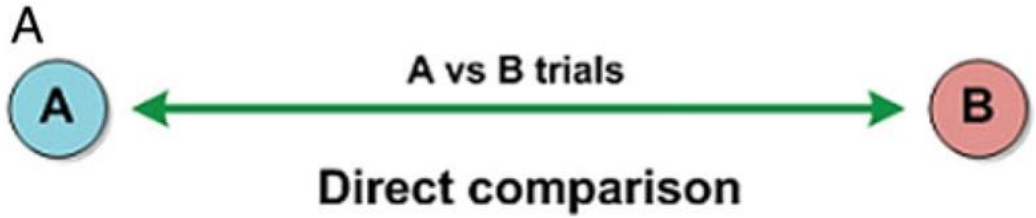
Generalità e requisiti

(G.L. Pappagallo)

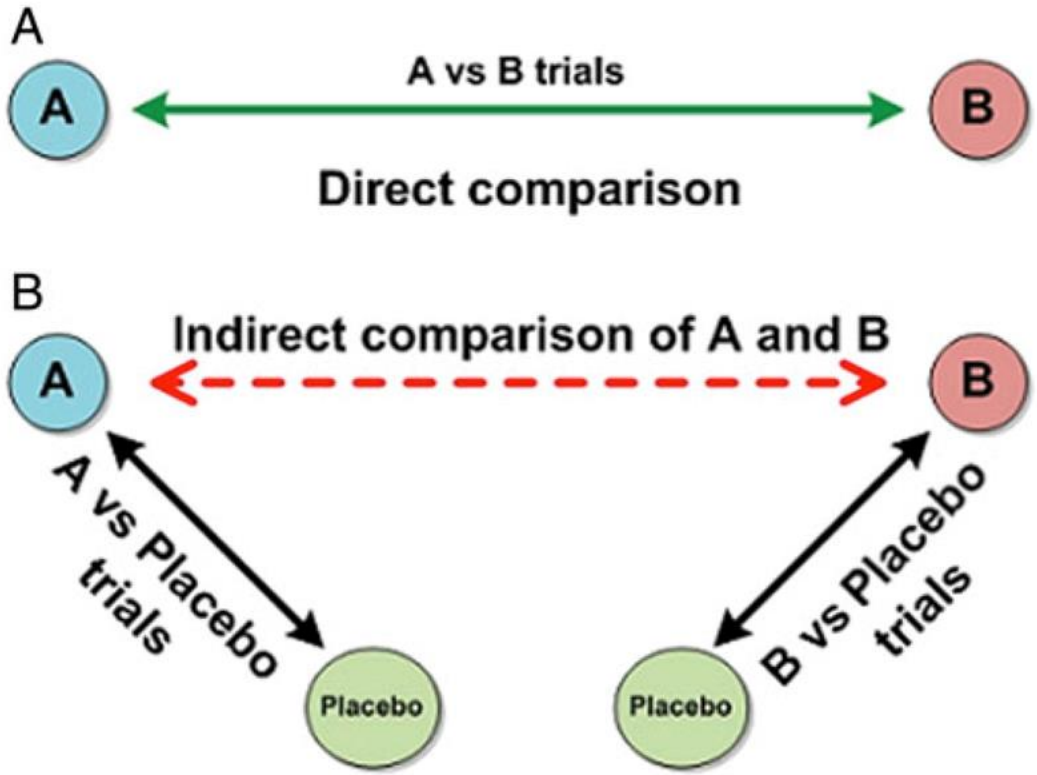


Indirect comparisons of competing interventions

AM Glenny,[1*] DG Altman,[2] F Song,[3] C Sakarovitch,[2] JJ Deeks,[2] R D'Amico,[2] M Bradburn[2] and AJ Eastwood[4]

Health Technology Assessment 2005; Vol. 9: No. 26

When conducting systematic reviews to evaluate the effectiveness of interventions, direct evidence from good-quality RCTs should be used wherever possible. If little or no such evidence exists, it may be necessary to look for indirect comparisons from RCTs. The reviewer needs, however, to be aware that the results may be susceptible to bias.



through a

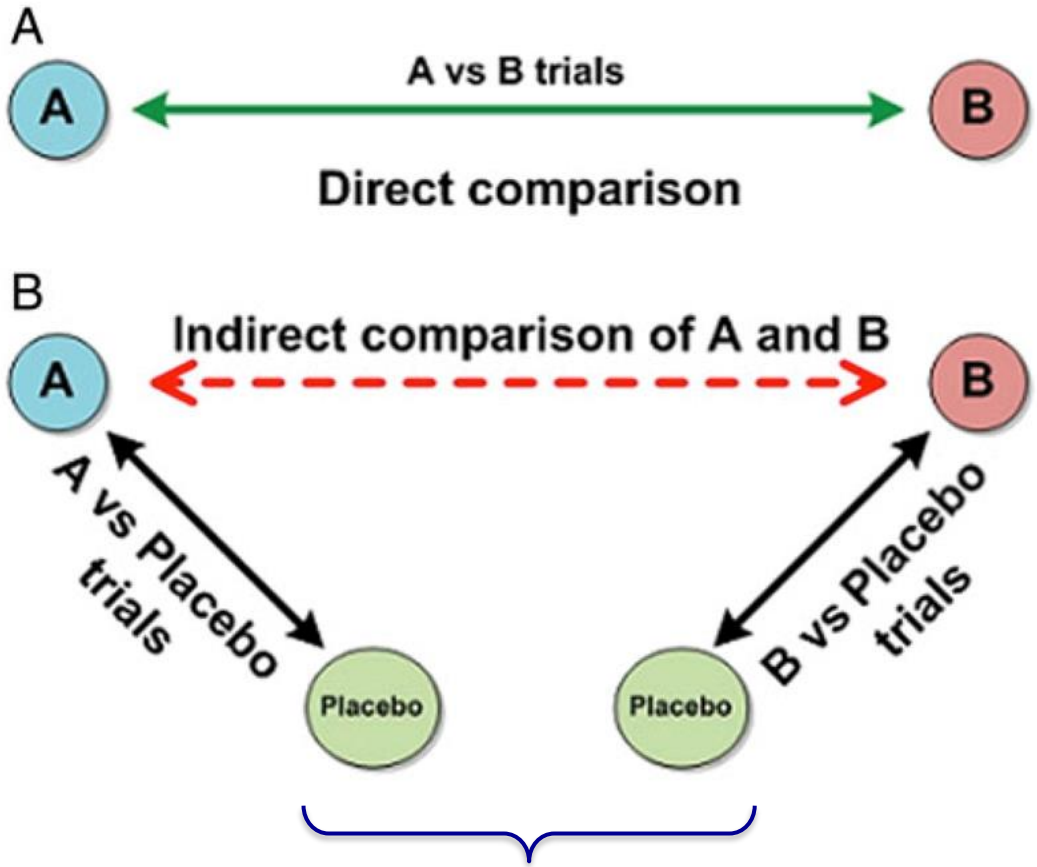
Common Comparator

Indirect comparisons of competing interventions

AM Glenny,[1*] DG Altman,[2] F Song,[3] C Sakarovitch,[2] JJ Deeks,[2] R D'Amico,[2] M Bradburn[2] and AJ Eastwood[4]

Health Technology Assessment 2005; Vol. 9: No. 26

When conducting systematic reviews to evaluate the effectiveness of interventions, direct evidence from good-quality RCTs should be used wherever possible. If little or no such evidence exists, it may be necessary to look for indirect comparisons from RCTs. The reviewer needs, however, to be aware that the results may be susceptible to bias.



Similarity Assumption

trials must be comparable on effect modifiers to obtain an unbiased pooled estimate.

Indirect comparisons of competing interventions

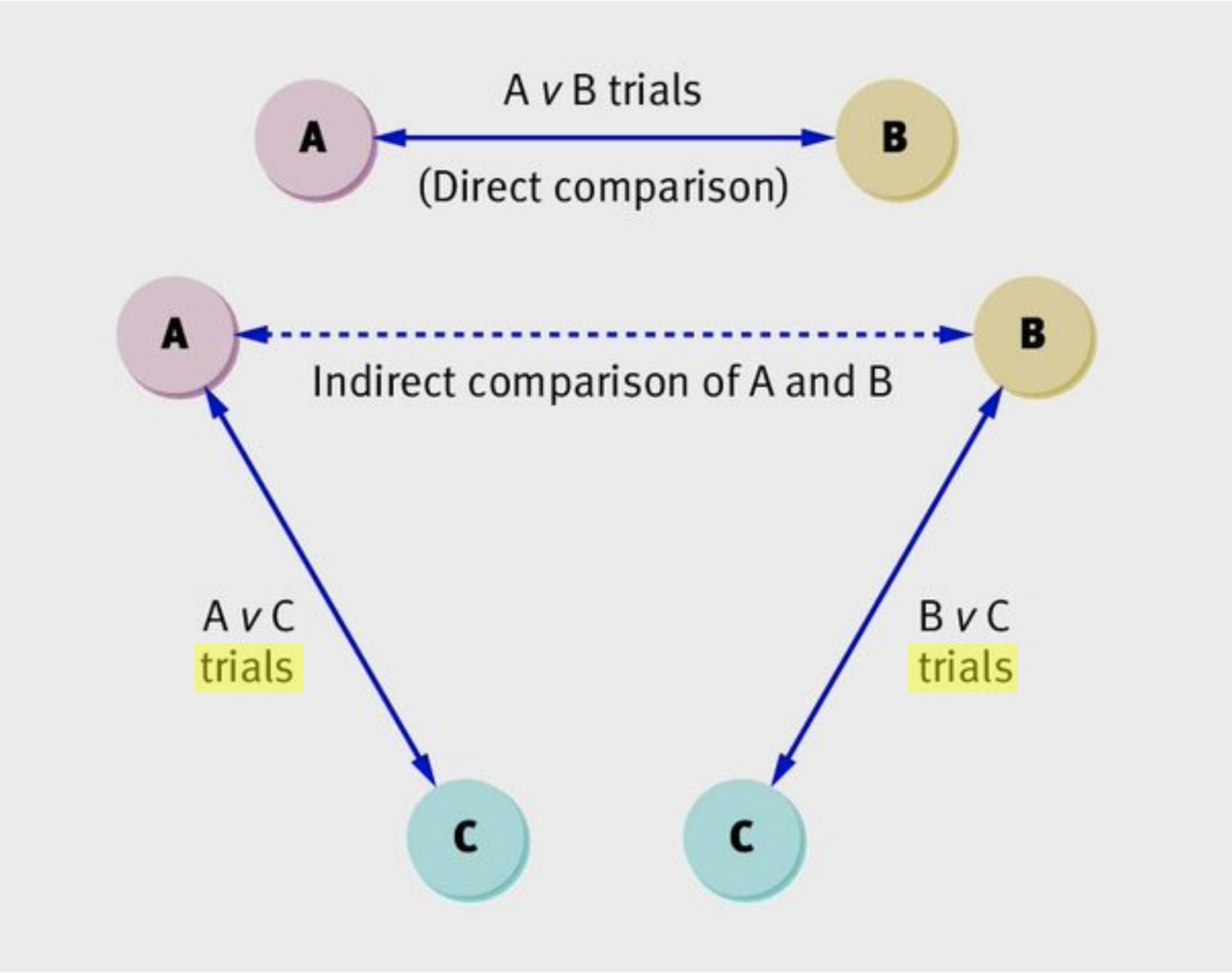
AM Glenny,[1*] DG Altman,[2] F Song,[3] C Sakarovitch,[2] JJ Deeks,[2] R D'Amico,[2] M Bradburn[2] and AJ Eastwood[4]

Health Technology Assessment 2005; Vol. 9: No. 26

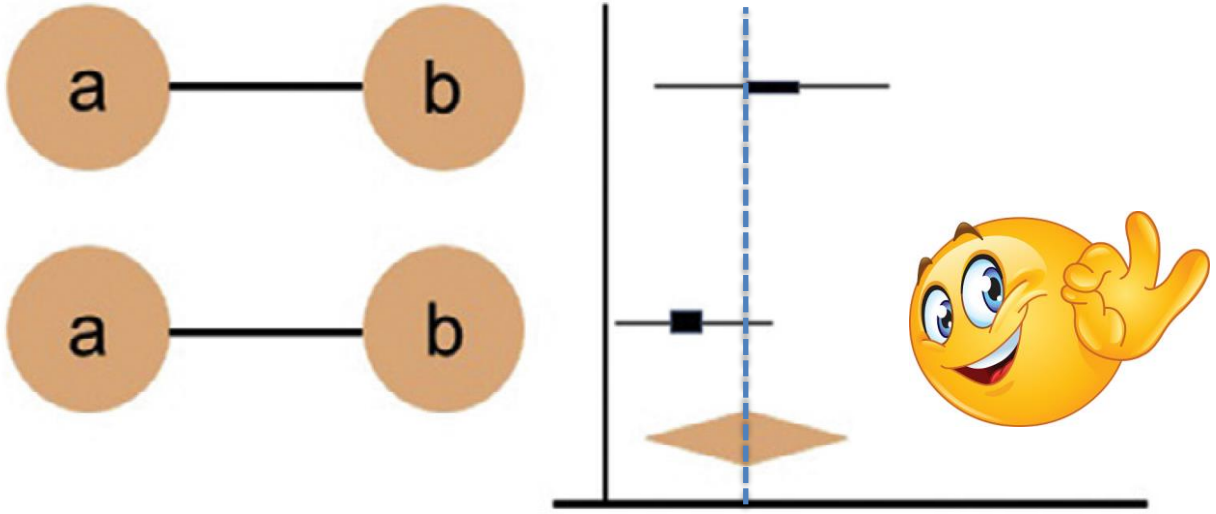
When conducting systematic reviews to evaluate the effectiveness of interventions, direct evidence from good-quality RCTs should be used wherever possible. If little or no such evidence exists, it may be necessary to look for indirect comparisons from RCTs. The reviewer needs, however, to be aware that the results may be susceptible to bias.

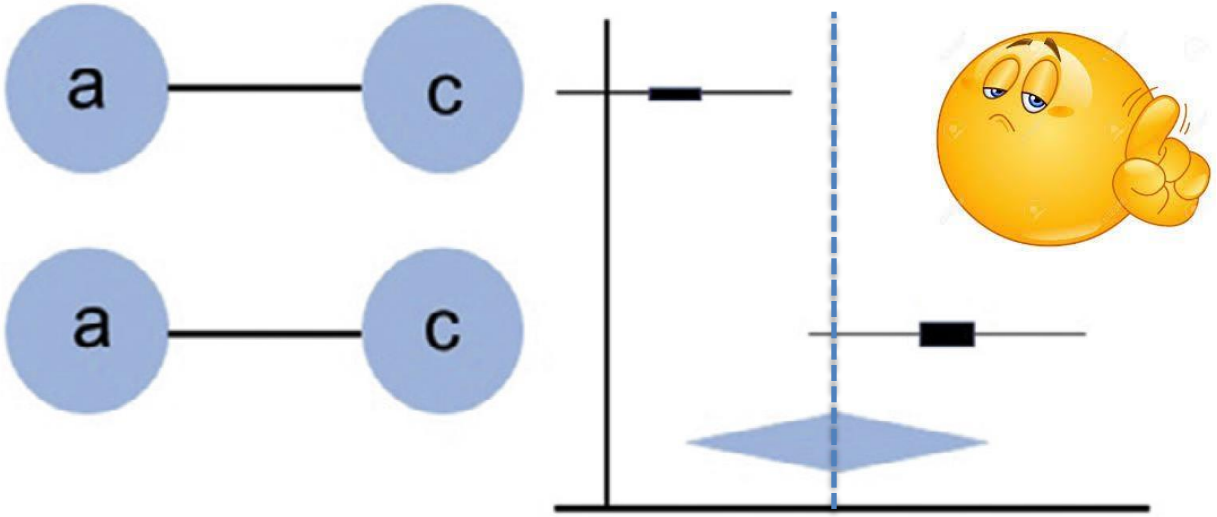
WHAT FACTORS DETERMINE SIMILARITY?

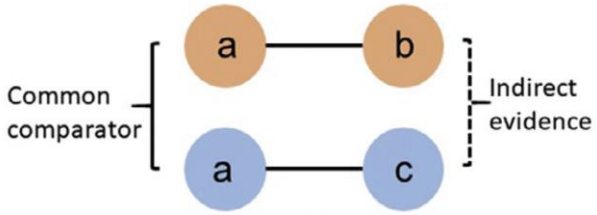
- Included trials should be “comparable” in terms of key factors that could affect the outcome of treatment
- If differences in patient or study characteristics would not be expected to influence treatment effect, the assumption of similarity is not violated
- **There are no statistical methods to test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



Quando
le evidenze dirette
sono costituite
da più trials…

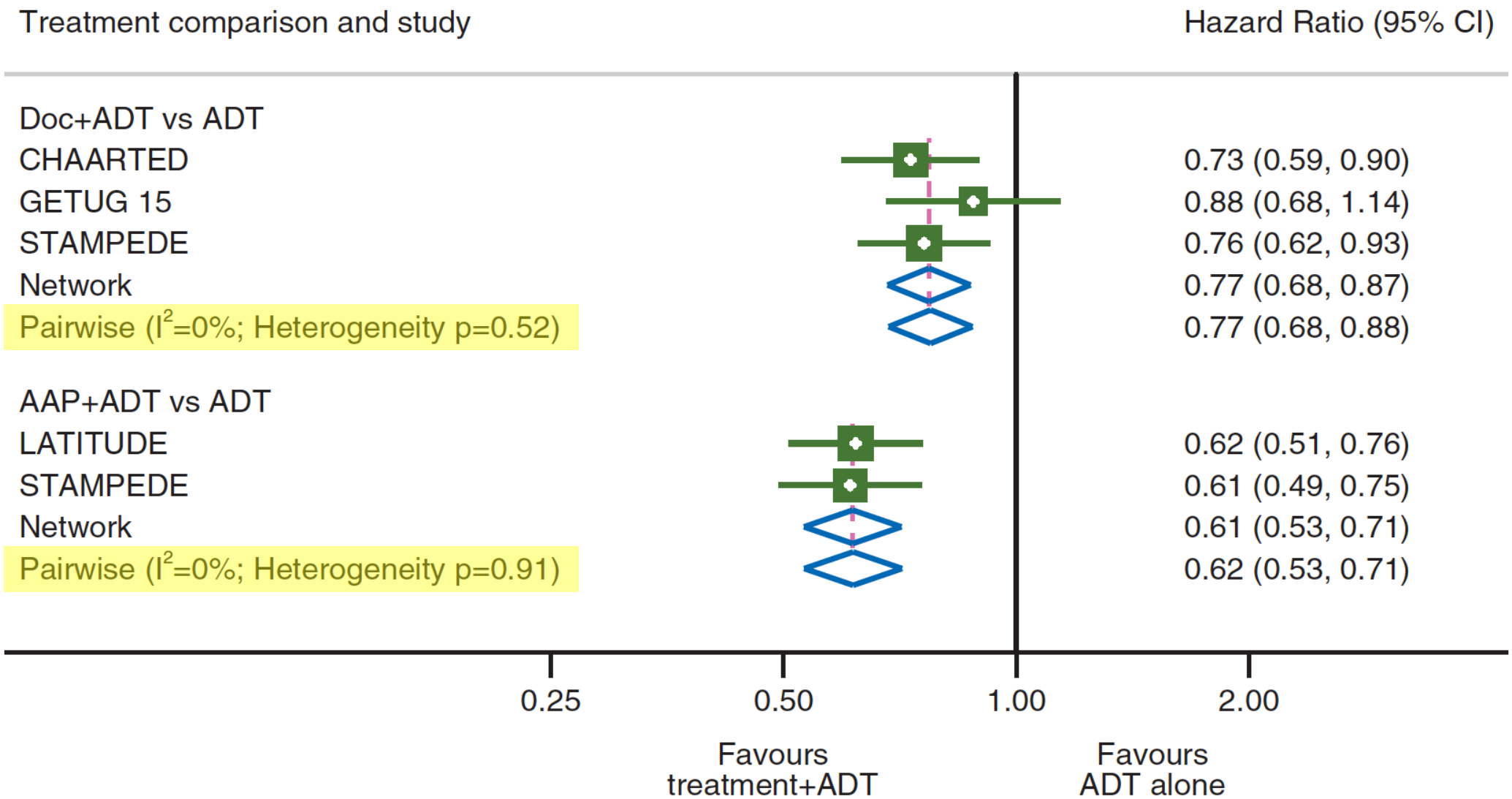






Homogeneity Assumption

there must be
no relevant heterogeneity
between trial results in
pairwise comparisons



Annals of Oncology 29: 1249–1257, 2018

Commonly applied methods

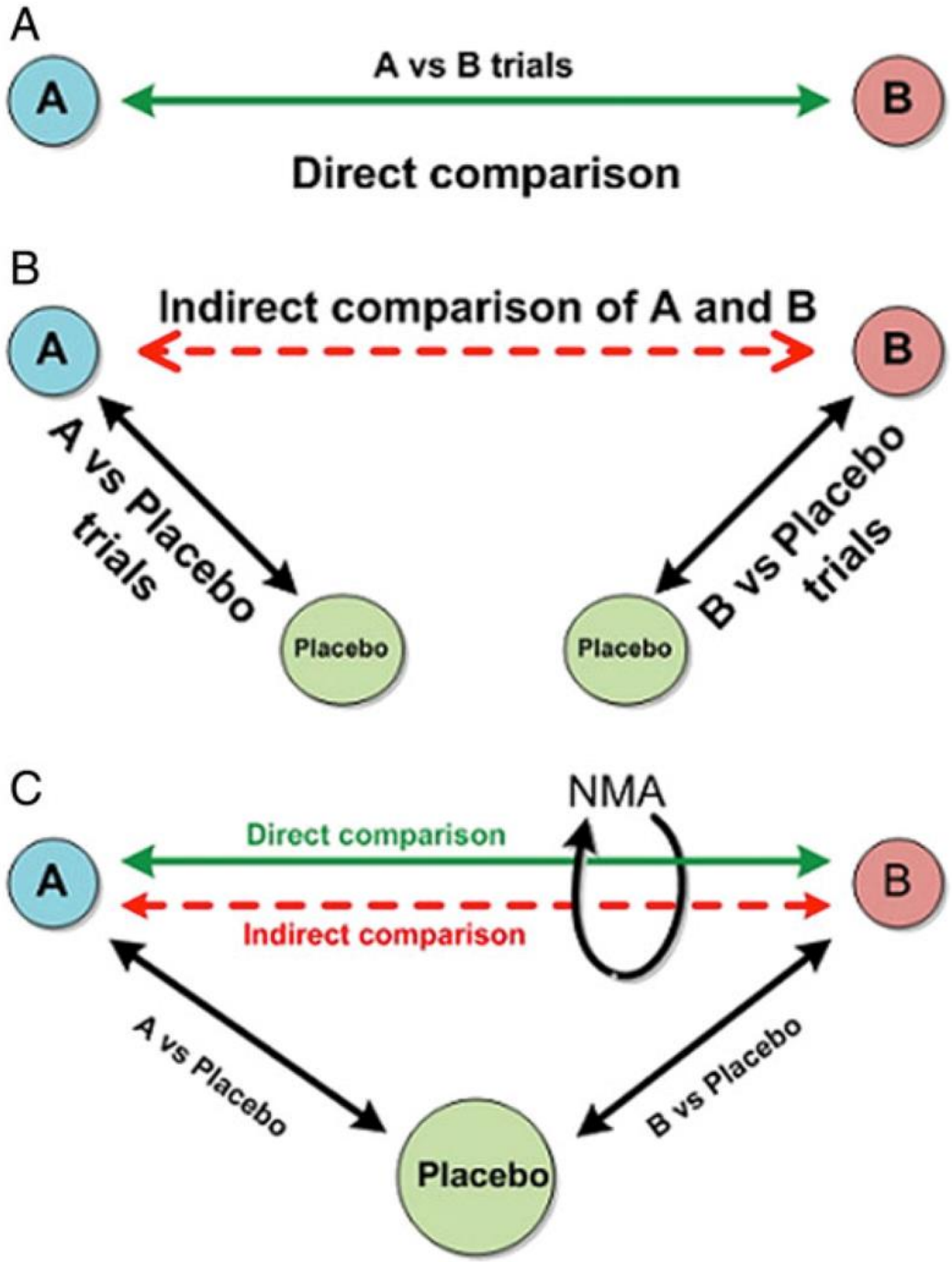
- **Bucher**
  - IPD not required
  - treatment effects calculated for each trial separately
  - within study randomization preserved
- **Population-adjusted indirect comparison (MAIC)**
  - IPD required for at least 1 trial
  - to match the IPD to the AgD of the other trial
- **Network Meta-Analysis (NMA)**
  - comparing interventions simultaneously in a single analysis by combining both direct and indirect evidence across a network of studies.

