

- There is no "base" treatment against which all others are compared
- NMA calculates $d_{AB}$, $d_{AC}$, and $d_{BC}$ independently from the trial data
- If model diagnostics indicate the inconsistency model fits the data better, then there is evidence of inconsistency

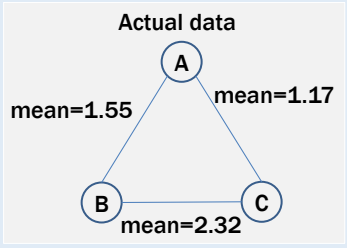


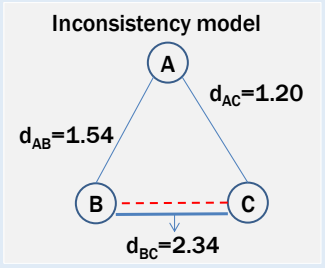

| Model | Residual Dev | pD | DIC |
|---|---|---|---|
| Standard NMA | 6.10 | 6.06 | 12.16 |
| Inconsistency model | 6.00 | 5.99 | 11.99 |

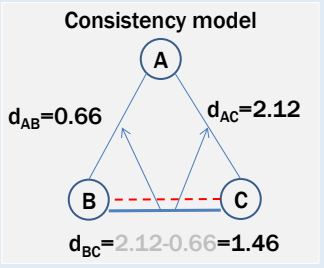
*Dias et al, NICE DSU TECHNICAL SUPPORT DOCUMENT 4: INCONSISTENCY IN NETWORKS OF EVIDENCE BASED ON RANDOMISED CONTROLLED TRIALS, May 2011*

# #3 INCONSISTENCY MODEL

Model diagnostics suggested consistency, however, also examine actual estimates for validity.




# WHAT TO DO IF INCONSISTENCY PRESENT

- Factors that cause inconsistency are the same that can cause heterogeneity
- Statistical methods cannot adjust for inconsistency since it is a property of networks of evidence and not any particular treatment comparison
- Empirical diagnostics for inconsistency are the same as similarity
  - Review for similarity in patient population or trial design
  - Potential causes are differing doses, differences in baseline severity, differences in concomitant medications allowed etc.
- Trials suspected of causing inconsistency/heterogeneity can then be excluded or their results can be used with an appropriate adjustment

54

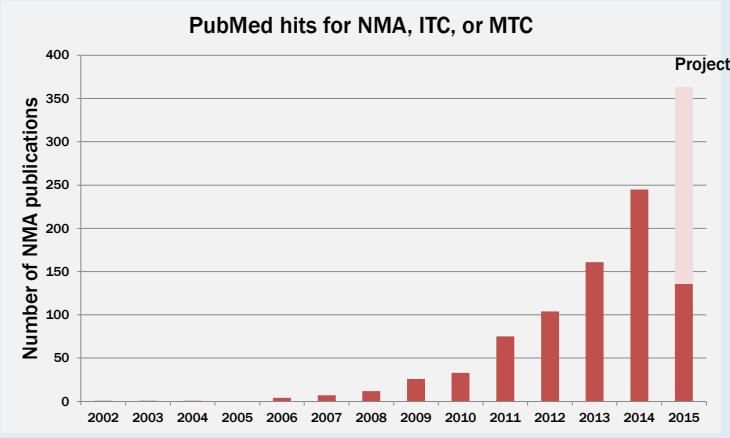
# SUMMARY OF INCONSISTENCY METHODS

| | Pros | Cons |
|---|---|---|
| **Bucher** | • Simple to perform and easy to understand<br>• Works best on "triangle structures" | • Low power when testing on loops with multiple edges or multi-arm trials<br>• Tests are correlated in complex networks |
| **Node-splitting** | • Works on networks of any complexity including multi-arm<br>• No need to summarize findings on each contrast first | • Need to be repeated for each potentially inconsistent node<br>• It can be difficult to detect inconsistency when the data is sparse |
| **Inconsistency model** | • Works on networks of any complexity including multi-arm<br>• No need to summarize findings on each contrast first | • Even if model diagnostics show no inconsistency it is important to look at actual results<br>• It can be difficult to detect inconsistency when the data is sparse |

# IMPORTANCE OF HIGH-QUALITY NMA

Berhanu Alemayehu
*Merck Inc*

# NMAs ARE INCREASINGLY COMMON




57

# QUALITY OF PUBLISHED NMAs
## *META-ANALYSES OF META-ANALYSES*

Two researchers asked the question, "Do published NMAs follow key methodological recommendations?"