Commonly applied methods

- **Bucher**
  - IPD not required
  - treatment effects calculated for each trial separately
  - within study randomization preserved
- **Population-adjusted indirect comparison (MAIC)**
  - IPD required for at least 1 trial
  - to match the IPD to the AgD of the other trial
- **Network Meta-Analysis (NMA)**
  - comparing interventions simultaneously in a single analysis by combining both direct and indirect evidence across a network of studies.

Commonly applied methods

- **Bucher**
  - IPD not required
  - treatment effects calculated for each trial separately
  - within study randomization preserved
- **Population-adjusted indirect comparison (MAIC)**
  - IPD required for at least 1 trial
  - to match the IPD to the AgD of the other trial
- **Network Meta-Analysis (NMA)**
  - comparing interventions simultaneously in a single analysis by combining both direct and indirect evidence across a network of studies.



Indirect comparisons of competing interventions

AM Glenny,[1*] DG Altman,[2] F Song,[3] C Sakarovitch,[2] JJ Deeks,[2] R D'Amico,[2] M Bradburn[2] and AJ Eastwood[4]

Health Technology Assessment 2005; Vol. 9: No. 26

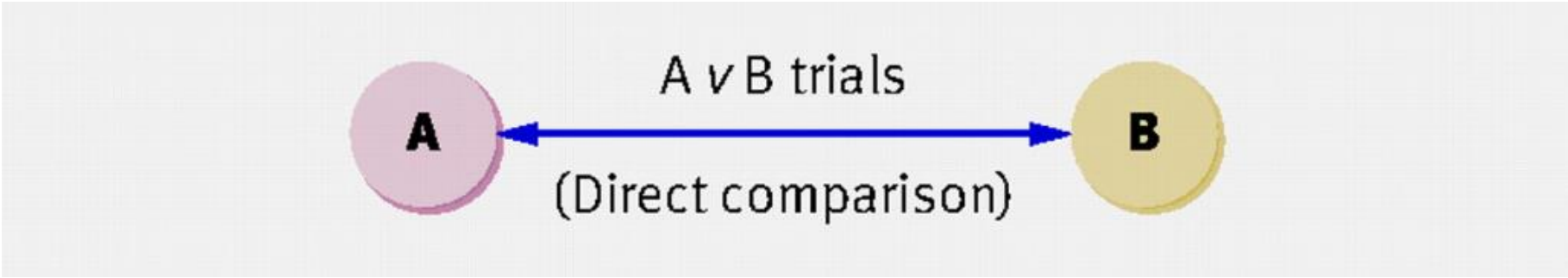
When conducting systematic reviews to evaluate the effectiveness of interventions, direct evidence from good-quality RCTs should be used wherever possible. If little or no such evidence exists, it may be necessary to look for indirect comparisons from RCTs. The reviewer needs, however, to be aware that the results may be susceptible to bias.

Consistency Assumption

there must be no relevant discrepancy between direct and indirect evidence

Indirect Treatment Comparison (Bucher)

(M. Cinquini)



The best?

No head-to-head comparison

ELSEVIER

Critical Reviews in Oncology/Hematology 94 (2015) 213–227

CRITICAL REVIEWS IN
Oncology
Hematology
Incorporating Geriatric Oncology

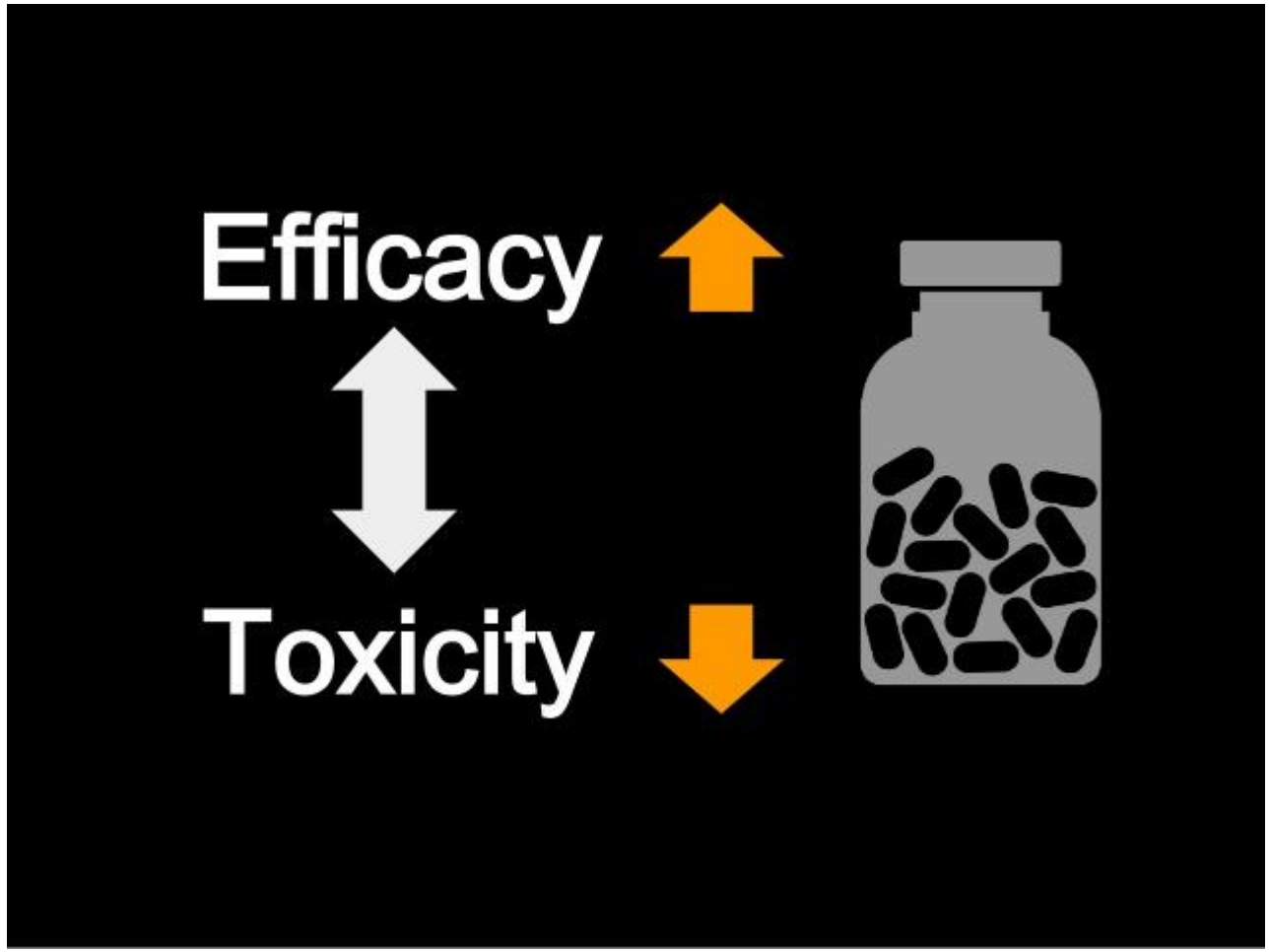
www.elsevier.com/locate/critrevonc

Is there evidence for different effects among EGFR-TKIs? Systematic review and meta-analysis of EGFR tyrosine kinase inhibitors (TKIs) *versus* chemotherapy as first-line treatment for patients harboring EGFR mutations

Eva Regina Haspinger[a], Francesco Agustoni[a], Valter Torri[b], Francesco Gelsomino[a], Marco Platania[a], Nicoletta Zilembo[a], Rosaria Gallucci[a], Marina Chiara Garassino[a,*], Michela Cinquini[b]

[a] *Medical Oncology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy*
[b] *Fondazione IRCCS Istituto di Ricerche Farmacologiche Mario Negri, Milan, Italy*

Accepted 11 November 2014


Efficacy
Toxicity

Population:

- ✓ previously untreated
- ✓ any age and race
- ✓ histologically proven NSCLC harbouring activating EGFR-mutation

Intervention:

- ✓ EGFR-TKIs (Erlotinib, Gefitinib, Afatinib)

Comparison:

- ✓ Platinum-based chemotherapy

Outcomes:

- ✓ PFS (whenever possible independently reviewed data)
- ✓ PFS in exon 19 deletion
- ✓ PFS in L858R mutation
- ✓ OS
- ✓ ORR (complete and/or partial and/or stable)
- ✓ Treatment related toxic events

Search strategy

PubMed, Cancer-Lit, Embase-databases and Cochrane-Library were searched for RCTs up to June 2014 with no language or publication status restrictions. Search terms were “TKI” [Substance Name] and “Carcinoma, NSCLC”[Substance Name]. The proceedings of the 2008–2014 conferences of the American Society of Clinical Oncology(ASCO), European Society of Medical Oncology (ESMO)and International Association for the Study of Lung Cancer (IASLC), World Conference of Lung Cancer were also searched for relevant abstracts. Any unpublished RCTs were considered for inclusion.

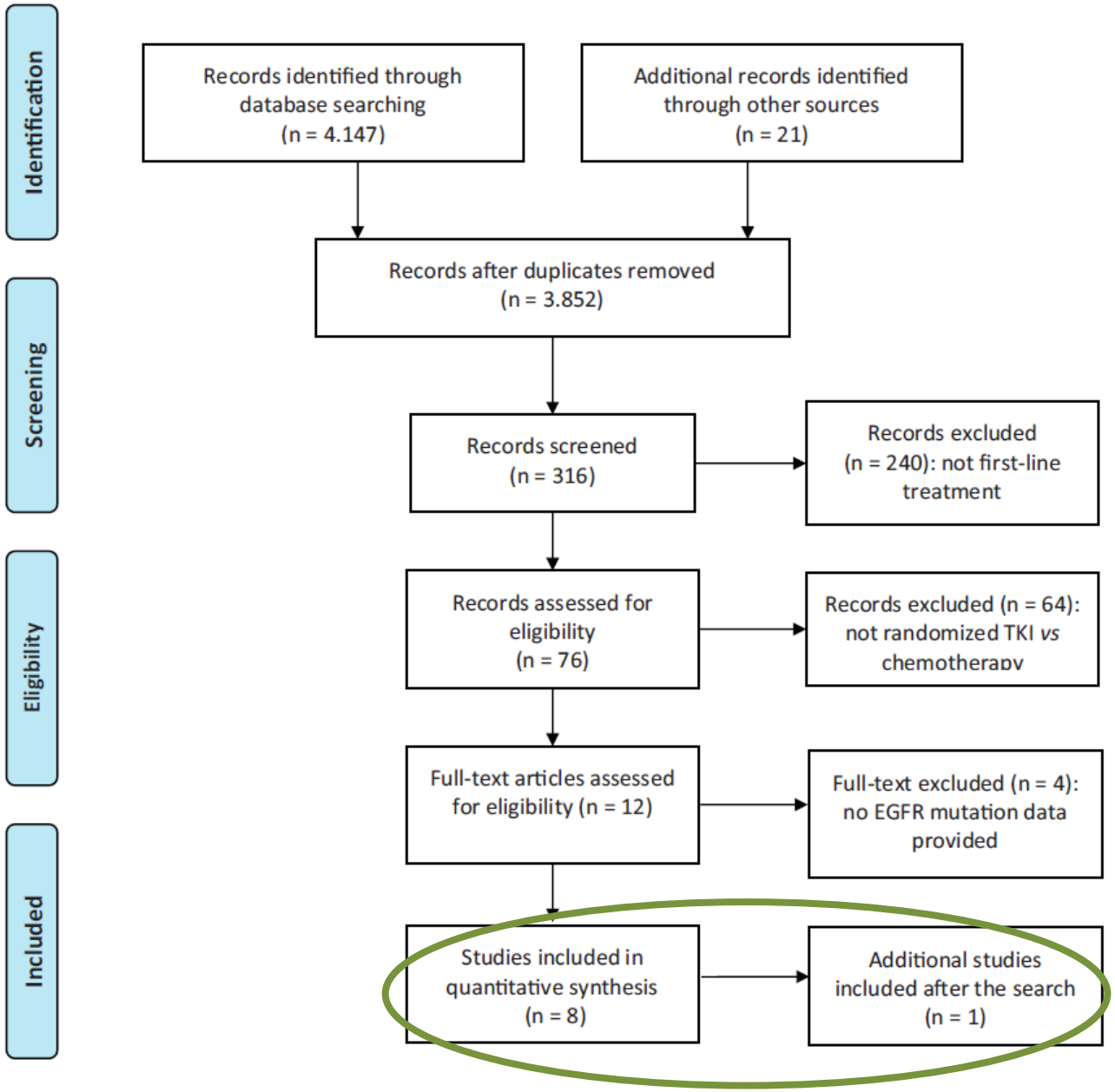


Fig. 1. Flow diagram for the selection of studies included in this meta-analysis.
From: Moher D, Liberati A, Telzlaff J, Altman DG, The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the Prisma statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097.

Indirect Comparisons

Basic assumptions underlying indirect comparisons include:

- ✓ **homogeneity** assumption for standard meta-analysis,
- ✓ **similarity** assumption for adjusted indirect comparison and
- ✓ **consistency** assumption for the combination of direct and indirect evidence. It is essential to fully understand and appreciate these basic assumptions in order to use adjusted indirect and mixed treatment comparisons appropriately.

What is indirect comparison? Fujian Song BMed MMed PhD Reader in Research Synthesis, Faculty of Health, University of East Anglia
www.whatisseries.co.uk http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/What_is_ind_comp.pdf

HOMOGENEITY ASSUMPTION

- When multiple trials are available for a given comparison, the results from multiple trials can be pooled in meta-analyses before an adjusted indirect comparison is conducted.
- For a meta-analysis to be valid, it is commonly established that results from different trials should be sufficiently homogeneous from a clinical and statistical perspective.
- This is usually demonstrated by a 2-tailed p value for homogeneity at Pearson chi-squared test or Cochran Q test > 0.10 and a I^2 (inconsistency) < 50%.
- When homogeneity is unlikely (e.g. I^2>50%) than heterogeneity and inconsistency are likely.

Song, What is …? 2009; Higgins et al, BMJ 2003

Data synthesis:

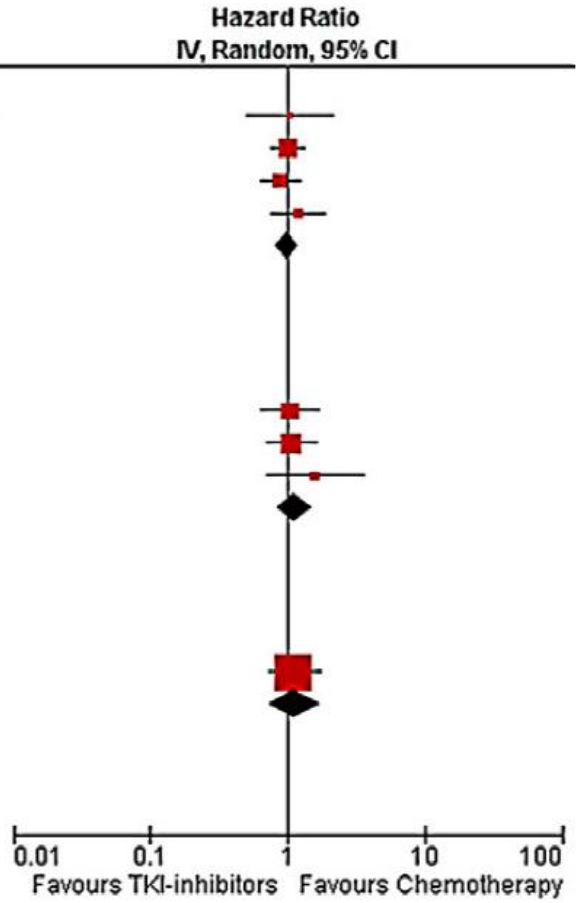
- ✓ HR for OS and PFS
- ✓ RR for the Others

OS

Panel B

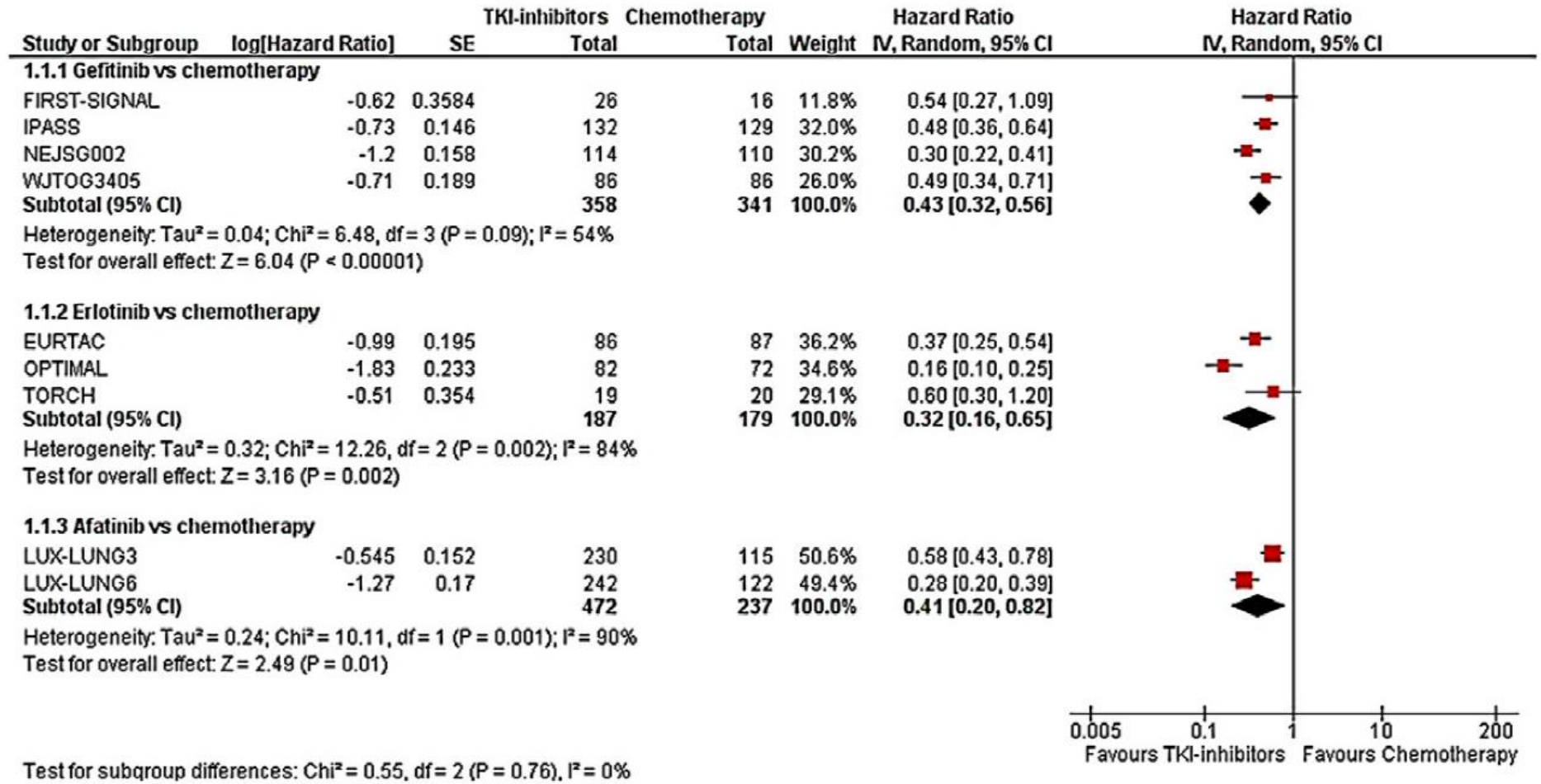
Study or Subgroup	log[Hazard Ratio]	SE	TKI-inhibitors Total	Chemotherapy Total	Weight	Hazard Ratio IV, Random, 95% CI
1.2.1 Gefitinib vs chemotherapy						
FIRST-SIGNAL	0.0392	0.3755	26	16	6.4%	1.04 [0.50, 2.17]
IPASS	0	0.143	132	129	44.3%	1.00 [0.76, 1.32]
NEJSG002	-0.12	0.171	114	110	31.0%	0.89 [0.63, 1.24]
WJTOG3405	0.17	0.223	86	86	18.2%	1.19 [0.77, 1.84]
Subtotal (95% CI)			**358**	**341**	**100.0%**	**1.00 [0.83, 1.20]**
Heterogeneity: Tau² = 0.00; Chi² = 1.08, df = 3 (P = 0.78); I² = 0%						
Test for overall effect: Z = 0.04 (P = 0.97)						
1.2.2 Erlotinib vs chemotherapy						
EURTAC	0.039	0.24	86	87	39.5%	1.04 [0.65, 1.66]
OPTIMAL	0.0677	0.219	82	72	47.4%	1.07 [0.70, 1.64]
TORCH	0.457	0.416	19	20	13.1%	1.58 [0.70, 3.57]
Subtotal (95% CI)			**187**	**179**	**100.0%**	**1.11 [0.83, 1.50]**
Heterogeneity: Tau² = 0.00; Chi² = 0.82, df = 2 (P = 0.66); I² = 0%						
Test for overall effect: Z = 0.71 (P = 0.48)						
1.2.3 Afatinib						
LUX-LUNG3	0.11	0.22	230	115	100.0%	1.12 [0.73, 1.72]
Subtotal (95% CI)			**230**	**115**	**100.0%**	**1.12 [0.73, 1.72]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.50 (P = 0.62)						

Test for subgroup differences: Chi² = 0.51, df = 2 (P = 0.77), I² = 0%

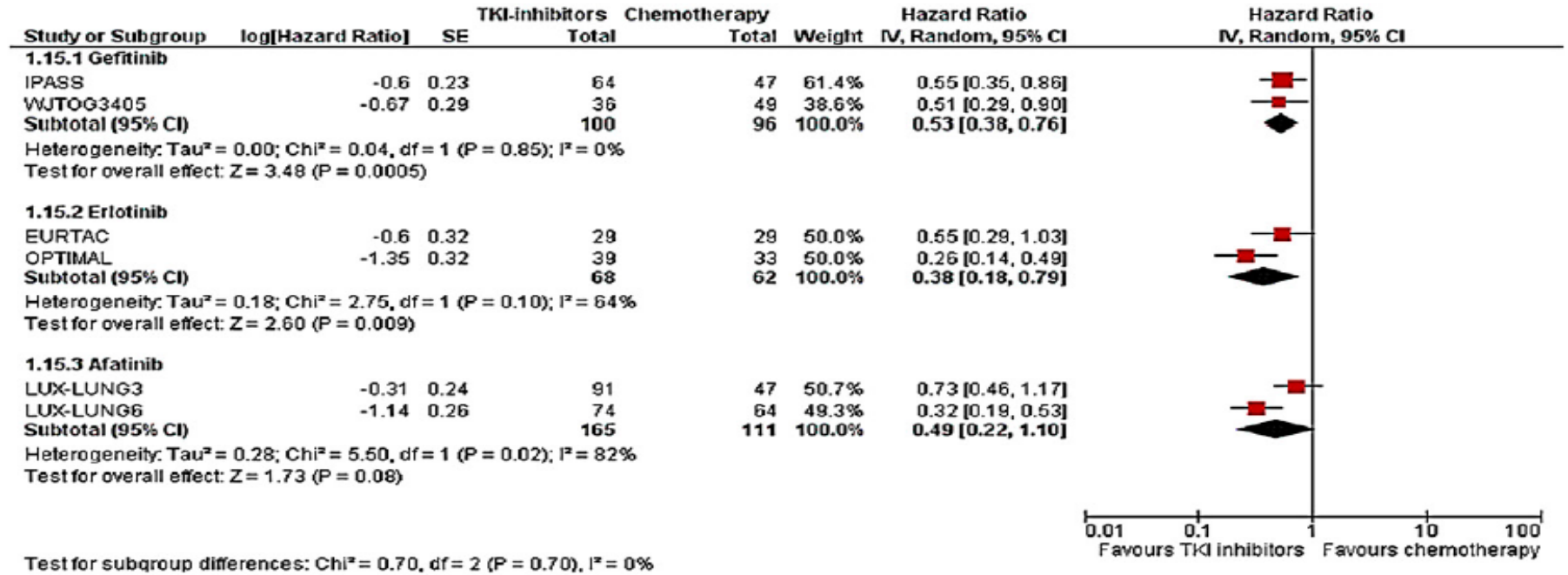


PFS

Panel A

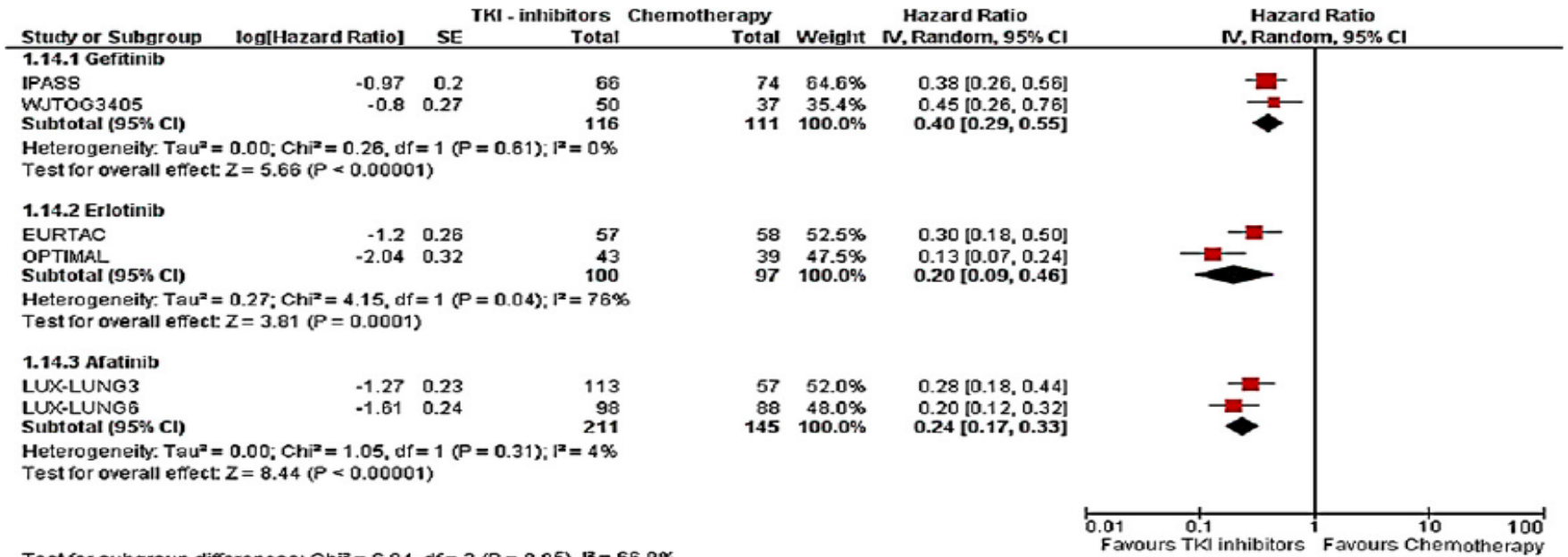

Study or Subgroup log[Hazard Ratio] SE TKI-inhibitors Total Chemotherapy Total Weight Hazard Ratio IV, Random, 95% CI
1.1.1 Gefitinib vs chemotherapy
FIRST-SIGNAL -0.62 0.3584 26 16 11.8% 0.54 [0.27, 1.09]
IPASS -0.73 0.146 132 129 32.0% 0.48 [0.36, 0.64]
NEJSG002 -1.2 0.158 114 110 30.2% 0.30 [0.22, 0.41]
WJTOG3405 -0.71 0.189 86 86 26.0% 0.49 [0.34, 0.71]
Subtotal (95% CI) 358 341 100.0% 0.43 [0.32, 0.56]
Heterogeneity: Tau² = 0.04; Chi² = 6.48, df = 3 (P = 0.09); I² = 54%
Test for overall effect: Z = 6.04 (P < 0.00001)
1.1.2 Erlotinib vs chemotherapy
EURTAC -0.99 0.195 86 87 36.2% 0.37 [0.25, 0.54]
OPTIMAL -1.83 0.233 82 72 34.6% 0.16 [0.10, 0.25]
TORCH -0.51 0.354 19 20 29.1% 0.60 [0.30, 1.20]
Subtotal (95% CI) 187 179 100.0% 0.32 [0.16, 0.65]
Heterogeneity: Tau² = 0.32; Chi² = 12.26, df = 2 (P = 0.002); I² = 84%
Test for overall effect: Z = 3.16 (P = 0.002)
1.1.3 Afatinib vs chemotherapy
LUX-LUNG3 -0.545 0.152 230 115 50.6% 0.58 [0.43, 0.78]
LUX-LUNG6 -1.27 0.17 242 122 49.4% 0.28 [0.20, 0.39]
Subtotal (95% CI) 472 237 100.0% 0.41 [0.20, 0.82]
Heterogeneity: Tau² = 0.24; Chi² = 10.11, df = 1 (P = 0.001); I² = 90%
Test for overall effect: Z = 2.49 (P = 0.01)
Test for subgroup differences: Chi² = 0.55, df = 2 (P = 0.76), I² = 0%
Hazard Ratio IV, Random, 95% CI
0.005
0.1
1
10
200
Favours TKI-inhibitors
Favours Chemotherapy

Exon 21

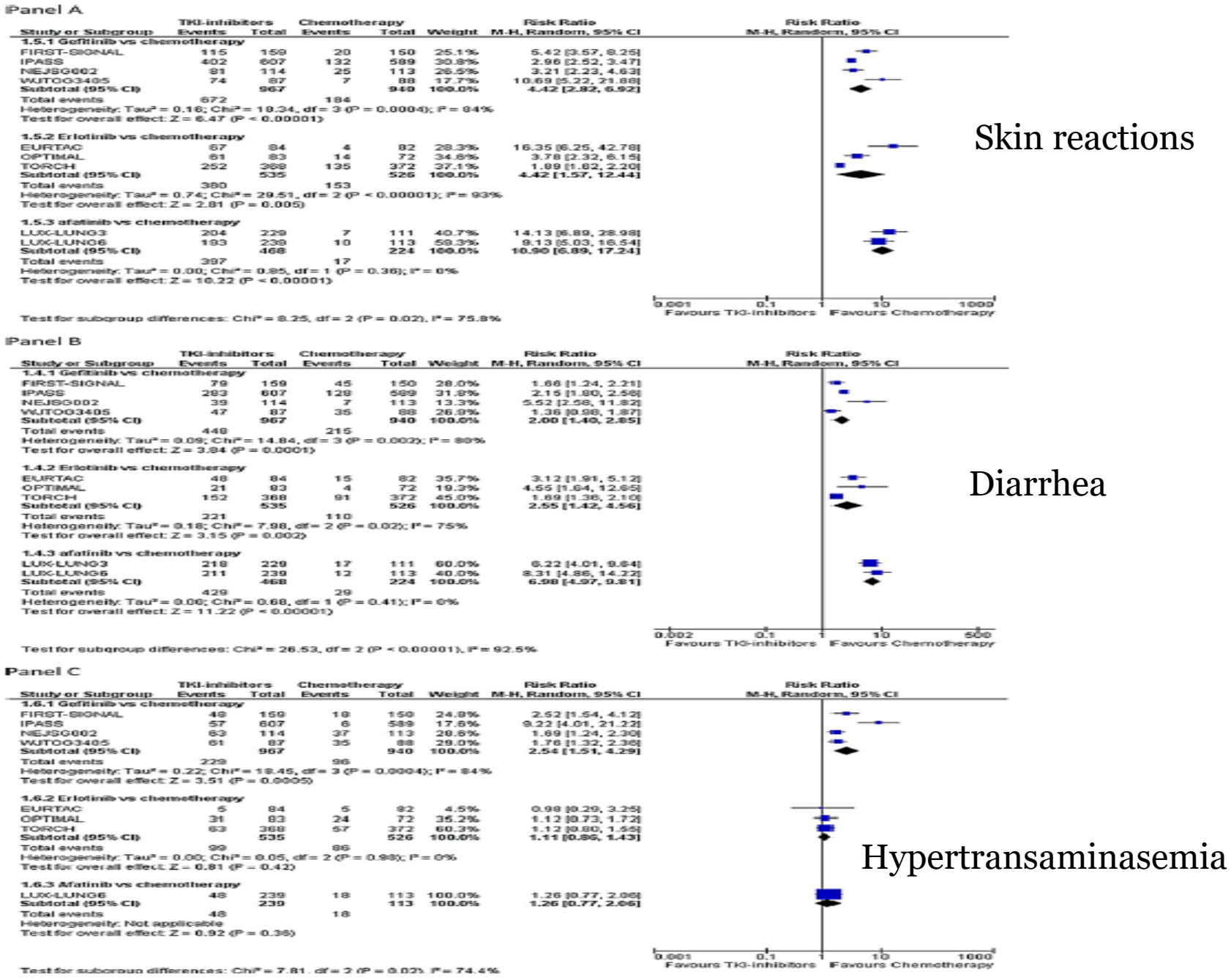


Test for subgroup differences: Chi² = 0.70, df = 2 (P = 0.70), I² = 0%

Exon 19



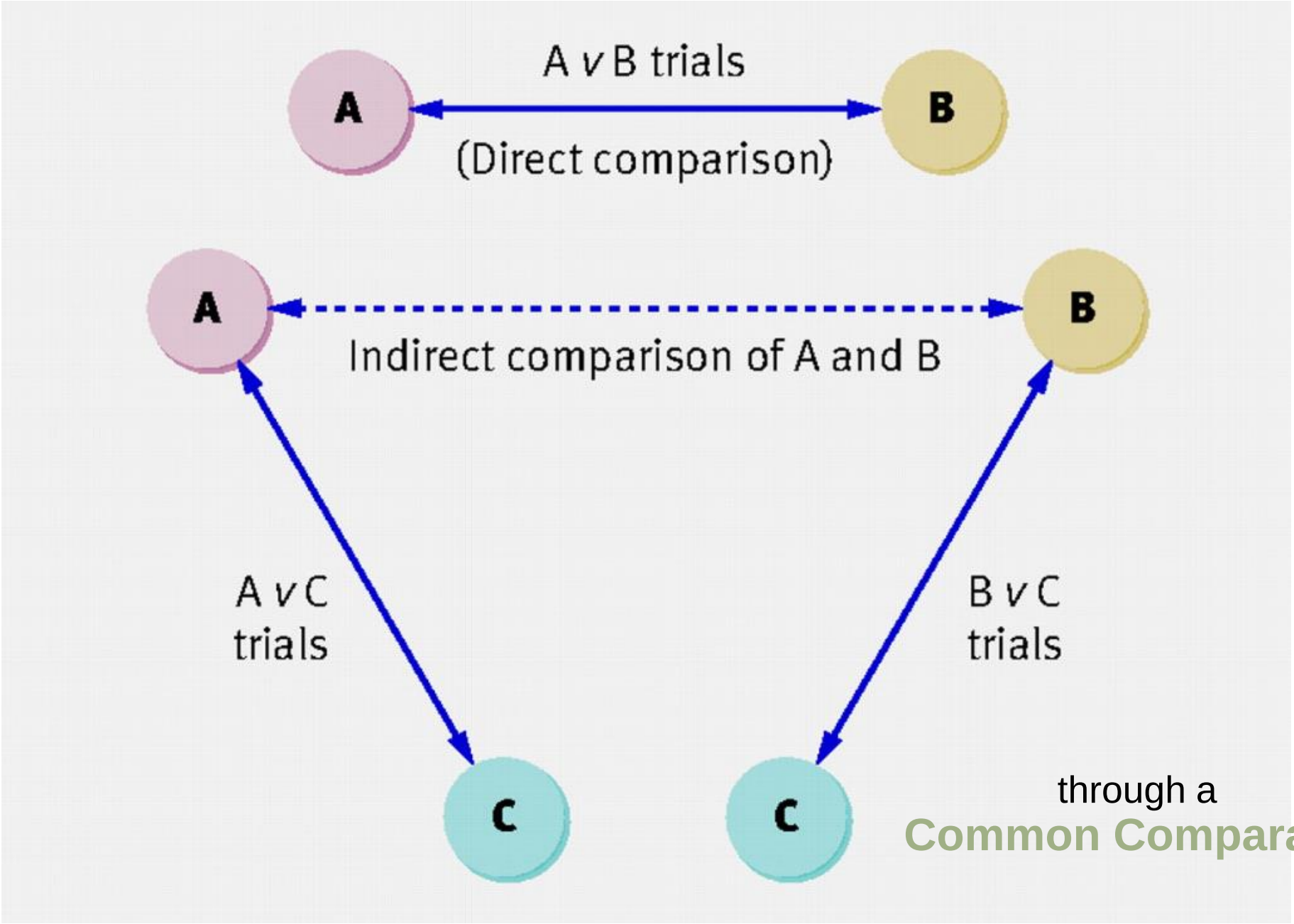
Test for subgroup differences: Chi² = 6.04, df = 2 (P = 0.05), I² = 66.9%


Panel A
TKI-inhibitors
Chemotherapy
Risk Ratio
Risk Ratio
Study or Subgroup Events Total Events Total Weight M-H, Random, 95% CI
M-H, Random, 95% CI
1.5.1 Gefitinib vs chemotherapy
FIRST-SIGNAL 115 159 20 150 25.1% 5.42 [3.57, 8.25]
IPASS 402 607 132 589 30.8% 2.96 [2.52, 3.47]
NEJSG002 81 114 25 113 26.5% 3.21 [2.23, 4.63]
WJTOG3405 74 87 7 88 17.7% 10.69 [5.22, 21.88]
Subtotal (95% CI) 967 940 100.0% 4.42 [2.82, 6.92]
Total events 672 184
Heterogeneity: Tau² = 0.16; Chi² = 19.24, df = 3 (P = 0.0004); I² = 84%
Test for overall effect: Z = 6.47 (P < 0.00001)
1.5.2 Erlotinib vs chemotherapy
EURTAC 67 84 4 82 28.3% 16.35 [6.25, 42.78]
OPTIMAL 61 83 14 72 34.6% 3.78 [2.32, 6.15]
TORCH 252 368 135 372 37.1% 1.89 [1.62, 2.20]
Subtotal (95% CI) 535 526 100.0% 4.42 [1.57, 12.44]
Total events 380 153
Heterogeneity: Tau² = 0.74; Chi² = 29.51, df = 2 (P < 0.00001); I² = 93%
Test for overall effect: Z = 2.81 (P = 0.005)
1.5.3 afatinib vs chemotherapy
LUX-LUNG3 204 229 7 111 40.7% 14.13 [6.89, 28.98]
LUX-LUNG6 193 239 10 113 59.3% 9.13 [5.03, 16.54]
Subtotal (95% CI) 468 224 100.0% 10.90 [6.89, 17.24]
Total events 397 17
Heterogeneity: Tau² = 0.00; Chi² = 0.85, df = 1 (P = 0.36); I² = 0%
Test for overall effect: Z = 10.22 (P < 0.00001)
Test for subgroup differences: Chi² = 8.25, df = 2 (P = 0.02), I² = 75.8%
0.001 0.1 1 10 1000
Favours TKI-inhibitors Favours Chemotherapy
Skin reactions
Panel B
TKI-inhibitors
Chemotherapy
Risk Ratio
Risk Ratio
Study or Subgroup Events Total Events Total Weight M-H, Random, 95% CI
M-H, Random, 95% CI
1.4.1 Gefitinib vs chemotherapy
FIRST-SIGNAL 79 159 45 150 28.0% 1.66 [1.24, 2.21]
IPASS 283 607 128 589 31.8% 2.15 [1.80, 2.56]
NEJSG002 39 114 7 113 13.3% 5.52 [2.58, 11.82]
WJTOG3405 47 87 35 88 26.9% 1.36 [0.98, 1.87]
Subtotal (95% CI) 967 940 100.0% 2.00 [1.40, 2.85]
Total events 448 215
Heterogeneity: Tau² = 0.09; Chi² = 14.84, df = 3 (P = 0.002); I² = 80%
Test for overall effect: Z = 3.94 (P = 0.0001)
1.4.2 Erlotinib vs chemotherapy
EURTAC 48 84 15 82 35.7% 3.12 [1.91, 5.12]
OPTIMAL 21 83 4 72 19.3% 4.55 [1.64, 12.65]
TORCH 152 368 91 372 45.0% 1.69 [1.36, 2.10]
Subtotal (95% CI) 535 526 100.0% 2.55 [1.42, 4.56]
Total events 221 110
Heterogeneity: Tau² = 0.18; Chi² = 7.98, df = 2 (P = 0.02); I² = 75%
Test for overall effect: Z = 3.15 (P = 0.002)
1.4.3 afatinib vs chemotherapy
LUX-LUNG3 218 229 17 111 60.0% 6.22 [4.01, 9.64]
LUX-LUNG6 211 239 12 113 40.0% 8.31 [4.86, 14.22]
Subtotal (95% CI) 468 224 100.0% 6.98 [4.97, 9.81]
Total events 429 29
Heterogeneity: Tau² = 0.00; Chi² = 0.68, df = 1 (P = 0.41); I² = 0%
Test for overall effect: Z = 11.22 (P < 0.00001)
Test for subgroup differences: Chi² = 26.53, df = 2 (P < 0.00001), I² = 92.5%
0.002 0.1 1 10 500
Favours TKI-inhibitors Favours Chemotherapy
Diarrhea
Panel C
TKI-inhibitors
Chemotherapy
Risk Ratio
Risk Ratio
Study or Subgroup Events Total Events Total Weight M-H, Random, 95% CI
M-H, Random, 95% CI
1.6.1 Gefitinib vs chemotherapy
FIRST-SIGNAL 48 159 19 150 24.8% 2.52 [1.54, 4.12]
IPASS 57 607 6 589 17.6% 9.22 [4.01, 21.22]
NEJSG002 63 114 37 113 28.6% 1.69 [1.24, 2.30]
WJTOG3405 61 87 35 88 29.0% 1.76 [1.32, 2.36]
Subtotal (95% CI) 967 940 100.0% 2.54 [1.51, 4.29]
Total events 229 96
Heterogeneity: Tau² = 0.22; Chi² = 18.45, df = 3 (P = 0.0004); I² = 84%
Test for overall effect: Z = 3.51 (P = 0.0005)
1.6.2 Erlotinib vs chemotherapy
EURTAC 5 84 5 82 4.5% 0.98 [0.29, 3.25]
OPTIMAL 31 83 24 72 35.2% 1.12 [0.73, 1.72]
TORCH 63 368 57 372 60.3% 1.12 [0.80, 1.56]
Subtotal (95% CI) 535 526 100.0% 1.11 [0.86, 1.43]
Total events 99 86
Heterogeneity: Tau² = 0.00; Chi² = 0.05, df = 2 (P = 0.98); I² = 0%
Test for overall effect: Z = 0.81 (P = 0.42)
1.6.3 Afatinib vs chemotherapy
LUX-LUNG6 48 239 18 113 100.0% 1.26 [0.77, 2.06]
Subtotal (95% CI) 239 113 100.0% 1.26 [0.77, 2.06]
Total events 48 18
Heterogeneity: Not applicable
Test for overall effect: Z = 0.92 (P = 0.36)
Test for subgroup differences: Chi² = 7.81, df = 2 (P = 0.02), I² = 74.4%
0.001 0.1 1 10 1000
Favours TKI-inhibitors Favours Chemotherapy
Hypertransaminasemia

SIMILARITY (TRANSITIVITY) ASSUMPTION

- For an adjusted indirect comparison (A vs B) to be valid, a similarity assumption is required in terms of moderators of relative treatment effect.
- That is, patients included should be sufficiently similar in the two sets of control arms (C_1 from the trial comparing A vs C_1, and C_2, from the trial comparing B vs C_2).
- This is crucial as only a large theoretical overlap between patients enrolled in C_1 and C_2 enables the relative effect estimated by trials of A versus C_1 to be generalizable to patients in trials of B versus C_1, and the relative effect estimated by trials of B versus C_2 to be generalizable to patients in trials of A versus C_2.

Song, What is …? 2009



through a

Common Comparator

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.

Table 1
Characteristics of the 9 clinical trials included in the meta-analysis.

Trial	Primary end-point	TKI	Chemotherapy	Patients (TKI/CT)	EGFR + patients (%)	Asiatic patients (%)	Crossover (%)[a]
IPASS Mok, 2009	Progression-free survival	Gefitinib	Carboplatin + paclitaxel	1.217 (609/608)	21.4	99.8	39.5
WJTOG3405 Mitsudomi, 2010	Progression-free survival	Gefitinib	Cisplatin + paclitaxel	177 (88/89)	100	100	59.3
NEJ002 Maemondo, 2010	Progression-free survival	Gefitinib	Carboplatin + paclitaxel	228 (114/114)	100	100	94.6
First-SIGNAL Han, 2012	Overall survival	Gefitinib	Cisplatin + gemcitabine	309 (159/150)	13.6	100	75.0
TORCH Gridelli, 2012	Overall survival	Erlotinib	Cisplatin + gemcitabine	760 (380/380)	5.1	0	60.9
OPTIMAL Zhou, 2011	Progression-free survival	Erlotinib	Carboplatin + gemcitabine	154 (82/72)	100	100	NA
EURTAC Rosell, 2011	Progression-free survival	Erlotinib	Cisplatin/carboplatin + docetaxel/gemcitabine	173 (86/87)	100	0	76.0
LUX-Lung 3 Sequist, 2012	Progression-free survival	Afatinib	Cisplatin + pemetrexed	345 (230/115)	100	100	75.0
LUX-Lung 6 Wu, 2013	Progression-free survival	Afatinib	Cisplatin + gemcitabine	364 (242/122)	100	100	56.0

[a] Patients who have been treated with crossover from chemotherapy to TKI in second-line.

Study		
FIRST-SIGNAL	Cisplatin 75 mg/m2 Gemcitabine 1	ery 3 weeks ycles
IPASS	Carboplati mg/mi Pac	veeks up to 6 weeks
NEJG002		
WJTOG3405		to 6 weeks
EURTAC		
OPTI		.v. 4 cycles
TORC	g/m2	i.v. every 3 weeks up to 6 weeks
LUX-LUNG I	m2 Pemetrexed	i.v. 6 cycles
LUX-LUNG VI	75 mg/m2 Gemcitabine mg/m2 day 1&8	i.v. Up to 6 cycles

STANDARD CHEMOTHERAPY

So, who's the best?

COMPUTATIONS

- The log relative risk of the adjusted indirect comparison of A and B ($\ln RR_{A \text{ vs } B}$) can be estimated by:

$$\ln RR_{A \text{ vs } B} = \ln RR_{A \text{ vs } C1} - \ln RR_{B \text{ vs } C2}$$

- and its standard error is:

$$SE(\ln RR_{A \text{ vs } B}) = \sqrt{[SE(\ln RR_{A \text{ vs } C1})^2 + SE(\ln RR_{B \text{ vs } C2})^2]}$$

- Similar computations can be envisioned for odds ratio, absolute risk reductions, weighted mean differences, and standardized mean differences.