Bafeta et al. 2013 BMJ

- Of 121 NMAs identified up to 12 July 2012
  - 34% reported on similarity assumption
  - 88% reported on homogeneity assumption
  - 56% with ≥1 closed loop reported consistency or exchangeability assumption

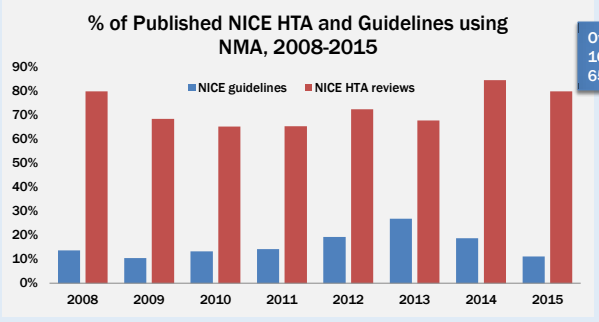
Chambers et al. 2015 PLOS One

- Of 318 NMAs identified up to 30 July 2014 (48% since Jan 2013)
  - 167 closed loop studies
  - 69% reported consistency of direct and indirect evidence
  - **Non-industry supported (79%) vs. Industry supported (39%)**

58

# NMAs ARE FREQUENTLY USED IN HTAs

- Payers in many countries are increasingly using the evidence generated from a NMA to guide their coverage and reimbursement decisions
- NICE and CADTH promote the use of NMAs when technologies compared have not been evaluated within a single RCT




*NICE Guide to the methods of technology appraisal 2013*

59

# APPROPRIATE CONDUCT AND INVESTIGATION OF NMA METHODOLOGY IS IMPORTANT TO SUPPORT TECHNOLOGY SUBMISSIONS

- NICE Decision Support Unit checklist to provide guidance on what questions should be asked by a reviewer of an evidence synthesis
  - (Similarity) A4.1. Treatments restricted to specific doses, or have different doses been "lumped" together? If the latter, is it adequately justified?
  - (Similarity) A7.2. Apparent or potential differences between trials in their patient populations? If so, has this been adequately taken into account?
  - (Homogeneity) B2.1. Have numerical estimates been provided of the degree of heterogeneity in the relative treatment effects?
  - (Consistency) C4.3. Have adequate checks for inconsistency been made?
  - (Consistency) C4.4 If inconsistency was detected, what adjustments were made to the analysis, and how was this justified?

*www.nicedsu.org.uk/TSD7%20reviewer%20checklist.final.08.05.12.pdf*

60

# ERG COMMENTS HIGHLIGHT IMPORTANCE OF EXAMINING NMA ASSUMPTIONS

## Certolizumab pegol for the treatment rheumatoid arthritis

- ITC to generate comparative effectiveness against other biologic therapies
- ERG comments:
  - "…insufficient consideration and exploration of underlying heterogeneity…"
  - "The validity of [ITC] rests on an assumption of exchangeability between trials such that the placebo arms of the trials are interchangeable. **The submission lacked an assessment or discussion of clinical or statistical heterogeneity amongst the trials used for [ITC] and did not comment on whether baseline characteristics of participants were similar across these RCTs**."

*Connock et al. HTA 2010; Vol 14:Suppl. 2*

61

# ERG COMMENTS HIGHLIGHT IMPORTANCE OF EXAMINING NMA ASSUMPTIONS

## Golimumab for the treatment psoriatic arthritis

- MTC to generate comparative effectiveness against other biologic therapies
- ERG comments:
  - "The validity of the MTC of meta-analysis is built on the assumptions that no important differences exist between trials in terms of baseline characteristics such as disease severity. The population characteristics of all seven trials are…similar in terms of patients' joint disease severity at baseline; this is important to ensure the validity of the PsARC outcome in the MTC."
  - "Overall, in the MTC analyses in the [manufacturer's submission], the degree of clinical heterogeneity between the included trials…was reasonable. Therefore, the assumption of exchangeability between the trials for the purposes of the MTC was acceptable."

*CRD and CHE Technology Assessment Group. Golimumab for the treatment of psoriatic arthritis: a single technology appraisal. Centre for Reviews and Dissemination/ Centre for Health Economics, 2010.*

62

# SUMMARY

- NMAs are useful tools, important for comparative evidence
- If not executed correctly, analysis results may lack internal and external validity
- Several meta-analysis attributes, in addition to those presented, should be examined
- At minimum, these three basic NMA assumptions should be tested, documented, and the analyses should be adjusted, as appropriate