Higgins et al, BMJ 2003; Song, What is …? 2009;
http://www.metcardio.org/macros/IMT.xls

Panel A

Study or Subgroup	log[Hazard Ratio\Risk ratio]	SE	Weight	Hazard Ratio\Risk ratio IV, Random, 95% CI
Progression-free survival	0.295	0.385		1.34 [0.63, 2.86]
PFS-exon 19	0.693	0.447		2.00 [0.83, 4.80]
PFS-L858R	0.332	0.417		1.39 [0.62, 3.16]
Overall survival	-0.104	0.177		0.90 [0.64, 1.27]
Objective response rate	-0.036	0.168		0.96 [0.69, 1.34]
Diarrhea	-0.223	0.121		0.80 [0.63, 1.01]
Rash	0	0.101		1.00 [0.82, 1.22]
Hypertransaminasemia	0.83	0.175		2.29 [1.63, 3.23]
Treatment discontinuation	-0.019	0.384		0.98 [0.46, 2.08]
Treatment-related death	1.05	1.295		2.86 [0.23, 36.17]

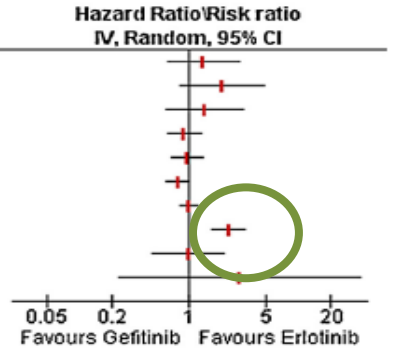
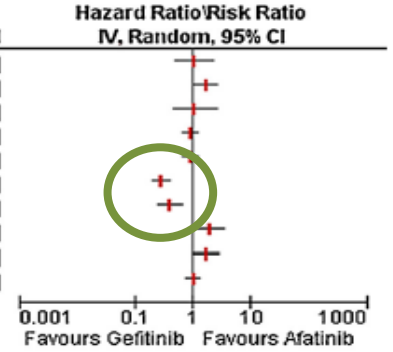


Image of Fig. 5

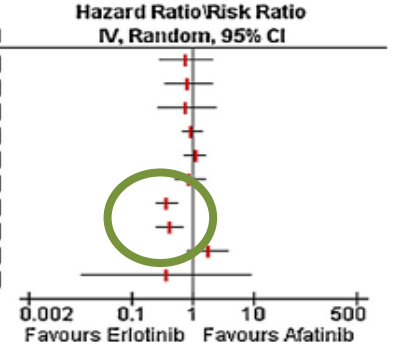
Panel B

Study or Subgroup	log[Hazard Ratio\Risk Ratio]	SE	Weight	Hazard Ratio\Risk Ratio IV, Random, 95% CI
Progression-free survival	0.048	0.387		1.05 [0.49, 2.24]
PFS-exon 19	0.511	0.235		1.67 [1.05, 2.64]
PFS-L858R	0.078	0.447		1.08 [0.45, 2.60]
Overall survival	-0.099	0.167		0.91 [0.65, 1.26]
Objective response rate	-0.097	0.157		0.91 [0.67, 1.23]
Diarrhea	-1.25	0.187		0.29 [0.20, 0.41]
Rash	-0.903	0.244		0.41 [0.25, 0.65]
Hypertransaminasemia	0.701	0.276		2.02 [1.17, 3.46]
Treatment discontinuation	0.531	0.273		1.70 [1.00, 2.90]
Treatment-related death	0.022	0.136		1.02 [0.78, 1.33]



Panel C

Study or Subgroup	log[Hazard Ratio\Risk Ratio]	SE	Weight	Hazard Ratio\Risk Ratio IV, Random, 95% CI
Progression-free survival	-0.248	0.507		0.78 [0.29, 2.11]
PFS-exon 19	-0.182	0.449		0.83 [0.35, 2.01]
PFS-L858R	-0.254	0.558		0.78 [0.26, 2.32]
Objective response rate	-0.061	0.186		0.94 [0.65, 1.35]
Overall survival	0.094	0.204		1.10 [0.74, 1.64]
Hypertransaminasemia	-0.127	0.285		0.88 [0.50, 1.54]
Diarrhea	-1.01	0.2		0.36 [0.25, 0.54]
Rash	-0.903	0.245		0.41 [0.25, 0.66]
Treatment discontinuation	0.55	0.395		1.73 [0.80, 3.76]
Treatment-related death	-1.03	1.637		0.36 [0.01, 8.83]



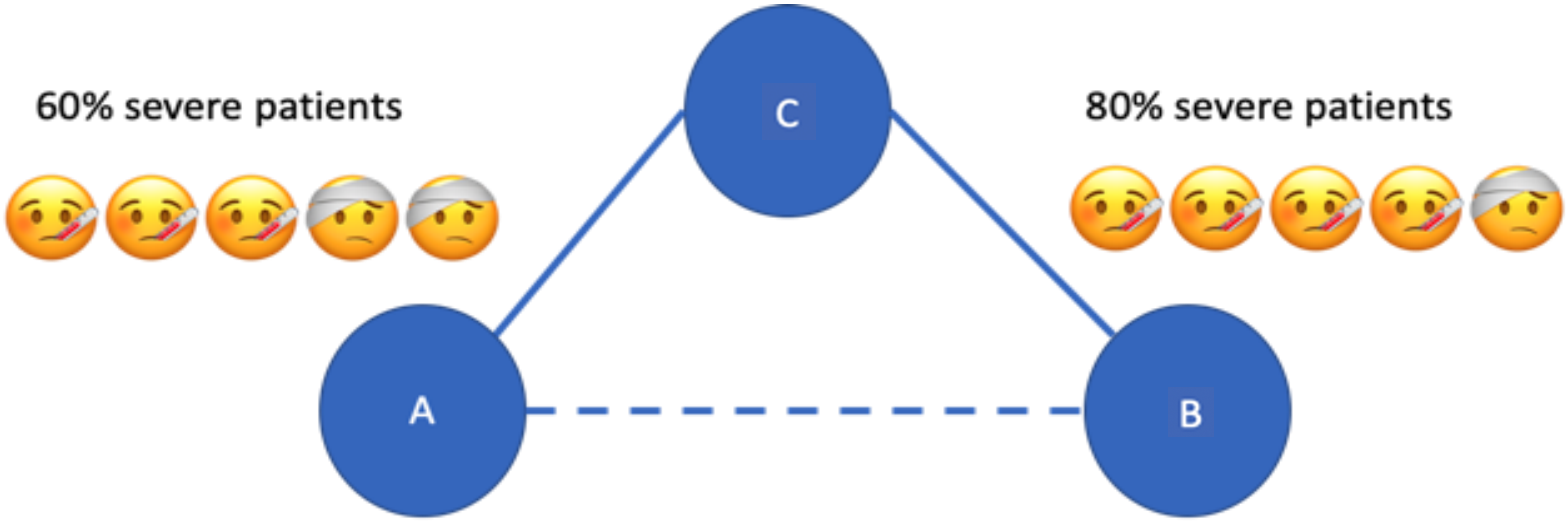
TAKE HOME MESSAGES

- Adjusted indirect comparison meta-analysis represents a simple yet robust tool to make statistical and clinical inference despite the lack of conclusive evidence from head-to-head randomized clinical trials.

- Despite being not at the uppermost level of the hierarchy of evidence based medicine, it can often provide results equivalent to those of subsequent direct comparisons.

Population-Adjusted Indirect Comparison

(G.L. Pappagallo)



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

Population-adjusted Indirect Comparisons

- Population-adjusted Indirect Comparisons use patient-level data from a trial of a given treatment (referred to as the index trial) to derive a comparison of outcomes with competing treatments, based on published information from similarly designed studies, after adjusting for differences in the characteristics of the populations.
- *in other words*: individual patient data (IPD) in one or more trials are used to adjust for between-trial differences in the distribution of variables that influence outcome
- "anchored" indirect comparison (common comparator arm in each trial) Vs "unanchored" indirect comparison (disconnected treatment network or single-arm studies)
  - *an unanchored comparison assumes that all effect modifiers and prognostic factors are accounted for*
- unanchored methods for population adjustment are problematic and should not be used when anchored methods can be applied

https://www.nicedsu.org.uk/wp-content/uploads/2017/05/Population-adjustment-TSD-FINAL.pdf.

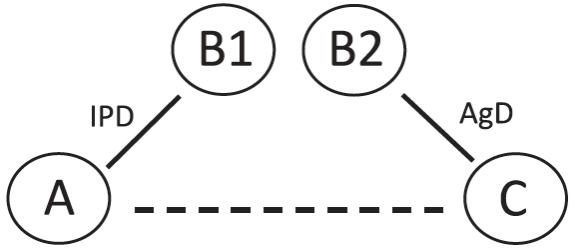
Population-adjusted Indirect Comparisons

- Population-adjusted Indirect Comparisons use patient-level data from a trial of a given treatment (referred to as the index trial) to derive a comparison of outcomes with competing treatments, based on published information from similarly designed studies, after adjusting for differences in the characteristics of the populations.
- *in other words*: individual patient data (IPD) in one or more trials are used to adjust for between-trial differences in the distribution of variables that influence outcome
- "anchored" indirect comparison (common comparator arm in each trial) Vs "unanchored" indirect comparison (disconnected treatment network or single-arm studies)
  - *an unanchored comparison assumes that all effect modifiers and prognostic factors are accounted for*
- unanchored methods for population adjustment are problematic and should not be used when anchored methods can be applied

https://www.nicedsu.org.uk/wp-content/uploads/2017/05/Population-adjustment-TSD-FINAL.pdf.

Matching-adjusted indirect comparison (MAIC)

- needs IPD for at least 1 trial, because the aim is to match the IPD to the AgD of the other trial



- the matching procedure selects a weight for each patient to reach similarity in the summary measures of the baseline characteristics of the IPD and AgD trial and follows the idea of *propensity score matching*
- the odds between being a patient in trial AB *Vs* trial CB provides the weights for balancing the populations

Comparative Effectiveness Research in Oncology Methodology: Observational Data

Dawn L. Hershman and Jason D. Wright

J Clin Oncol 30:4215-4222. © 2012 by American Society of Clinical Oncology

Propensity Score Analysis

Propensity score analyses attempt to balance covariates between experimental groups. Using multivariable modeling, the characteristics of a cohort are used to calculate the probability of receiving the intervention. This probability is the propensity score.

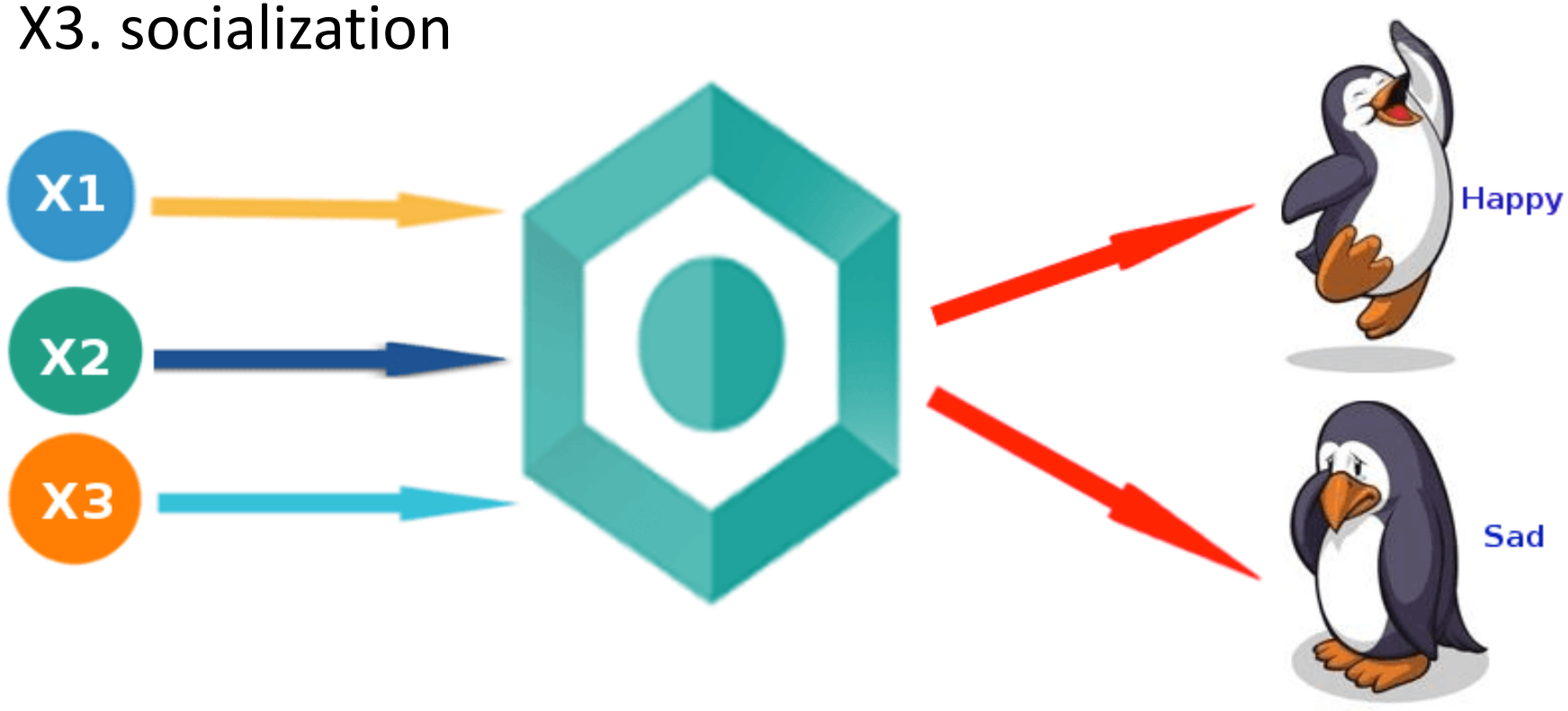
MULTIVARIATE ANALYSIS

Independent variables

X1. food
X2. water temperature
X3. socialization

Dependent variable

Penguin mood



Multivariate Behavioral Research, 46:399–424, 2011

An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies

Peter C. Austin

All measured baseline covariates, all baseline covariates that are associated with treatment assignment, all covariates that affect the outcome (i.e., the potential confounders), and all covariates that affect both treatment assignment and the outcome (i.e., the true confounders).

Propensity Score adjustment

Variabili indipendenti:
Fattori basali tali da influenzare la proposta terapeutica

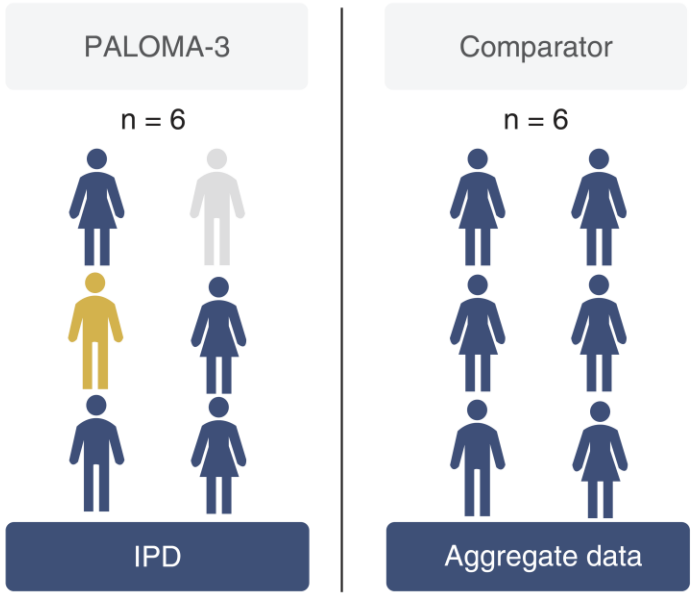
↓

Variabile dipendente:
Trattamento assegnato

Matching-adjusted indirect comparison of palbociclib versus ribociclib and abemaciclib in hormone receptor-positive/HER2-negative advanced breast cancer

Hope S Rugo*,[1], Anja Haltner[2], Lin Zhan[3], Anh Tran[2], Eustratios Bananis[3], Becky Hooper[2], Debanjali Mitra[3] & Chris Cameron[2]

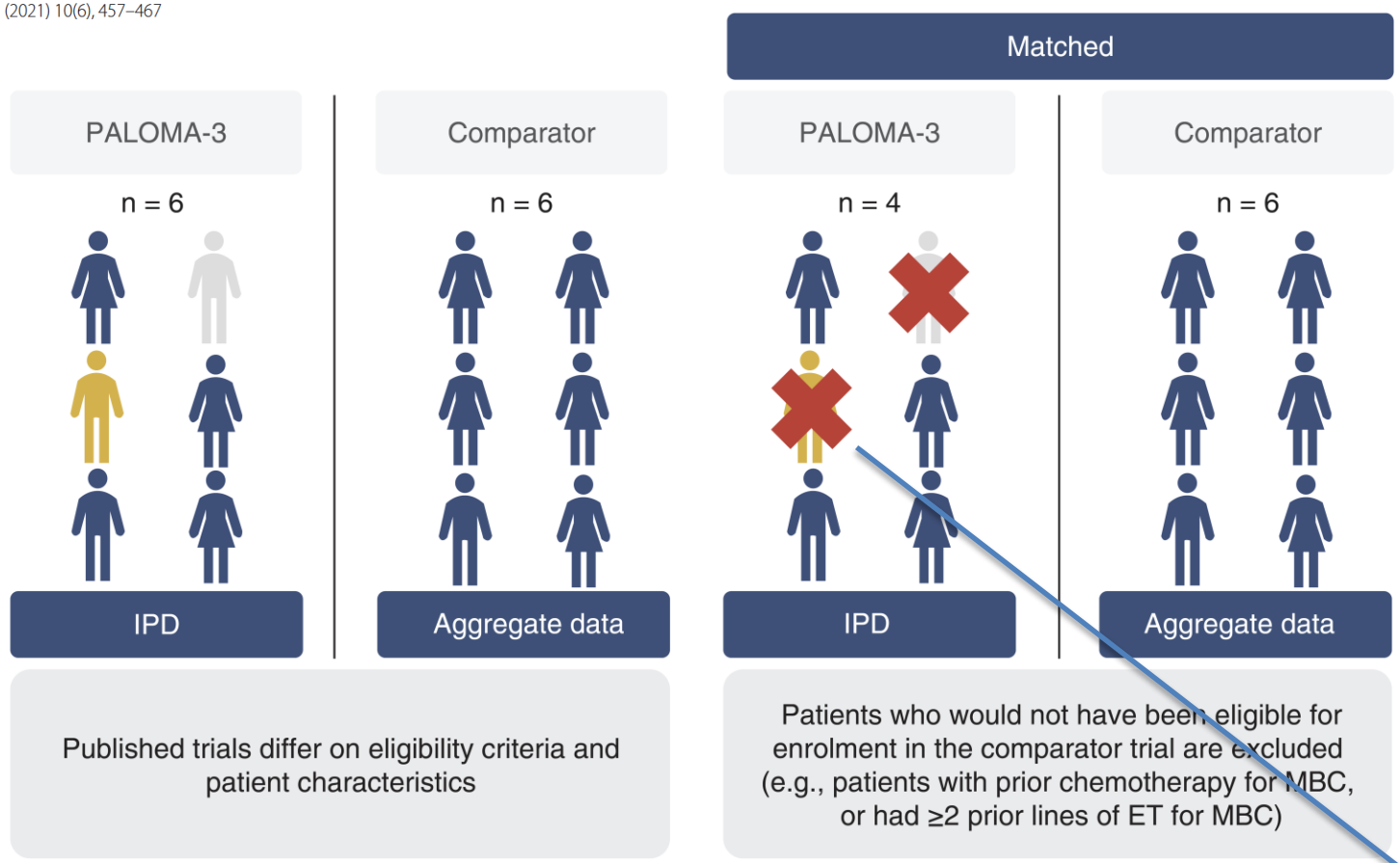
J. Comp. Eff. Res. (2021) 10(6), 457–467

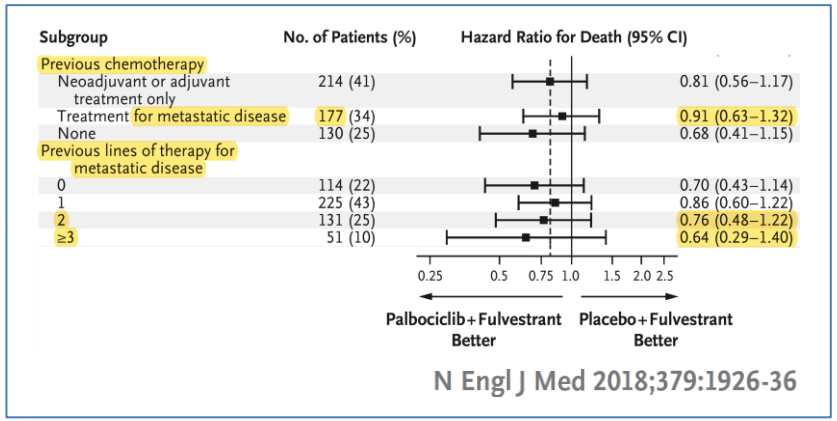


Matching-adjusted indirect comparison of palbociclib versus ribociclib and abemaciclib in hormone receptor-positive/HER2-negative advanced breast cancer

Hope S Rugo*,[1], Anja Haltner[2], Lin Zhan[3], Anh Tran[2], Eustratios Bananis[3], Becky Hooper[2], Debanjali Mitra[3] & Chris Cameron[2]

J. Comp. Eff. Res. (2021) 10(6), 457–467

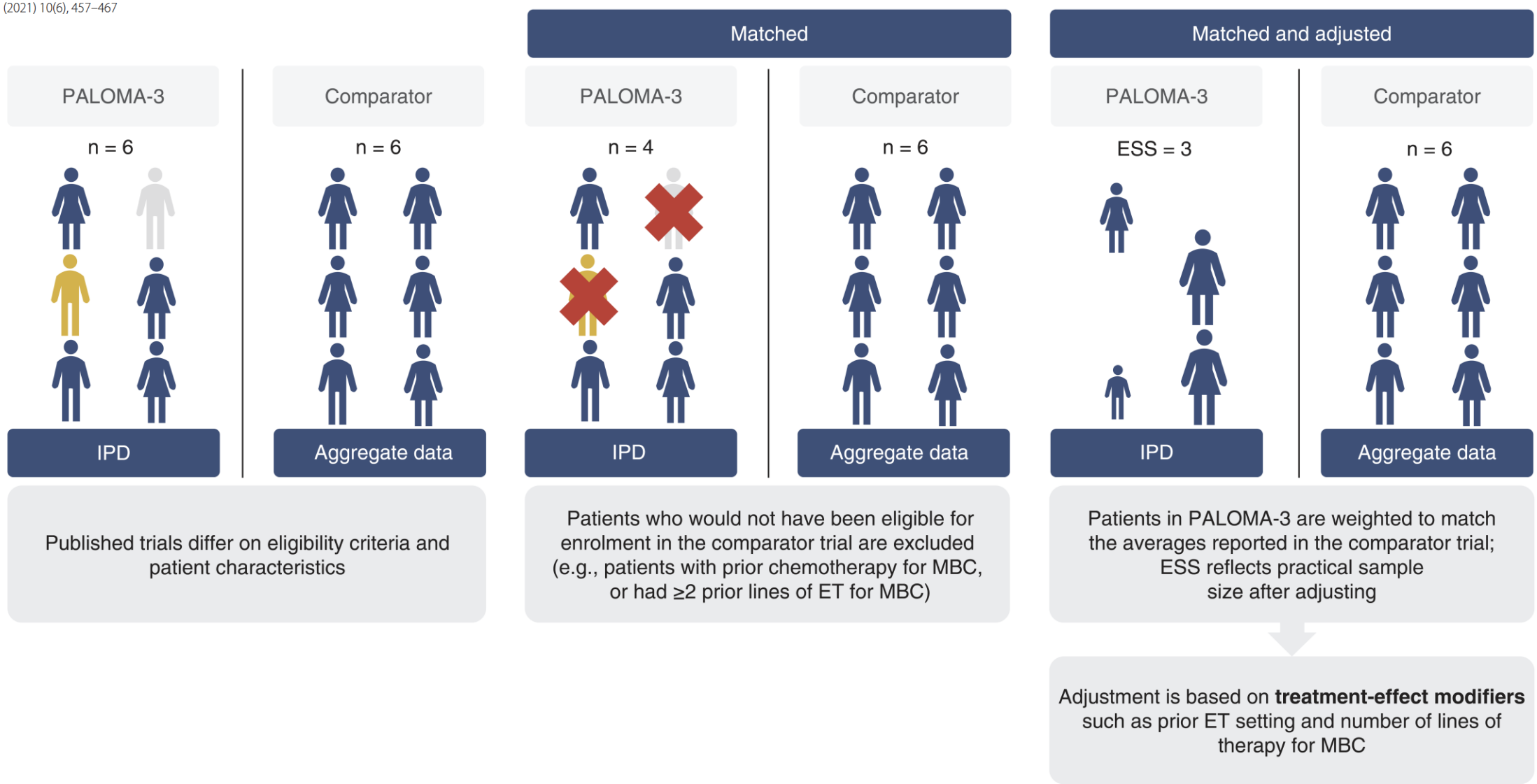


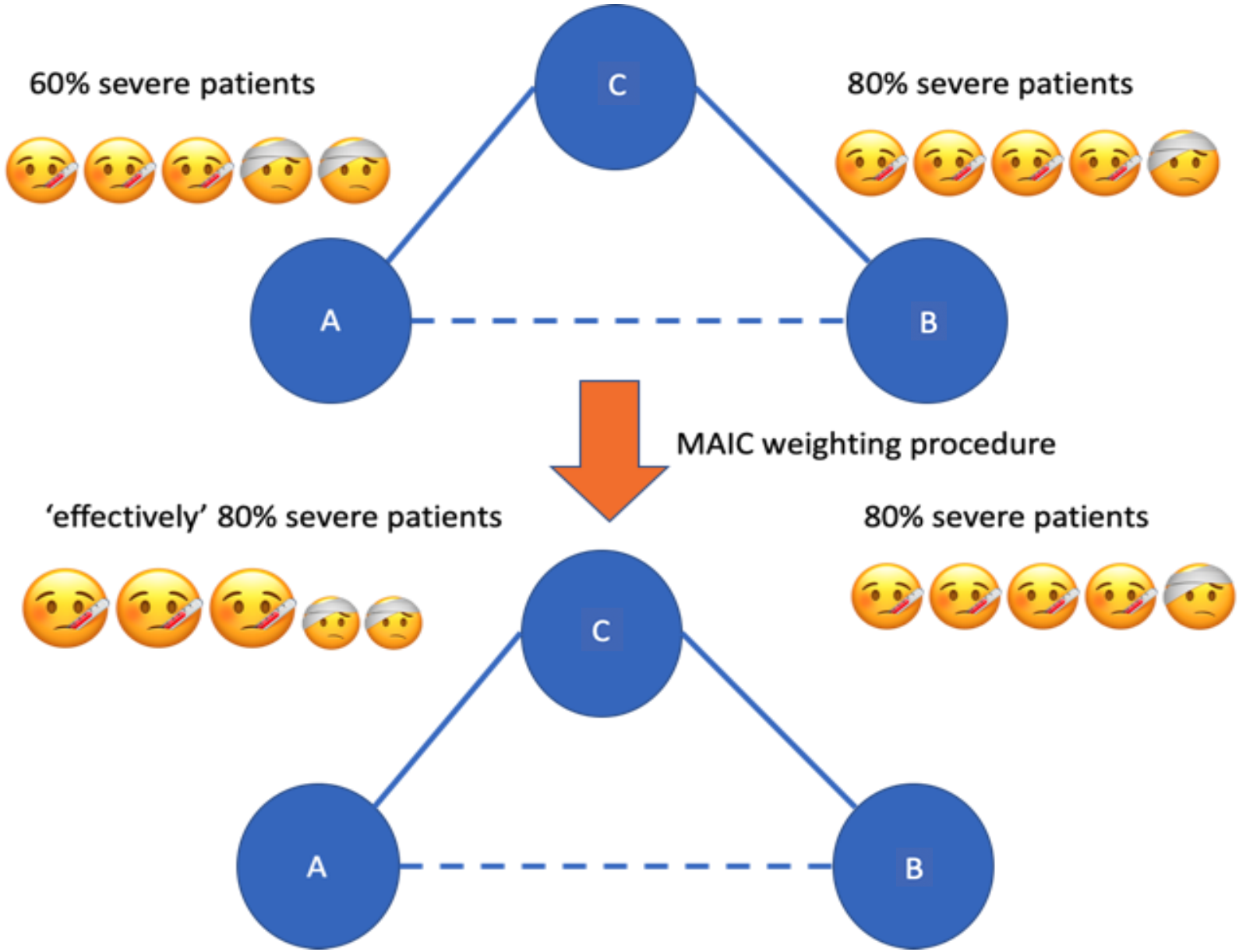


Matching-adjusted indirect comparison of palbociclib versus ribociclib and abemaciclib in hormone receptor-positive/HER2-negative advanced breast cancer

Hope S Rugo*,[1], Anja Haltner[2], Lin Zhan[3], Anh Tran[2], Eustratios Bananis[3], Becky Hooper[2], Debanjali Mitra[3] & Chris Cameron[2]

J. Comp. Eff. Res. (2021) 10(6), 457–467



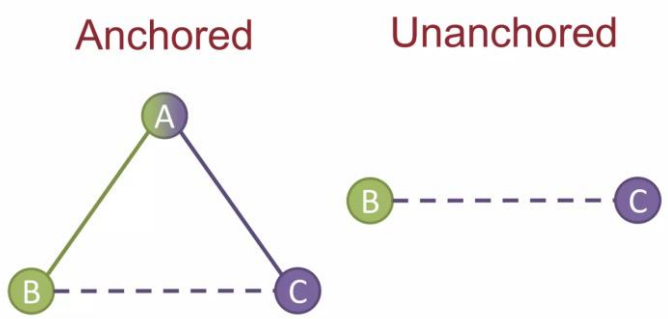

60% severe patients
C
80% severe patients
A
B
MAIC weighting procedure
'effectively' 80% severe patients
80% severe patients
C
A
B

The main limitation relates to the inherent challenge of MAIC in that it is only possible to adjust baseline variables that are mutually reported between trials, and therefore it cannot address the potential unmeasurable differences between the trials.

LEUKEMIA & LYMPHOMA
https://doi.org/10.1080/10428194.2024.2313628

Population-adjusted Indirect Comparisons

- MAICs and STCs use patient-level data from a trial of a given treatment (referred to as the index trial) to derive a comparison of outcomes with competing treatments, based on published information from similarly designed studies, after adjusting for differences in the characteristics of the populations.
- *in other words*: individual patient data (IPD) in one or more trials are used to adjust for between-trial differences in the distribution of variables that influence outcome
- "anchored" indirect comparison (common comparator arm in each trial) Vs "unanchored" indirect comparison (disconnected treatment network or single-arm studies)
  - *an unanchored MAIC or STC assumes that all effect modifiers and prognostic factors are accounted for*
- unanchored methods for population adjustment are problematic and should not be used when anchored methods can be applied



https://www.nicedsu.org.uk/wp-content/uploads/2017/05/Population-adjustment-TSD-FINAL.pdf.

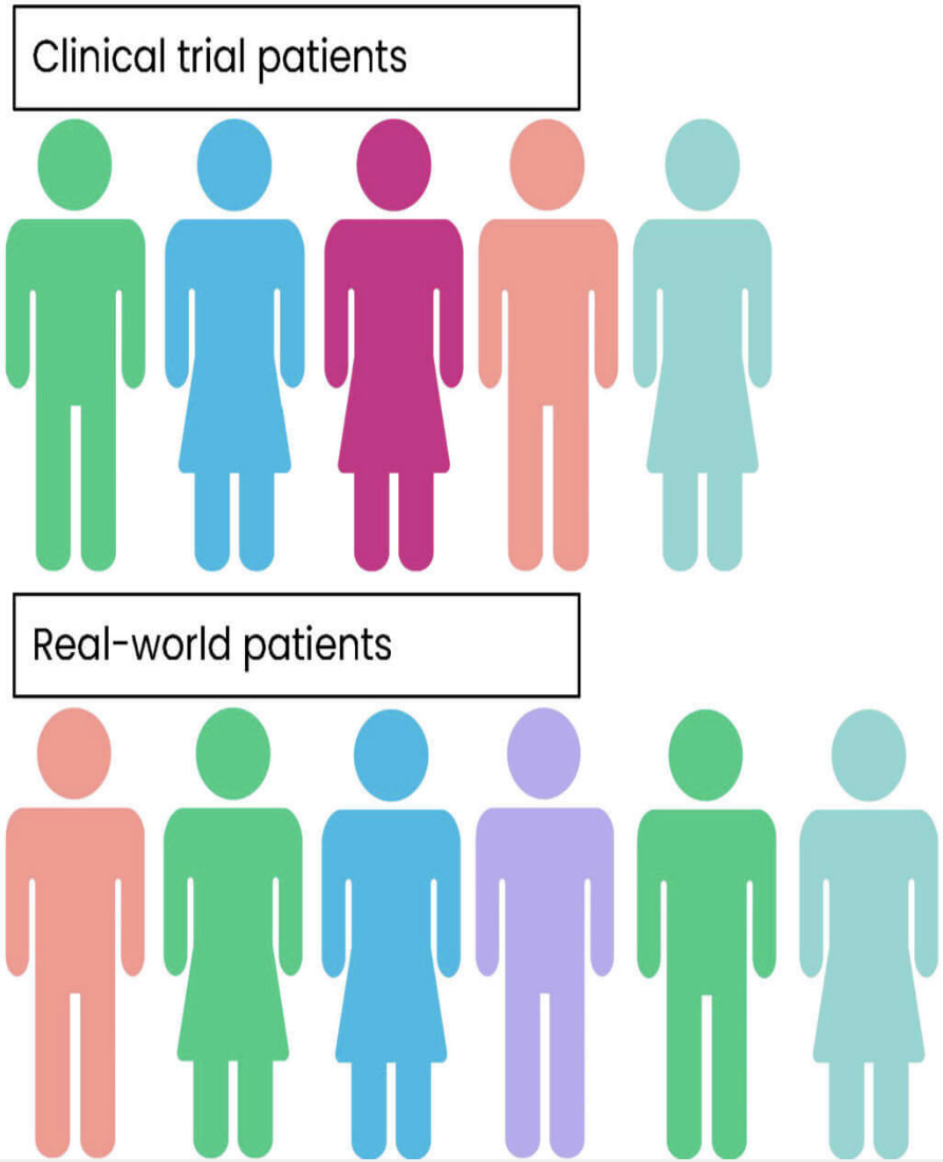
Elranatamab efficacy in MagnetisMM-3 compared with real-world control arms in triple-class refractory multiple myeloma

Luciano J Costa[1], Thomas W LeBlanc[2], Hans Tesch[3], Pieter Sonneveld[4], Ryan P Kyle[5], Liliya Sinyavskaya[5], Patrick Hlavacek[6], Aster Meche[6], Jinma Ren[7], Alex Schepart[6], Didem Aydin[6], Guido Nador[8] & Marco daCosta DiBonaventura*,[6]

Future Oncol. 2024 Feb 28. doi: 10.2217/fon-2023-0995. Epub ahead of print. PMID: 38415370

Elranatamab efficacy in the single-arm, registrational MagnetisMM-3 trial (NCT04649359) was compared with that of physician's choice of treatment (PCT) for triple-class refractory multiple myeloma. MagnestisMM-3 eligibility criteria were applied to two USA-based oncology electronic health record databases, COTA and Flatiron Health (FH), to identify cohorts for this study (NCT05932290). Applied statistical techniques accounted for cohort imbalances. MagnetisMM-3 (BCMA-naive; n = 123) outcomes were compared with those from COTA (n = 239) and FH (n = 152).

Original, unadjusted population



Statistical analyses

To summarize our statistical analysis approach, we first controlled for baseline confounding variables by estimating propensity scores (PSs) using age, sex, race, ISS disease stage, ECOG performance status, time from initial MM diagnosis to index date, bone lesions, EMD (COTA database only), high-risk cytogenetics, CCI score, number of LOTs used prior to index, penta-drug refractory status, SCT and levels of aspartate aminotransferase, alanine aminotransferase, hemoglobin, creatinine clearance, calcium, bilirubin and serum albumin. PSs were then used to calculate inverse probability of treatment (IPT) weights, which balanced the distributions of these confounding variables across treatment groups.

The propensity score gives the probability of an individual being exposed (i.e. assigned to the intervention or risk factor) given their baseline characteristics.
The aim of the propensity score in observational research is to control for measured confounders by achieving balance in characteristics between exposed and unexposed groups.

Inverse probability of treatment weighting (IPTW) can be used to adjust for confounding in observational studies.
IPTW uses the propensity score to balance baseline patient characteristics in the exposed and unexposed groups by weighting each individual in the analysis by the inverse probability of receiving his/her actual exposure.

Future Oncol. 2024 Feb 28. doi: 10.2217/fon-2023-0995. Epub ahead of print. PMID: 38415370

Statistical analyses

To summarize our statistical analysis approach, we first controlled for baseline confounding variables by estimating propensity scores (PSs) using age, sex, race, ISS disease stage, ECOG performance status, time from initial MM diagnosis to index date, bone lesions, EMD (COTA database only), high-risk cytogenetics, CCI score, number of LOTs used prior to index, penta-drug refractory status, SCT and levels of aspartate aminotransferase, alanine aminotransferase, hemoglobin, creatinine clearance, calcium, bilirubin and serum albumin. PSs were then used to calculate inverse probability of treatment (IPT) weights, which balanced the distributions of these confounding variables across treatment groups.

The propensity score gives the probability of an individual being exposed (i.e. assigned to the intervention or risk factor) given their baseline characteristics.

The aim of the propensity score in observational research is to control for measured confounders by achieving balance in characteristics between exposed and unexposed groups.

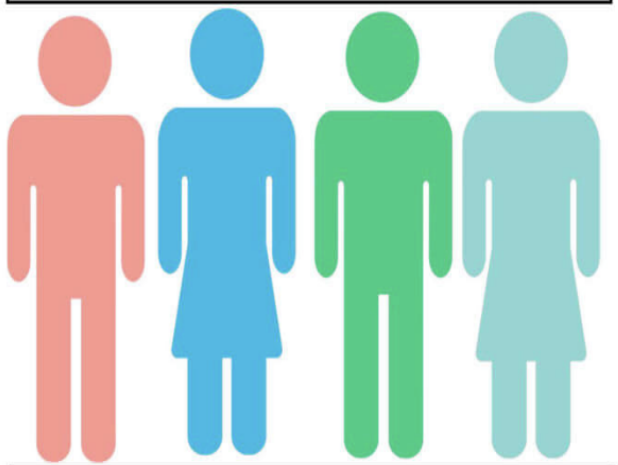
Inverse probability of treatment weighting (IPTW) can be used to adjust for confounding in observational studies.

IPTW uses the propensity score to balance baseline patient characteristics in the exposed and unexposed groups by weighting each individual in the analysis by the inverse probability of receiving his/her actual exposure.

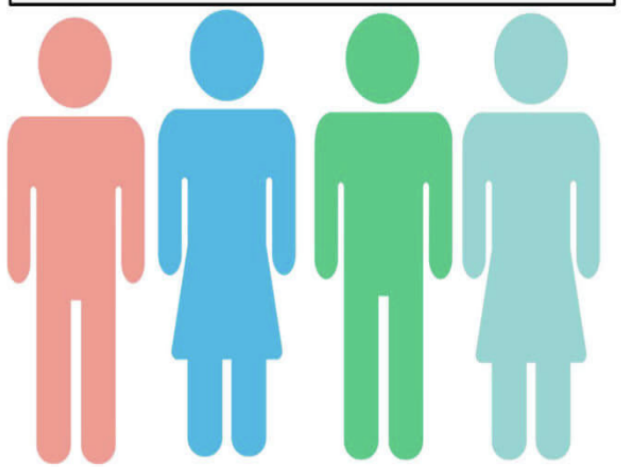
Future Oncol. 2024 Feb 28. doi: 10.2217/fon-2023-0995. Epub ahead of print. PMID: 38415370

Adjusted population after PS matching

Retained clinical trial patients

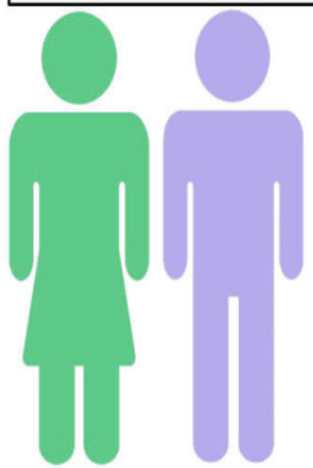


Retained real-world patients



Dropped clinical trial patient

Dropped real-word patients



Patients in the clinical trial get "paired up" with a similar real-world patient. Patients that don't find a match get dropped.

Statistical analyses

To summarize our statistical analysis approach, we first controlled for baseline confounding variables by estimating propensity scores (PSs) using age, sex, race, ISS disease stage, ECOG performance status, time from initial MM diagnosis to index date, bone lesions, EMD (COTA database only), high-risk cytogenetics, CCI score, number of LOTs used prior to index, penta-drug refractory status, SCT and levels of aspartate aminotransferase, alanine aminotransferase, hemoglobin, creatinine clearance, calcium, bilirubin and serum albumin. PSs were then used to calculate inverse probability of treatment (IPT) weights, which balanced the distributions of these confounding variables across treatment groups.

The propensity score gives the probability of an individual being exposed (i.e. assigned to the intervention or risk factor) given their baseline characteristics.

The aim of the propensity score in observational research is to control for measured confounders by achieving balance in characteristics between exposed and unexposed groups.

Inverse probability of treatment weighting (IPTW) can be used to adjust for confounding in observational studies.

IPTW uses the propensity score to balance baseline patient characteristics in the exposed and unexposed groups by weighting each individual in the analysis by the inverse probability of receiving his/her actual exposure.

Future Oncol. 2024 Feb 28. doi: 10.2217/fon-2023-0995. Epub ahead of print. PMID: 38415370

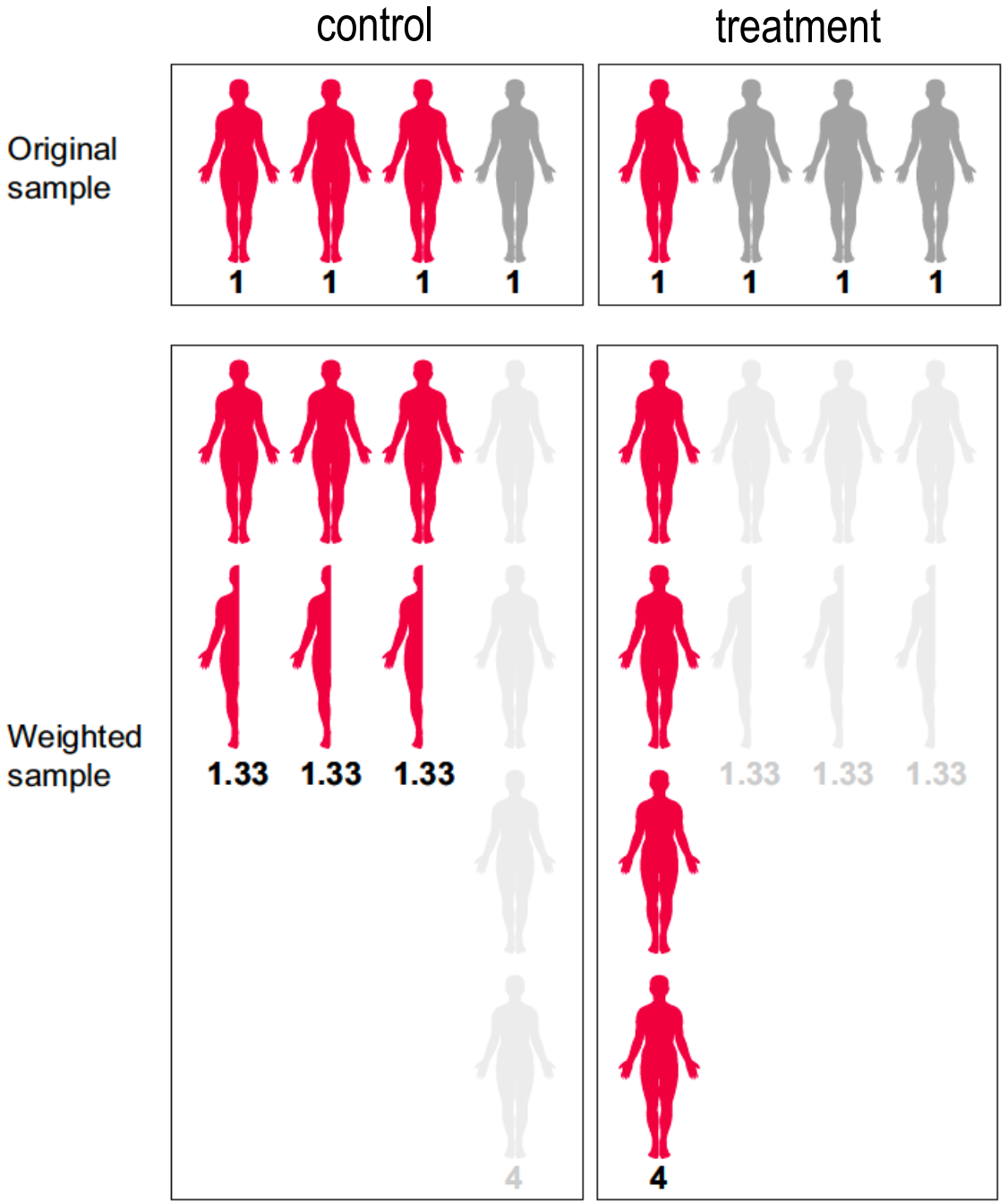
- probability of High-Risk (red figures) in treatment arm: 25%
- inverse probability weight of High-Risk in treatment arm: 1/0.25 = 4
- inverse probability weight of High-Risk in control arm: 1/(1-0.25) = 1.33
- probability of Low-Risk (grey figures) in treatment arm: 75%
- inverse probability weight of Low-Risk in treatment arm: 1/0.75 = 1.33
- inverse probability weight of Low-Risk in control arm: 1/(1-0.75) = 4

After applying the inverse probability weights to create a weighted pseudopopulation, the High-Risk characteristic is equally distributed across treatment groups (50% in each group).

Clinical Kidney Journal, 2022, vol. 15, no. 1, 14–20

- probability of High-Risk (red figures) in treatment arm: 25%
- inverse probability weight of High-Risk in treatment arm: 1/0.25 = 4
- inverse probability weight of High-Risk in control arm: 1/(1-0.25) = 1.33
- probability of Low-Risk (grey figures) in treatment arm: 75%
- inverse probability weight of Low-Risk in treatment arm: 1/0.75 = 1.33
- inverse probability weight of Low-Risk in control arm: 1/(1-0.75) = 4

After applying the inverse probability weights to create a weighted pseudopopulation, the High-Risk characteristic is equally distributed across treatment groups (50% in each group).



Clinical Kidney Journal, 2022, vol. 15, no. 1, 14–20

- probability of High-Risk (red figures) in treatment arm: 25%
- inverse probability weight of High-Risk in treatment arm: 1/0.25 = 4
- inverse probability weight of High-Risk in control arm: 1/(1-0.25) = 1.33
- probability of Low-Risk (grey figures) in treatment arm: 75%
- inverse probability weight of Low-Risk in treatment arm: 1/0.75 = 1.33
- inverse probability weight of Low-Risk in control arm: 1/(1-0.75) = 4

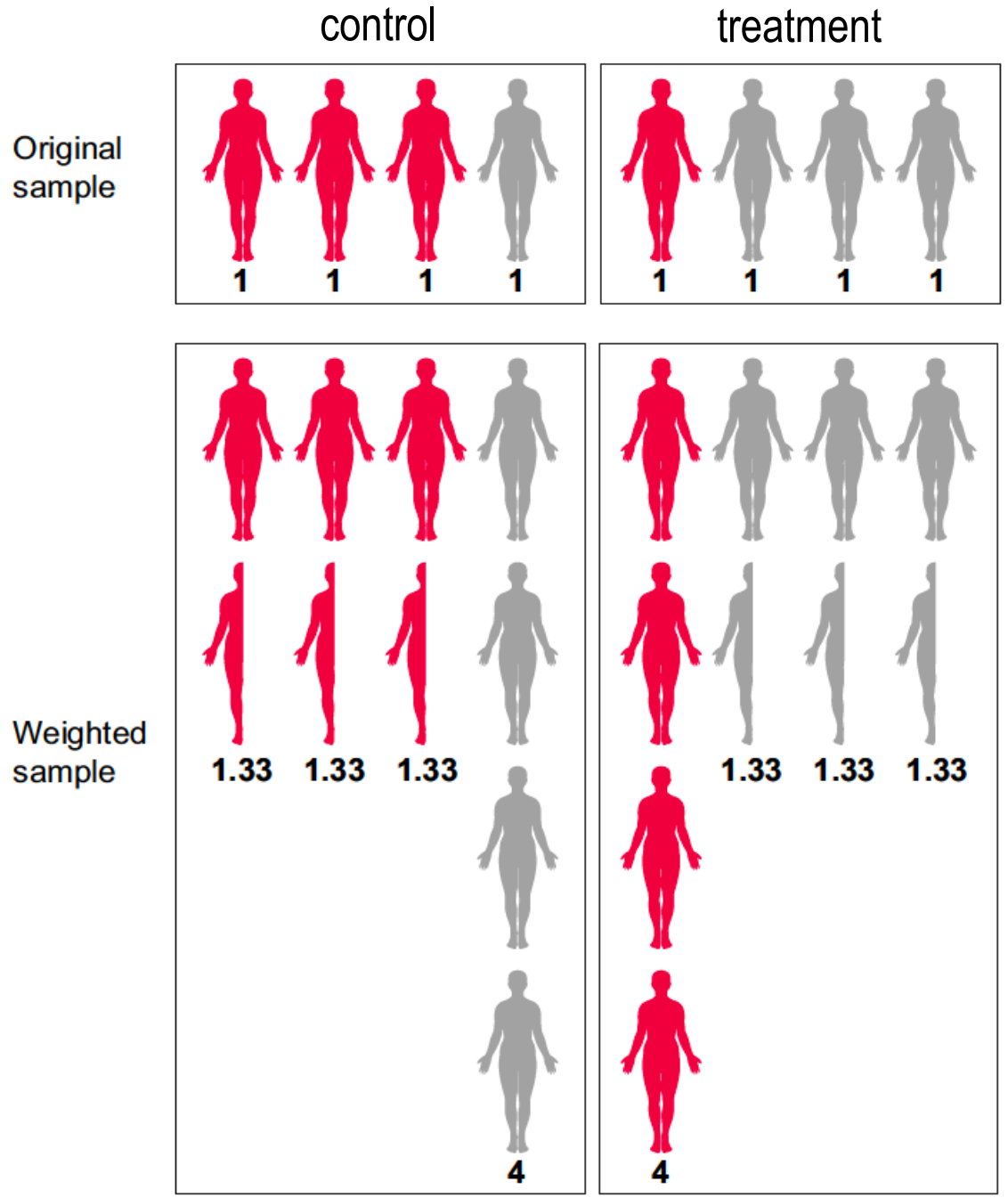
After applying the inverse probability weights to create a weighted pseudopopulation, the High-Risk characteristic is equally distributed across treatment groups (50% in each group).

Clinical Kidney Journal, 2022, vol. 15, no. 1, 14–20

- probability of High-Risk (red figures) in treatment arm: 25%
- inverse probability weight of High-Risk in treatment arm: 1/0.25 = 4
- inverse probability weight of High-Risk in control arm: 1/(1-0.25) = 1.33
- probability of Low-Risk (grey figures) in treatment arm: 75%
- inverse probability weight of Low-Risk in treatment arm: 1/0.75 = 1.33
- inverse probability weight of Low-Risk in control arm: 1/(1-0.75) = 4

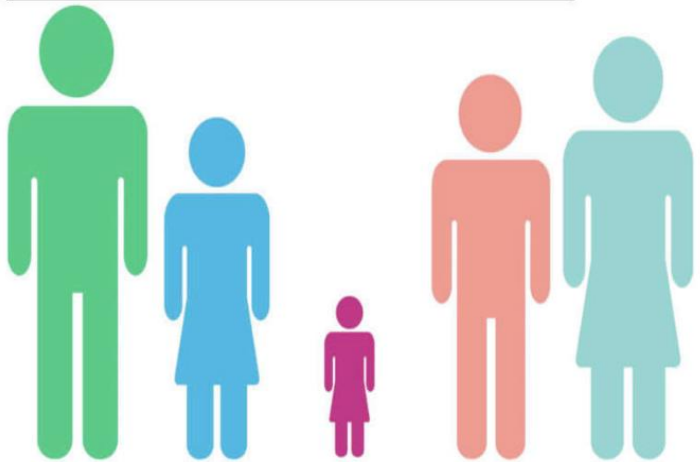
After applying the inverse probability weights to create a weighted pseudopopulation, the High-Risk characteristic is equally distributed across treatment groups (50% in each group).

Clinical Kidney Journal, 2022, vol. 15, no. 1, 14–20

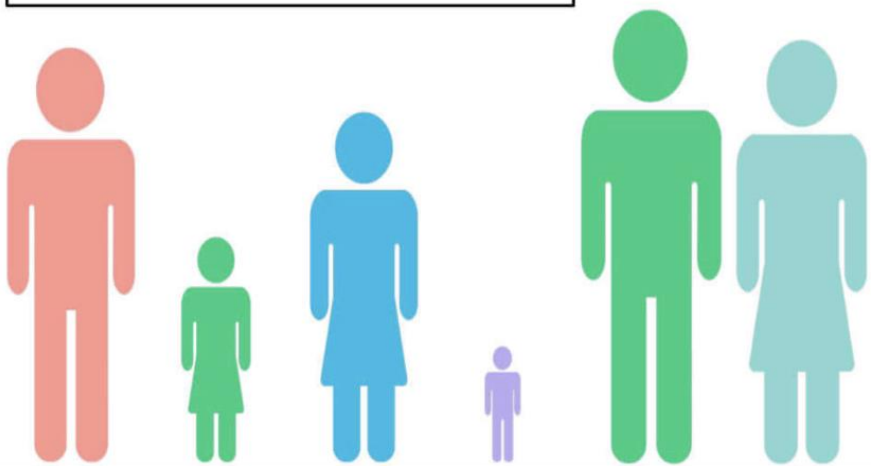


Adjusted population after PS weighting

Clinical trial patients



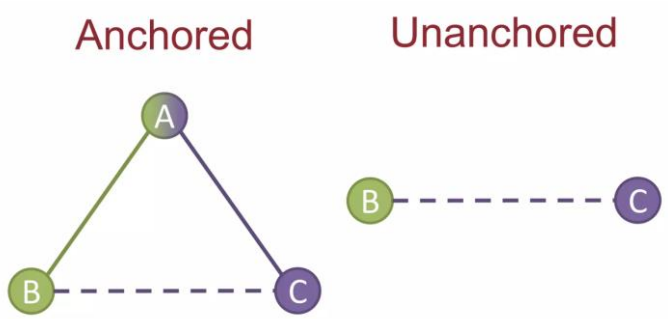
Real-world patients



More similar patients in each group get up-weighted, while the less similar patients get down-weighted. All patients are retained.

Population-adjusted Indirect Comparisons

- MAICs and STCs use patient-level data from a trial of a given treatment (referred to as the index trial) to derive a comparison of outcomes with competing treatments, based on published information from similarly designed studies, after adjusting for differences in the characteristics of the populations.
- *in other words*: individual patient data (IPD) in one or more trials are used to adjust for between-trial differences in the distribution of variables that influence outcome
- "anchored" indirect comparison (common comparator arm in each trial) Vs "unanchored" indirect comparison (disconnected treatment network or single-arm studies)
  - *an unanchored MAIC or STC assumes that all effect modifiers and prognostic factors are accounted for*
- unanchored methods for population adjustment are problematic and should not be used when anchored methods can be applied



https://www.nicedsu.org.uk/wp-content/uploads/2017/05/Population-adjustment-TSD-FINAL.pdf.

Network Meta-Analysis (NMA)

(M. Cinquini)

Network Meta-Analysis

(Multiple Treatments Meta-Analysis, Mixed Treatment Comparisons)

- Combine direct + indirect estimates of multiple treatment effects
- Internally consistent set of estimates that respects randomization
- Estimate effect of each intervention relative to every other **whether or not there is direct comparison in studies**
- Calculate probability that each treatment is most effective
- Compared to conventional pair-wise meta-analysis:
    - Greater precision in summary estimates
    - Ranking of treatments according to effectiveness

Indirect Comparisons of Multiple Treatments – Network Meta-Analysis

Trial			
1	A	B	
2	A	B	
3		B	C
4		B	C
5	A		C
6	A		C
7	A	B	C

- Want to compare A vs. B
  Direct evidence from trials 1, 2 and 7
  Indirect evidence from trials 3, 4, 5, 6 and 7

- Combining all “A” arms and comparing with all “B” arms destroys randomization

- Use indirect evidence of A vs. C and B vs. C comparisons as additional evidence to preserve randomization and within-study comparison

Indirect Comparisons

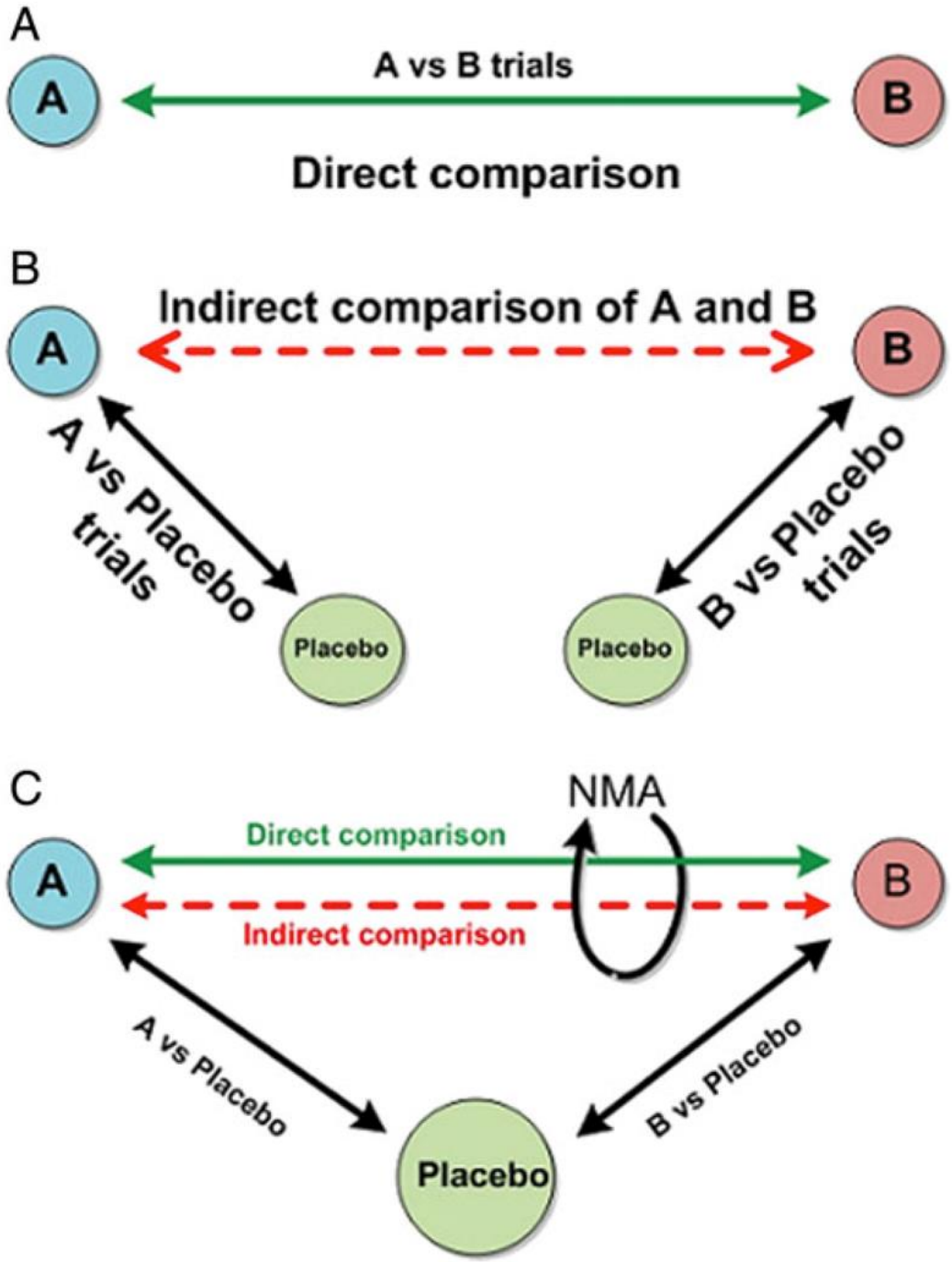
Basic assumptions underlying indirect comparisons include:

- ✓ **homogeneity** assumption for standard meta-analysis,
- ✓ **similarity** assumption for adjusted indirect comparison and
- ✓ **consistency** assumption for the combination of direct and indirect evidence. It is essential to fully understand and appreciate these basic assumptions in order to use adjusted indirect and mixed treatment comparisons appropriately.

CONSISTENCY ASSUMPTION

- When both direct and indirect evidence is available, an assumption of evidence consistency is required to quantitatively combine the direct and indirect estimates.
- It is important to investigate possible causes of discrepancy between the direct and indirect evidence, such as the play of chance, invalid indirect comparison, bias in head-to-head comparative trials, and clinically meaningful heterogeneity
- When the direct comparison differs from the adjusted indirect comparison, we should usually give more credibility to evidence from head-to-head comparative trials. However, evidence from direct comparative trials may not always be valid.

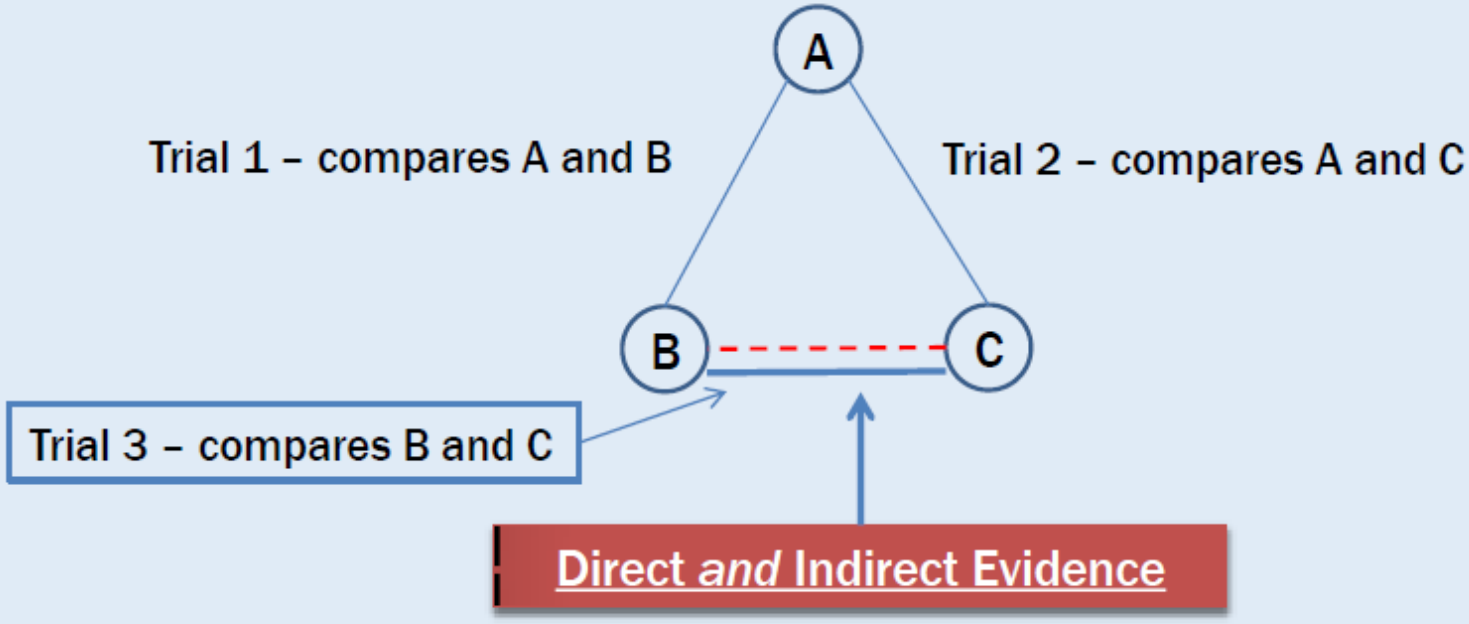
Song, What is …? 2009; Song et al, J Clin Epidemiol 2008



Consistency Assumption

there must be no relevant discrepancy between direct and indirect evidence

THERE ARE 2 TYPES OF TRIAL EVIDENCE



Consistency ⇒ Direct and indirect evidence agree

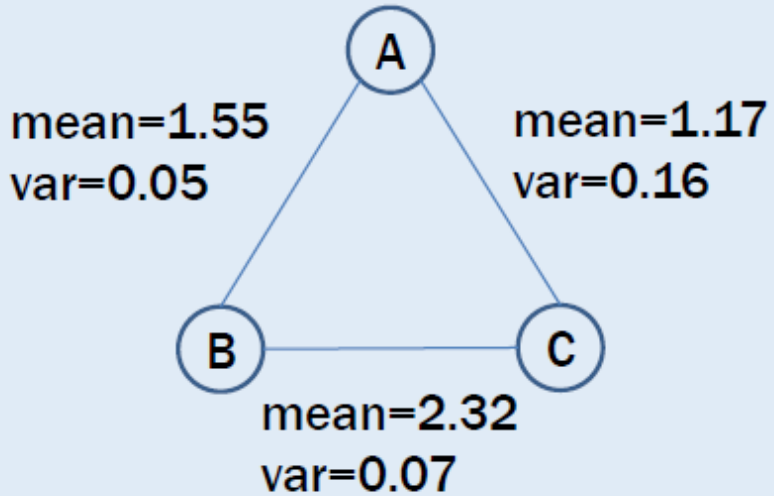
Inconsistency ⇒ Direct and indirect evidence disagree

Differing effect modifiers among the trials can cause inconsistency

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 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**

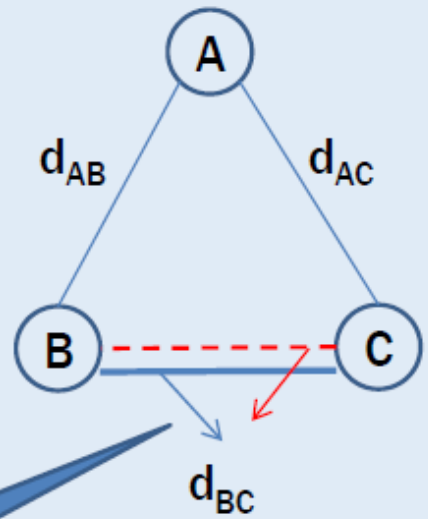
- Measure of inconsistency (Z):
  = 2.32 (Direct estimate) – (-0.38) (Indirect estimate) = **2.70**
  variance = 0.07 + 0.21 = **0.28**

- If $Z/\sqrt{Var(Z)}$ is rejected ($N(0,1)$) then the loop is inconsistent

In this case P<.000001, indicating inconsistency

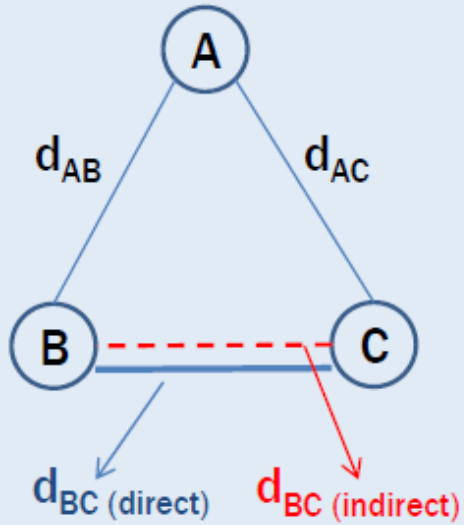
#2 NODE-SPLITTING

Full NMA estimates 3 parameters



Direct and indirect evidence inform this comparison

Node-splitting estimates separate parameters for direct and indirect evidence



Inconsistency is present if

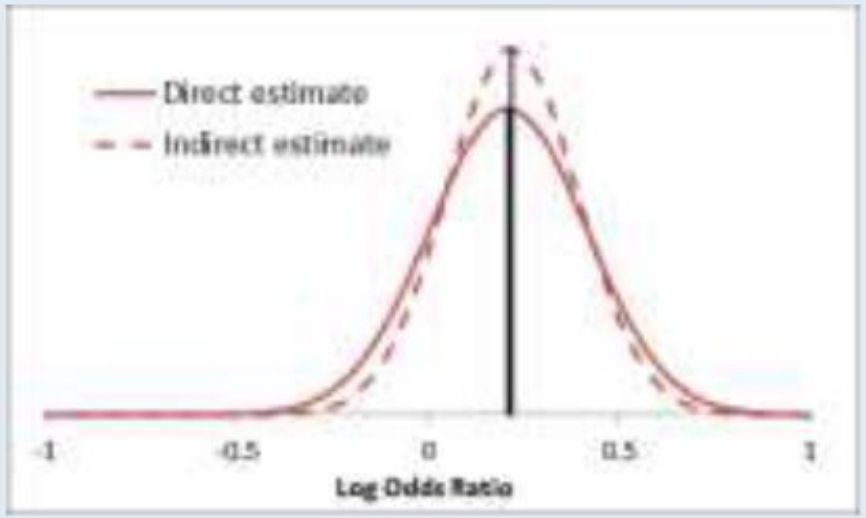
$$d_{BC\ (direct)} \neq d_{BC\ (indirect)}$$

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

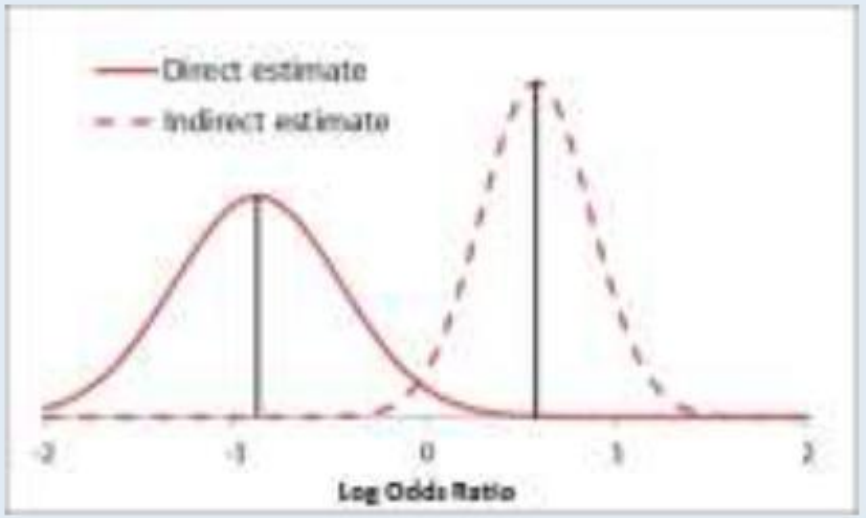
Example of posterior distributions with direct and indirect evidence

Consistent Evidence



Posterior densities overlap indicating absence of inconsistency

Inconsistent Evidence

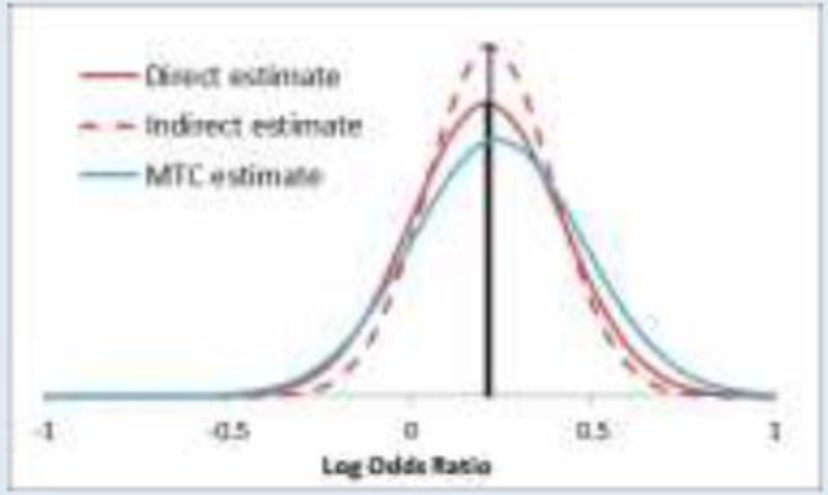


Posterior densities hardly overlap indicating presence of inconsistency

#2 NODE-SPLITTING

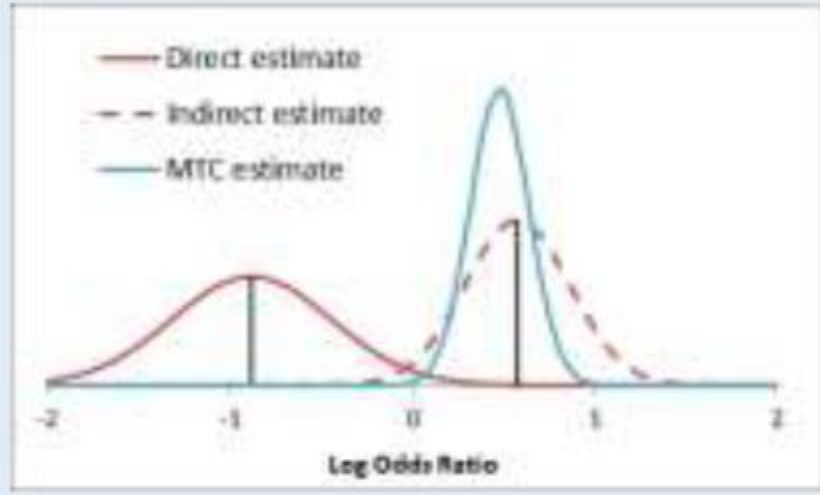
What do we do with this information?

Consistent Evidence



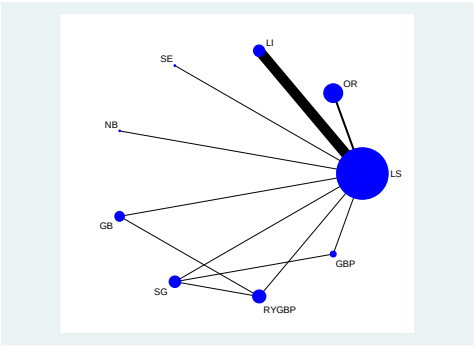
MTC estimate is similar to direct and indirect estimates

Inconsistent Evidence



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

Step 1: generating network geometry



Step 2: testing for inconsistency

```
Multivariate meta-analysis
Variance-covariance matrix = proportional .5*I(4)+.5*J(4,4,1)
Method = reml                                 Number of dimensions    =      4
Restricted log likelihood = -30.939719        Number of observations  =     25
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_y_B						
des_ABC	.2528377	.5704516	0.44	0.658	-.8652269	1.370902
des_ABD	-.7433714	.5269164	-1.41	0.158	-1.776108	.2893657
des_ABE	-.1959024	.5311986	-0.37	0.712	-1.237033	.8452278
_cons	-.9727775	.2201655	-4.42	0.000	-1.404294	-.5412611
_y_C						
des_AC	.217719	.6845858	0.32	0.750	-1.124045	1.559483
_cons	-1.58294	.6293945	-2.52	0.012	-2.816531	-.3493498
_y_D						
des_AD	.5489224	.5775957	0.95	0.342	-.5831443	1.680989
des_BDE	1.020097	.9029483	1.13	0.259	-.7496496	2.789843
des_CD	.633251	.9312281	0.68	0.496	-1.191923	2.458425
_cons	-1.72662	.4786004	-3.61	0.000	-2.66466	-.7885806
_y_E						
des_BDE	.4401131	1.862385	0.24	0.813	-3.210095	4.090321
_cons	-3.402272	1.051331	-3.24	0.001	-5.462844	-1.3417

```
Estimated between-studies SDs and correlation matrix:
             SD        _y_B        _y_C        _y_D        _y_E
_y_B   1.767e-09           1           .           .           .
_y_C   1.767e-09          .5           1           .           .
_y_D   1.767e-09          .5          .5           1           .
_y_E   1.767e-09          .5          .5          .5           1

Testing for inconsistency:
 ( 1)  [_y_B]des_ABC = 0
 ( 2)  [_y_B]des_ABD = 0
 ( 3)  [_y_B]des_ABE = 0
 ( 4)  [_y_C]des_AC = 0
 ( 5)  [_y_D]des_AD = 0
 ( 6)  [_y_D]des_BDE = 0
 ( 7)  [_y_E]des_BDE = 0
 ( 8)  [_y_D]des_CD = 0

           chi2(  8) =    4.00
         Prob > chi2 =    0.8567
```

Step 3: creating plots and league table of effect size by treatment

Table 1. Inconsistency test between direct and indirect treatment comparisons in mixed treatment comparison

Side	Direct		Indirect		Difference		p>z
	Coefficient	SE	Coefficient	SE	Coefficient	SE	
A B	-1.083	0.174	-0.877	0.620	-0.206	0.636	0.746
A C	-1.388	0.247	-1.869	0.493	0.481	0.542	0.375
A D	-1.378	0.265	-0.738	0.413	-0.640	0.479	0.182
A E	-3.425	0.940	-3.221	1.005	-0.204	0.937	0.828
B C	-0.894	0.655	-0.312	0.297	-0.581	0.715	0.416
B D	0.099	0.462	-0.241	0.329	0.340	0.567	0.548
B E	-2.152	0.881	-2.615	1.087	0.463	0.896	0.605
C D	0.490	0.492	0.177	0.350	0.313	0.604	0.605
D E	-2.550	1.254	-1.956	0.958	-0.595	1.314	0.651

SE, standard error; A, placebo; B, IV_single; C, IV_double; D, topical; E, combination.

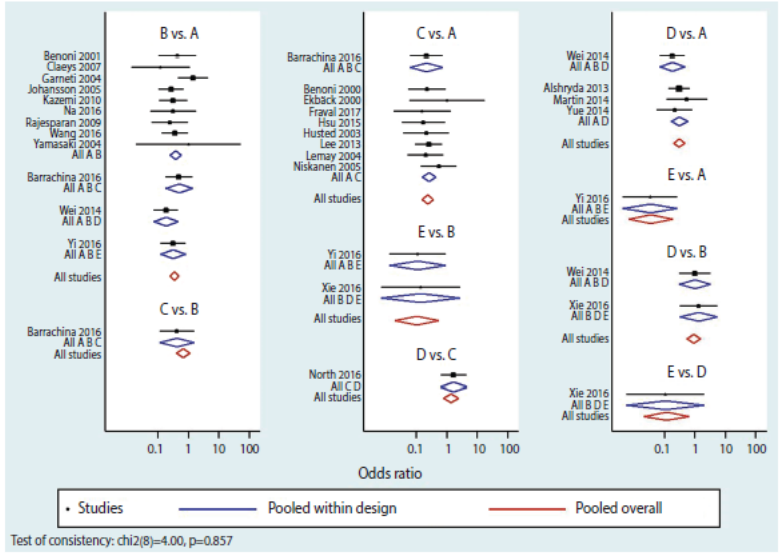


Figure 4. Network forest plot. A, placebo; B, IV_single; C, IV_double; D, topical; E, combination.

Step 4: determining relative rankings of treatment

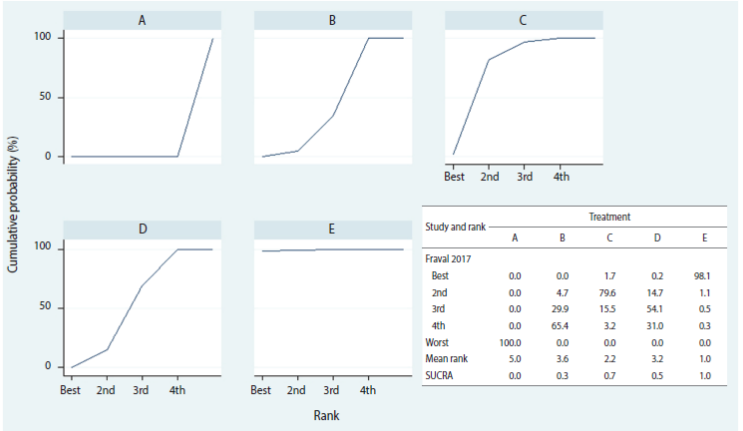


Figure 6. Results of network rank test. A, placebo; B, IV_single; C, IV_double; D, topical; E, combination; SCURA, surface under the cumulative ranking.

Presenting the data

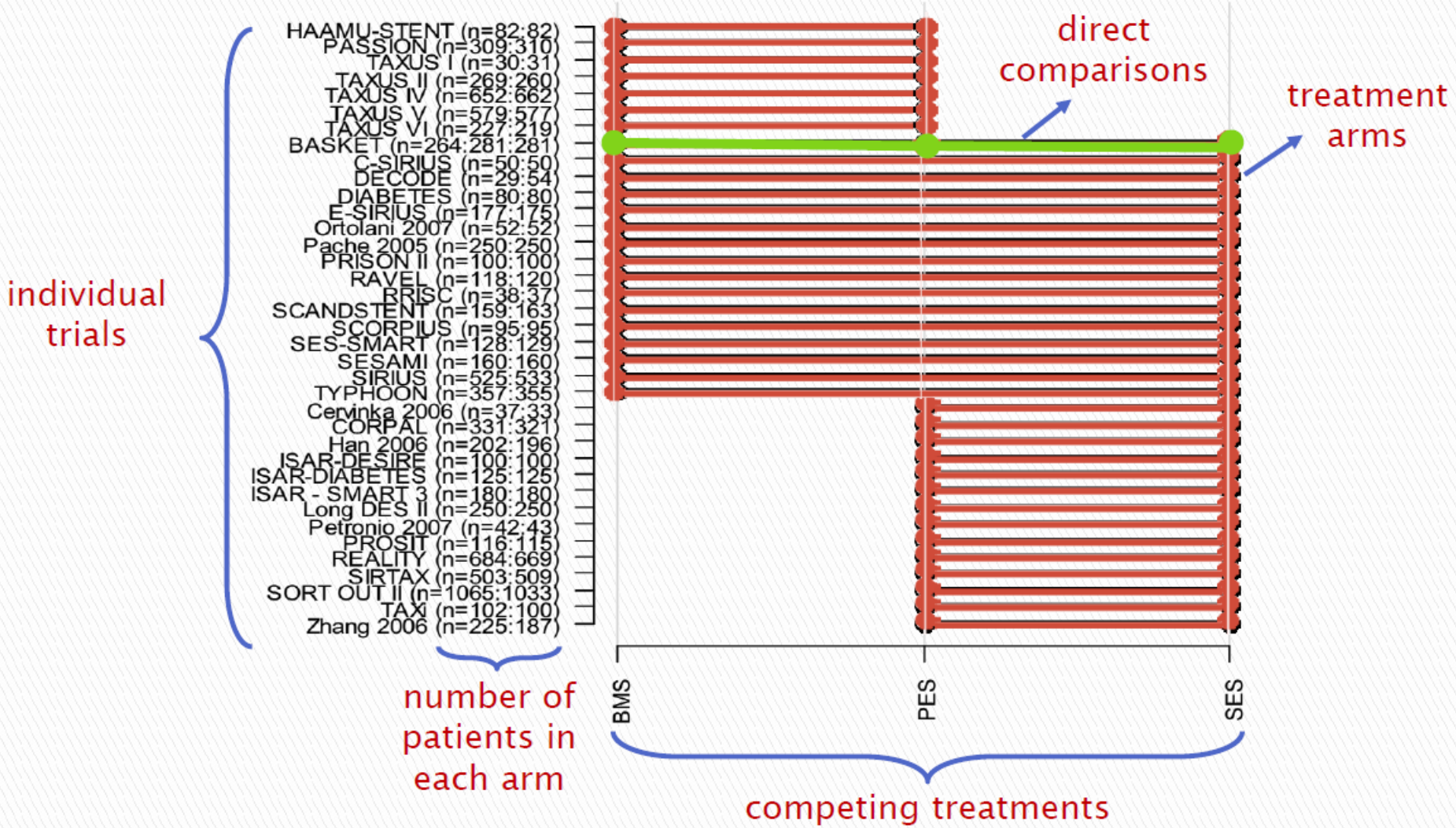
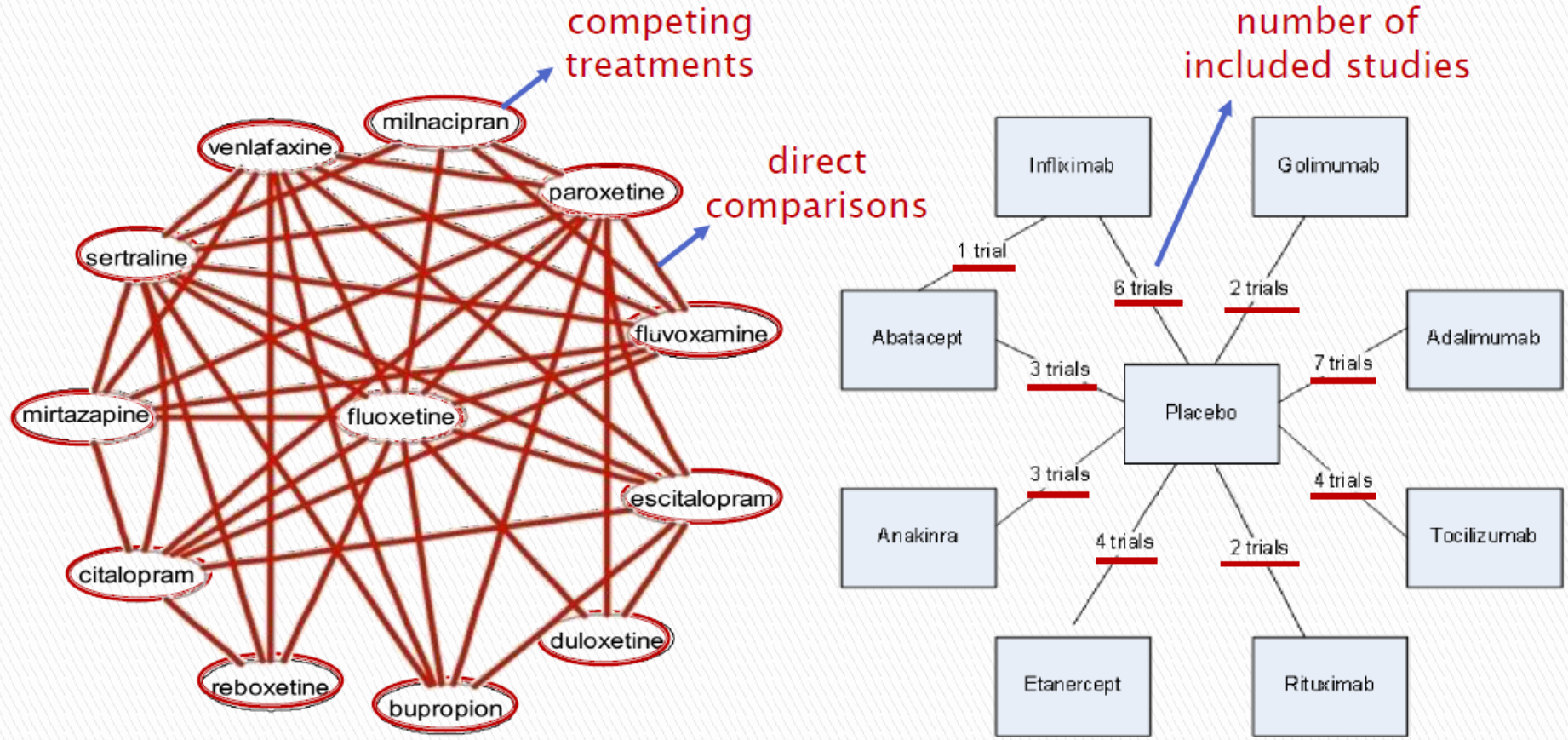


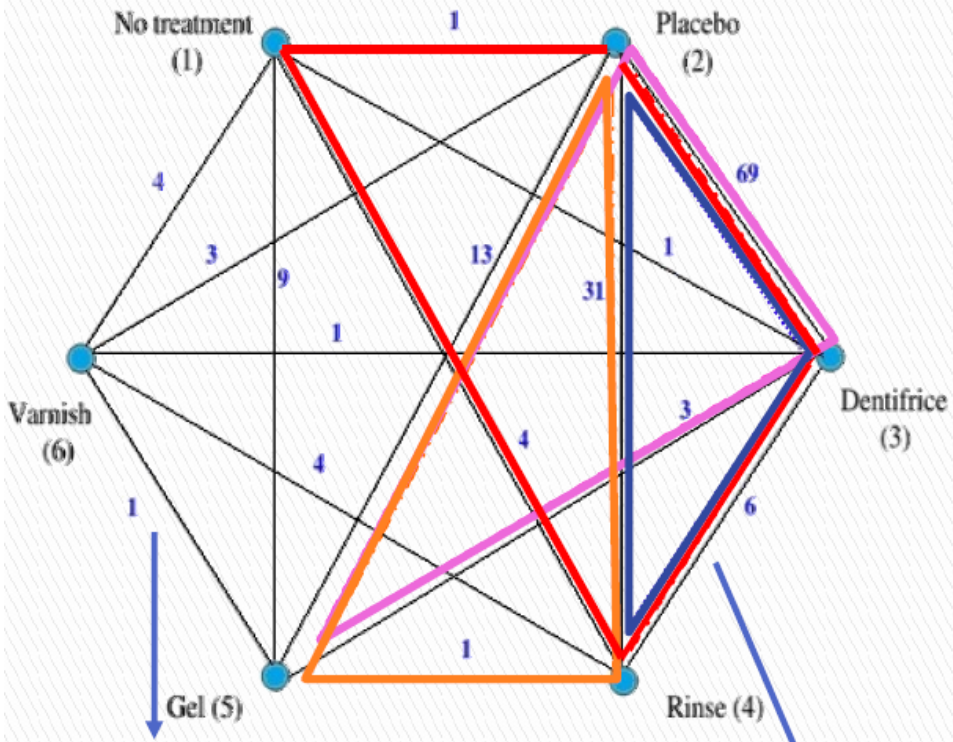
Diagram showing the comparisons involved in the individual studies of the network

[Example in Hoaglin et al. 2011]



Network graph showing the available direct comparisons in the network

[Examples in Hoaglin et al. 2011 & Jonas et al. 2013]



Type (g)	No. of studies	No. trt (1)	Placebo (2)	Dentifrice (3)	Rinse (4)	Gel (5)	Varnish (6)
1	9	X				X	
2	3	X			X		
3	4	X					X
4	61		X	X			
5	9		X			X	
6	25		X		X		
7	3		X				X
8	1			X	X		
9	1			X		X	
10	1					X	X
11	4				X		X
12	4		X	X	X		
13	3		X	X		X	
14	1		X		X	X	
15	1	X	X	X	X		

direct comparisons from two-arm trials

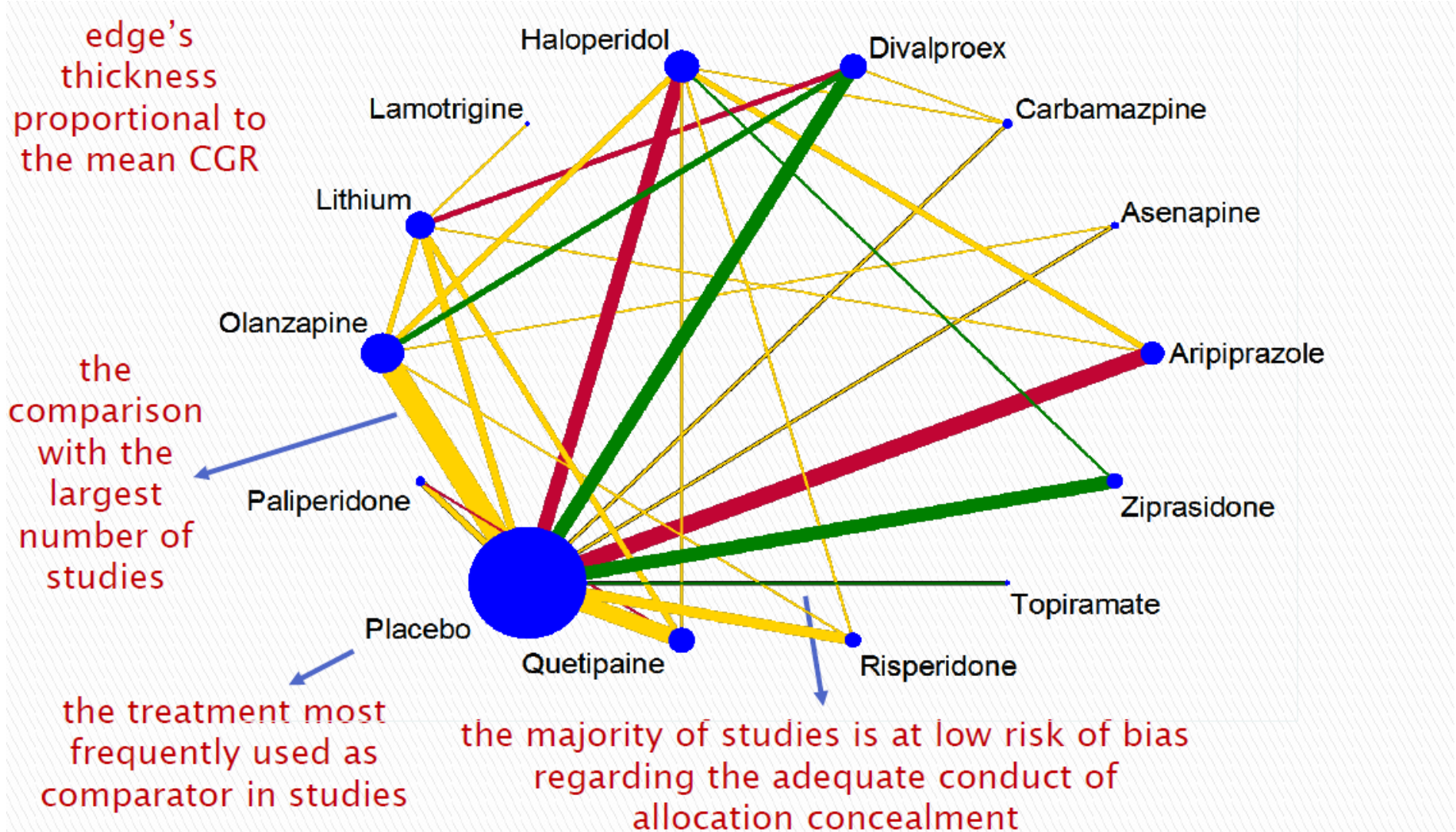
direct comparisons from multi-arm trials

15 different study designs

treatment arms in each design

Network graph showing the presence of multi-arm trials & table showing the network structure; the available study designs in the network

[Examples in Lu et al. 2011]

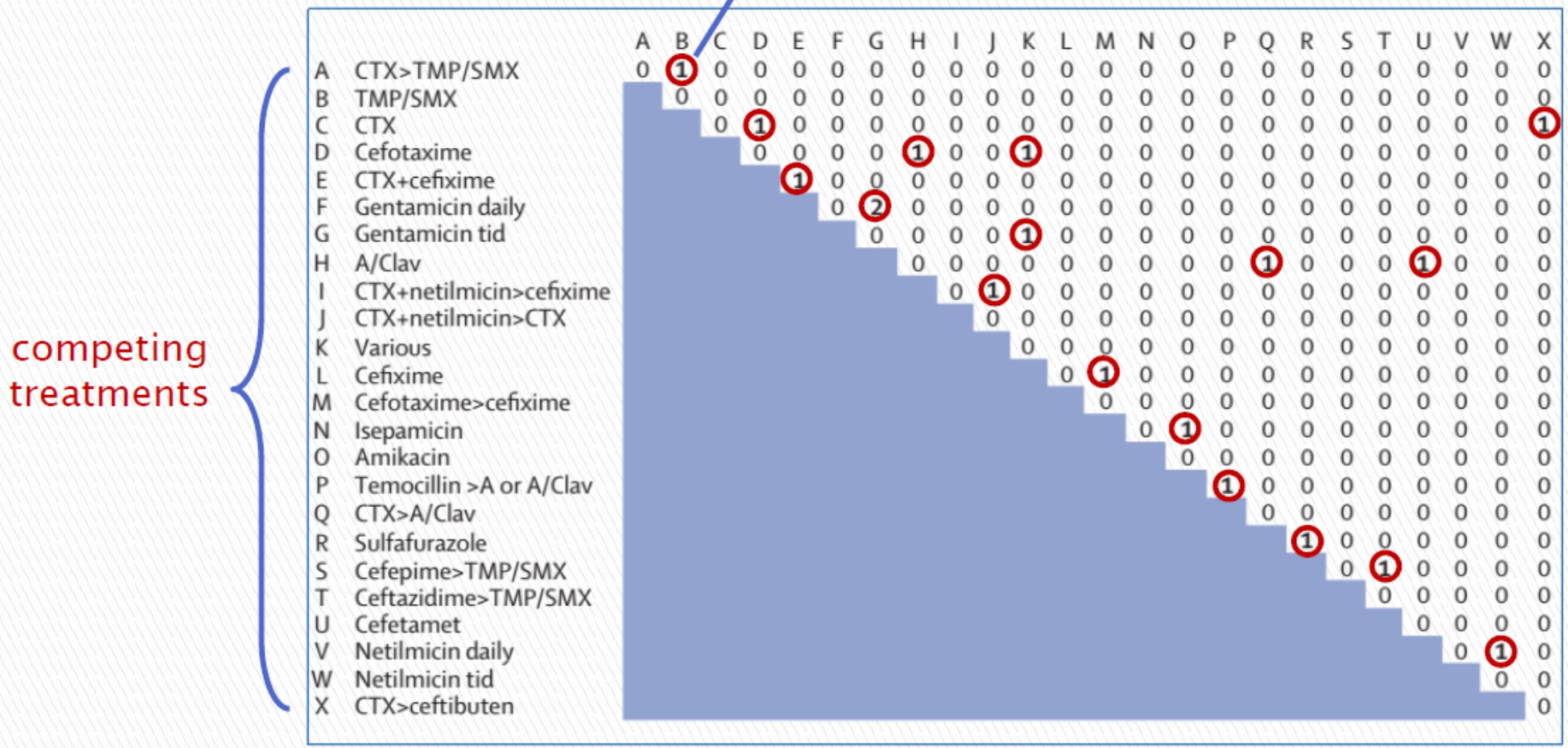


Network graph with weighted and/or colored nodes and edges

[Examples in Chaimani et al. 2013]

number of trials comparing the treatments in the respective row and column

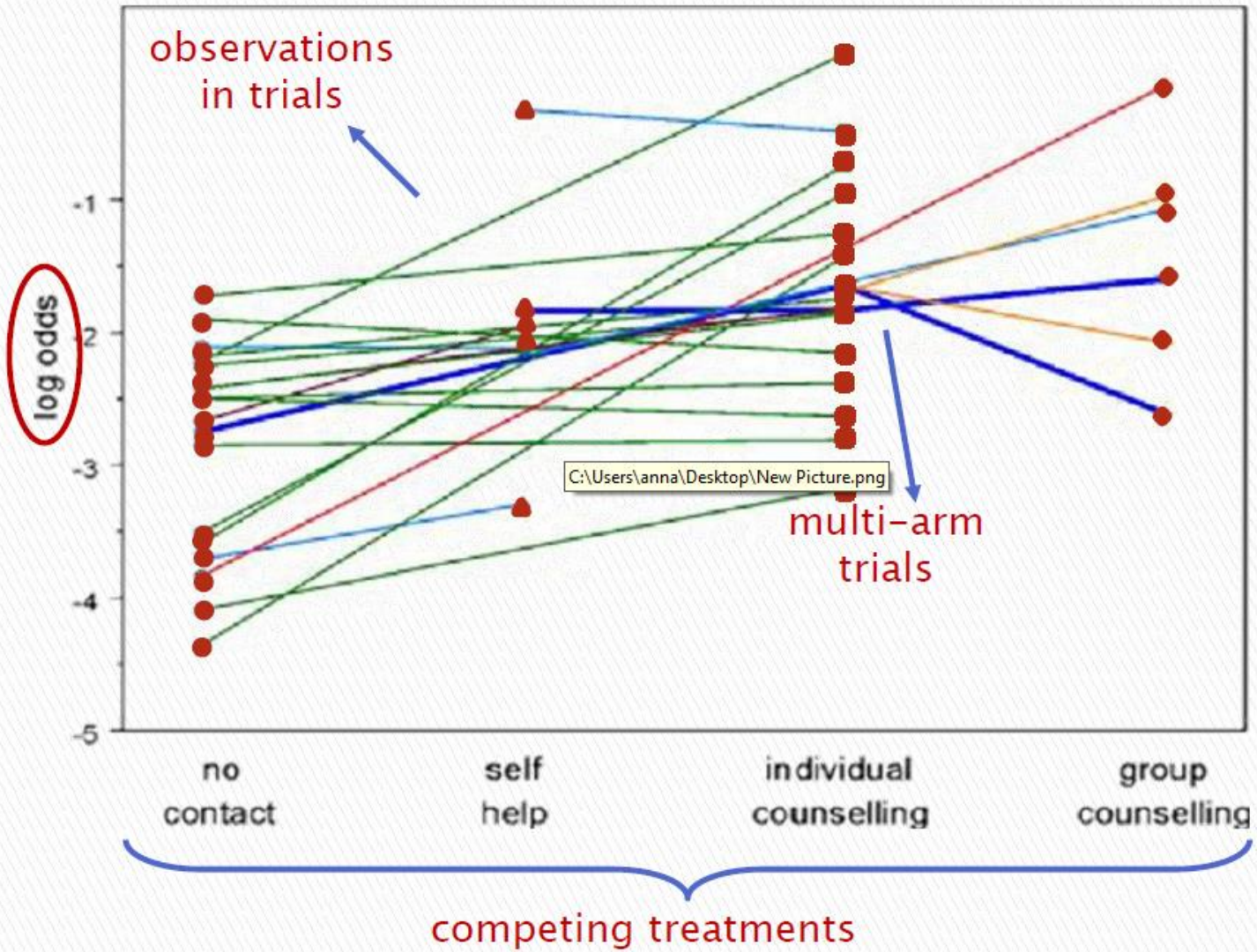
competing treatments



Matrix showing the available direct comparisons in the network

[Example in Ioannidis 2006]

the slope of the lines shows which treatments are favored in studies

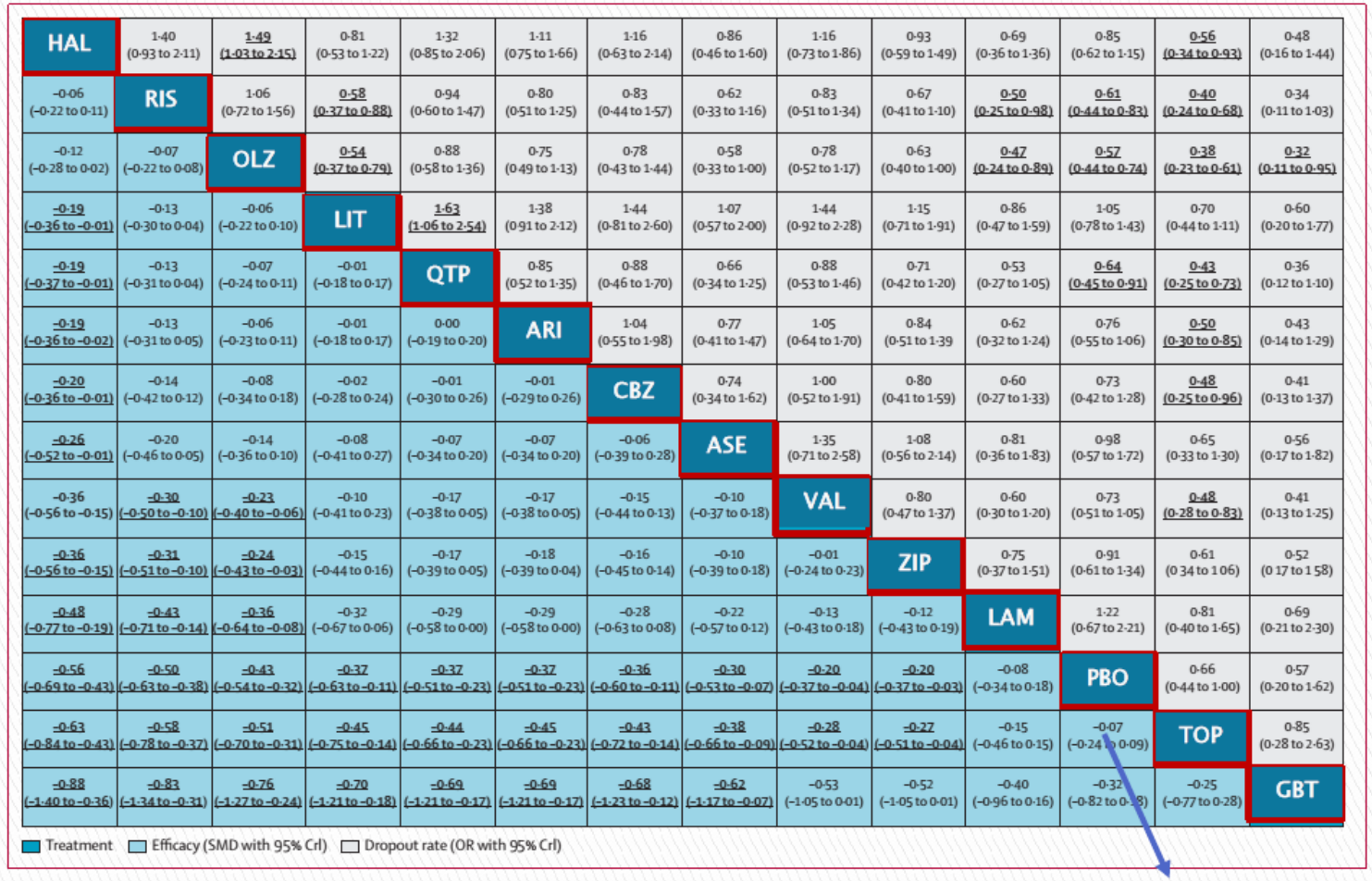


Graph showing the data provided by the individual studies of the network

[Example in Lu & Ades 2006]

Presenting the results measures of effect

relative treatment effects for efficacy
SMD<0 favor the treatment in column



HAL	1·40 (0·93 to 2·11)	**1·49 (1·03 to 2·15)**	0·81 (0·53 to 1·22)	1·32 (0·85 to 2·06)	1·11 (0·75 to 1·66)	1·16 (0·63 to 2·14)	0·86 (0·46 to 1·60)	1·16 (0·73 to 1·86)	0·93 (0·59 to 1·49)	0·69 (0·36 to 1·36)	0·85 (0·62 to 1·15)	**0·56 (0·34 to 0·93)**	0·48 (0·16 to 1·44)
−0·06 (−0·22 to 0·11)	**RIS**	1·06 (0·72 to 1·56)	**0·58 (0·37 to 0·88)**	0·94 (0·60 to 1·47)	0·80 (0·51 to 1·25)	0·83 (0·44 to 1·57)	0·62 (0·33 to 1·16)	0·83 (0·51 to 1·34)	0·67 (0·41 to 1·10)	**0·50 (0·25 to 0·98)**	**0·61 (0·44 to 0·83)**	**0·40 (0·24 to 0·68)**	0·34 (0·11 to 1·03)
−0·12 (−0·28 to 0·02)	−0·07 (−0·22 to 0·08)	**OLZ**	**0·54 (0·37 to 0·79)**	0·88 (0·58 to 1·36)	0·75 (0·49 to 1·13)	0·78 (0·43 to 1·44)	0·58 (0·33 to 1·00)	0·78 (0·52 to 1·17)	0·63 (0·40 to 1·00)	**0·47 (0·24 to 0·89)**	**0·57 (0·44 to 0·74)**	**0·38 (0·23 to 0·61)**	**0·32 (0·11 to 0·95)**
−0·19 (−0·36 to −0·01)	−0·13 (−0·30 to 0·04)	−0·06 (−0·22 to 0·10)	**LIT**	**1·63 (1·06 to 2·54)**	1·38 (0·91 to 2·12)	1·44 (0·81 to 2·60)	1·07 (0·57 to 2·00)	1·44 (0·92 to 2·28)	1·15 (0·71 to 1·91)	0·86 (0·47 to 1·59)	1·05 (0·78 to 1·43)	0·70 (0·44 to 1·11)	0·60 (0·20 to 1·77)
−0·19 (−0·37 to −0·01)	−0·13 (−0·31 to 0·04)	−0·07 (−0·24 to 0·11)	−0·01 (−0·18 to 0·17)	**QTP**	0·85 (0·52 to 1·35)	0·88 (0·46 to 1·70)	0·66 (0·34 to 1·25)	0·88 (0·53 to 1·46)	0·71 (0·42 to 1·20)	0·53 (0·27 to 1·05)	**0·64 (0·45 to 0·91)**	**0·43 (0·25 to 0·73)**	0·36 (0·12 to 1·10)
−0·19 (−0·36 to −0·02)	−0·13 (−0·31 to 0·05)	−0·06 (−0·23 to 0·11)	−0·01 (−0·18 to 0·17)	0·00 (−0·19 to 0·20)	**ARI**	1·04 (0·55 to 1·98)	0·77 (0·41 to 1·47)	1·05 (0·64 to 1·70)	0·84 (0·51 to 1·39	0·62 (0·32 to 1·24)	0·76 (0·55 to 1·06)	**0·50 (0·30 to 0·85)**	0·43 (0·14 to 1·29)
−0·20 (−0·36 to −0·01)	−0·14 (−0·42 to 0·12)	−0·08 (−0·34 to 0·18)	−0·02 (−0·28 to 0·24)	−0·01 (−0·30 to 0·26)	−0·01 (−0·29 to 0·26)	**CBZ**	0·74 (0·34 to 1·62)	1·00 (0·52 to 1·91)	0·80 (0·41 to 1·59)	0·60 (0·27 to 1·33)	0·73 (0·42 to 1·28)	**0·48 (0·25 to 0·96)**	0·41 (0·13 to 1·37)
−0·26 (−0·52 to −0·01)	−0·20 (−0·46 to 0·05)	−0·14 (−0·36 to 0·10)	−0·08 (−0·41 to 0·27)	−0·07 (−0·34 to 0·20)	−0·07 (−0·34 to 0·20)	−0·06 (−0·39 to 0·28)	**ASE**	1·35 (0·71 to 2·58)	1·08 (0·56 to 2·14)	0·81 (0·36 to 1·83)	0·98 (0·57 to 1·72)	0·65 (0·33 to 1·30)	0·56 (0·17 to 1·82)
−0·36 (−0·56 to −0·15)	**−0·30 (−0·50 to −0·10)**	**−0·23 (−0·40 to −0·06)**	−0·10 (−0·41 to 0·23)	−0·17 (−0·38 to 0·05)	−0·17 (−0·38 to 0·05)	−0·15 (−0·44 to 0·13)	−0·10 (−0·37 to 0·18)	**VAL**	0·80 (0·47 to 1·37)	0·60 (0·30 to 1·20)	0·73 (0·51 to 1·05)	**0·48 (0·28 to 0·83)**	0·41 (0·13 to 1·25)
−0·36 (−0·56 to −0·15)	**−0·31 (−0·51 to −0·10)**	**−0·24 (−0·43 to −0·03)**	−0·15 (−0·44 to 0·16)	−0·17 (−0·39 to 0·05)	−0·18 (−0·39 to 0·04)	−0·16 (−0·45 to 0·14)	−0·10 (−0·39 to 0·18)	−0·01 (−0·24 to 0·23)	**ZIP**	0·75 (0·37 to 1·51)	0·91 (0·61 to 1·34)	0·61 (0·34 to 1·06)	0·52 (0·17 to 1·58)
−0·48 (−0·77 to −0·19)	**−0·43 (−0·71 to −0·14)**	**−0·36 (−0·64 to −0·08)**	−0·32 (−0·67 to 0·06)	−0·29 (−0·58 to 0·00)	−0·29 (−0·58 to 0·00)	−0·28 (−0·63 to 0·08)	−0·22 (−0·57 to 0·12)	−0·13 (−0·43 to 0·18)	−0·12 (−0·43 to 0·19)	**LAM**	1·22 (0·67 to 2·21)	0·81 (0·40 to 1·65)	0·69 (0·21 to 2·30)
−0·56 (−0·69 to −0·43)	**−0·50 (−0·63 to −0·38)**	**−0·43 (−0·54 to −0·32)**	**−0·37 (−0·63 to −0·11)**	**−0·37 (−0·51 to −0·23)**	**−0·37 (−0·51 to −0·23)**	**−0·36 (−0·60 to −0·11)**	**−0·30 (−0·53 to −0·07)**	**−0·20 (−0·37 to −0·04)**	**−0·20 (−0·37 to −0·03)**	−0·08 (−0·34 to 0·18)	**PBO**	0·66 (0·44 to 1·00)	0·57 (0·20 to 1·62)
−0·63 (−0·84 to −0·43)	**−0·58 (−0·78 to −0·37)**	**−0·51 (−0·70 to −0·31)**	**−0·45 (−0·75 to −0·14)**	**−0·44 (−0·66 to −0·23)**	**−0·45 (−0·66 to −0·23)**	**−0·43 (−0·72 to −0·14)**	**−0·38 (−0·66 to −0·09)**	**−0·28 (−0·52 to −0·04)**	**−0·27 (−0·51 to −0·04)**	−0·15 (−0·46 to 0·15)	−0·07 (−0·24 to 0·09)	**TOP**	0·85 (0·28 to 2·63)
−0·88 (−1·40 to −0·36)	**−0·83 (−1·34 to −0·31)**	**−0·76 (−1·27 to −0·24)**	**−0·70 (−1·21 to −0·18)**	**−0·69 (−1·21 to −0·17)**	**−0·69 (−1·21 to −0·17)**	**−0·68 (−1·23 to −0·12)**	**−0·62 (−1·17 to −0·07)**	−0·53 (−1·05 to 0·01)	−0·52 (−1·05 to 0·01)	−0·40 (−0·96 to 0·16)	−0·32 (−0·82 to 0·18)	−0·25 (−0·77 to 0·28)	**GBT**

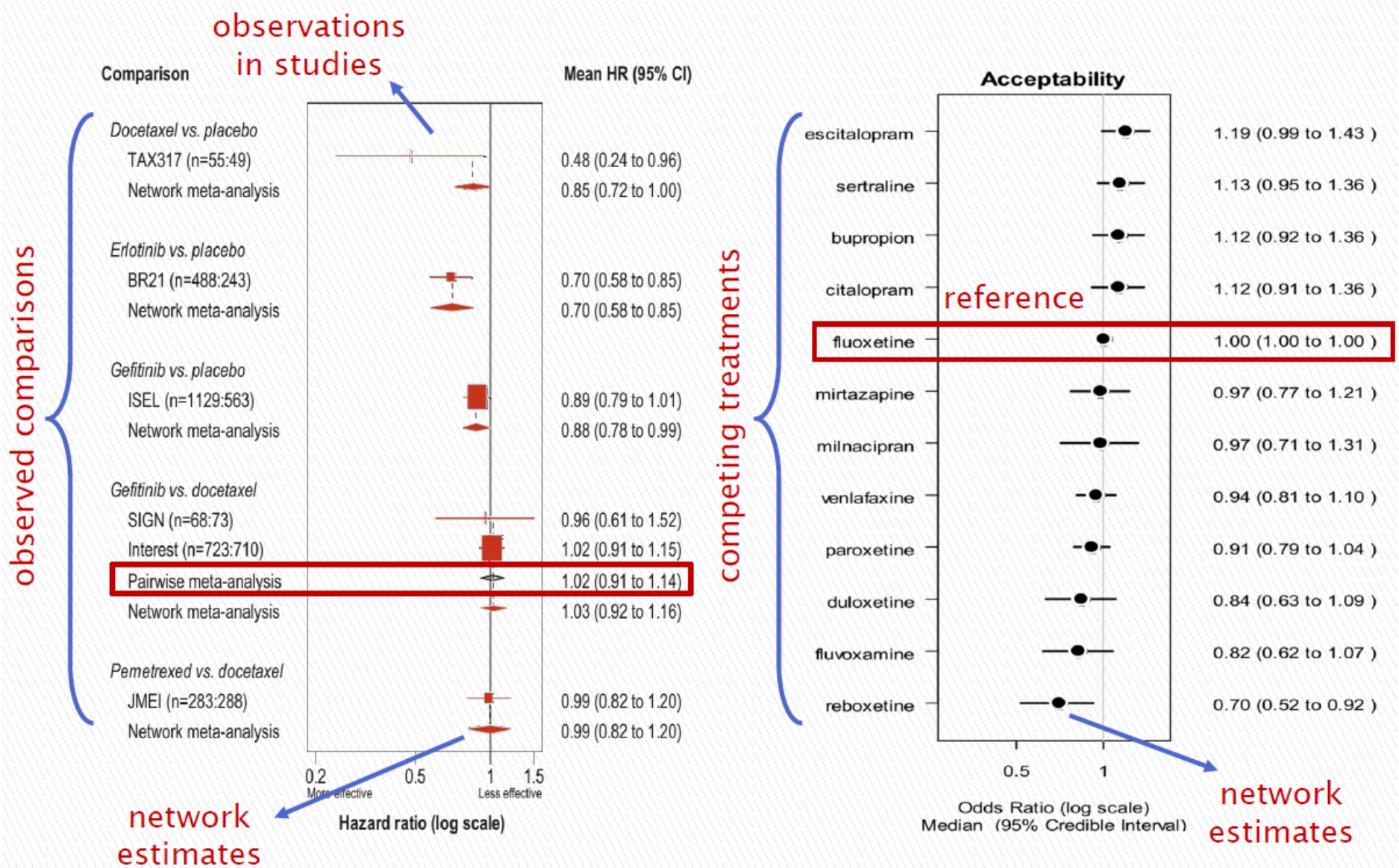
Treatment Efficacy (SMD with 95% CrI) Dropout rate (OR with 95% CrI)

relative treatment effects for dropout rate
OR>1 favor the treatment in column

significant effects are in bold and competing treatments *underscored font*

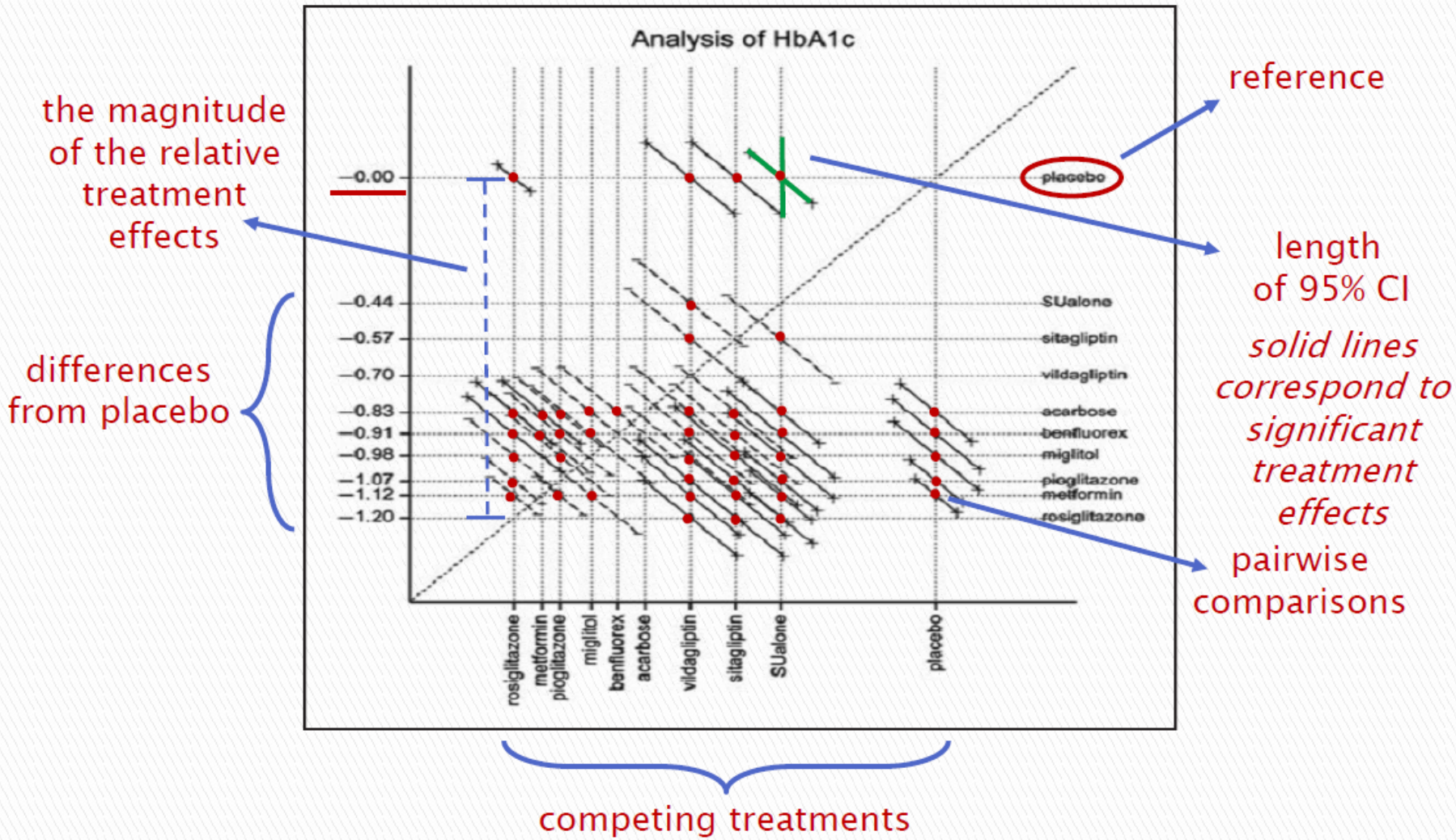
Table showing all the pairwise relative treatment effects with their 95% CI for one or two outcomes

[Example in Cipriani et al. 2011]



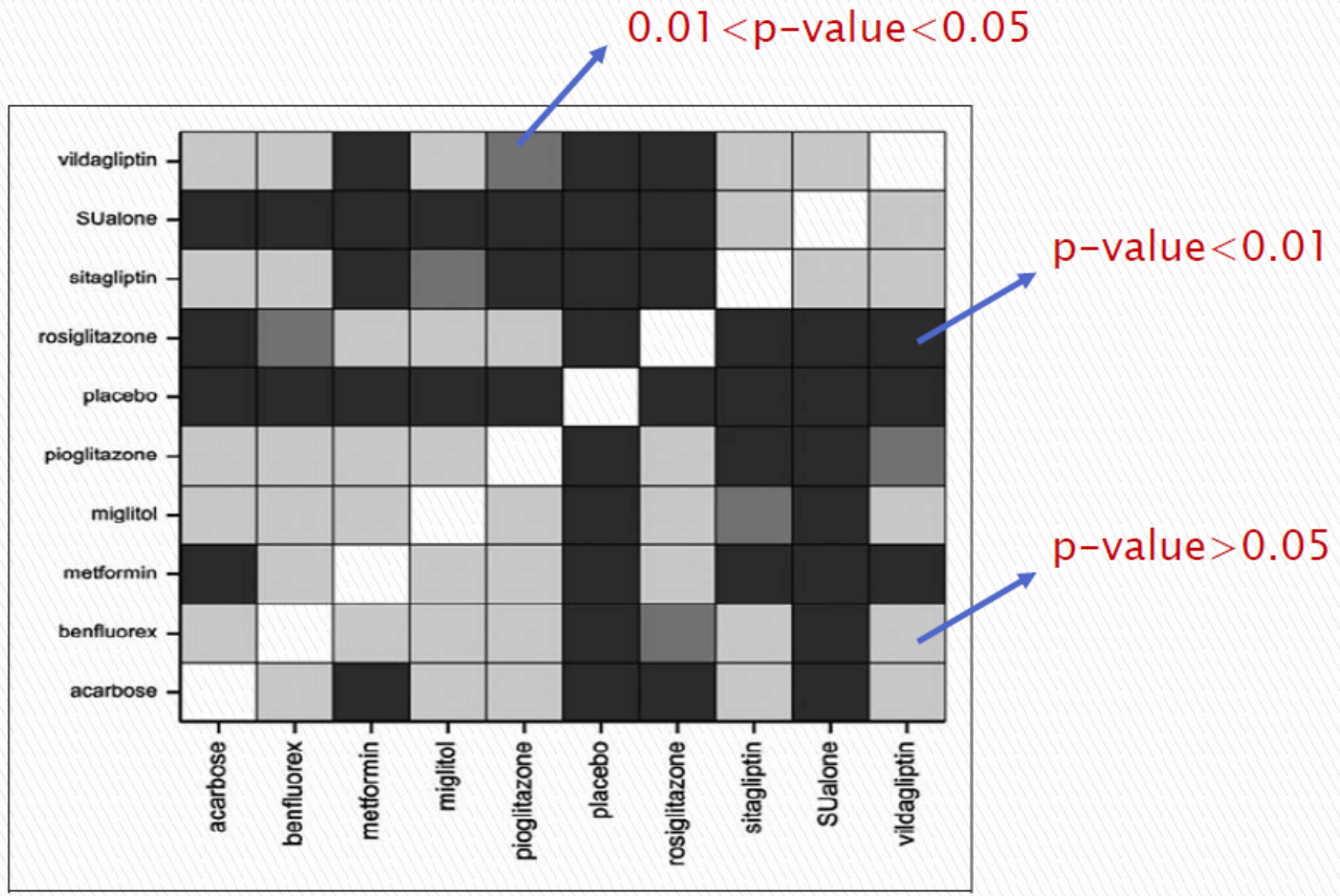
Forest plot with the treatment effects estimates for the pairwise comparisons

[Examples in Hawkins et al. 2009 & Hoaglin et al. 2011]



'Hsu mean-mean plot' showing the network estimates with the 95% CI for all pairwise comparisons

[Example in Senn et al. 2013]

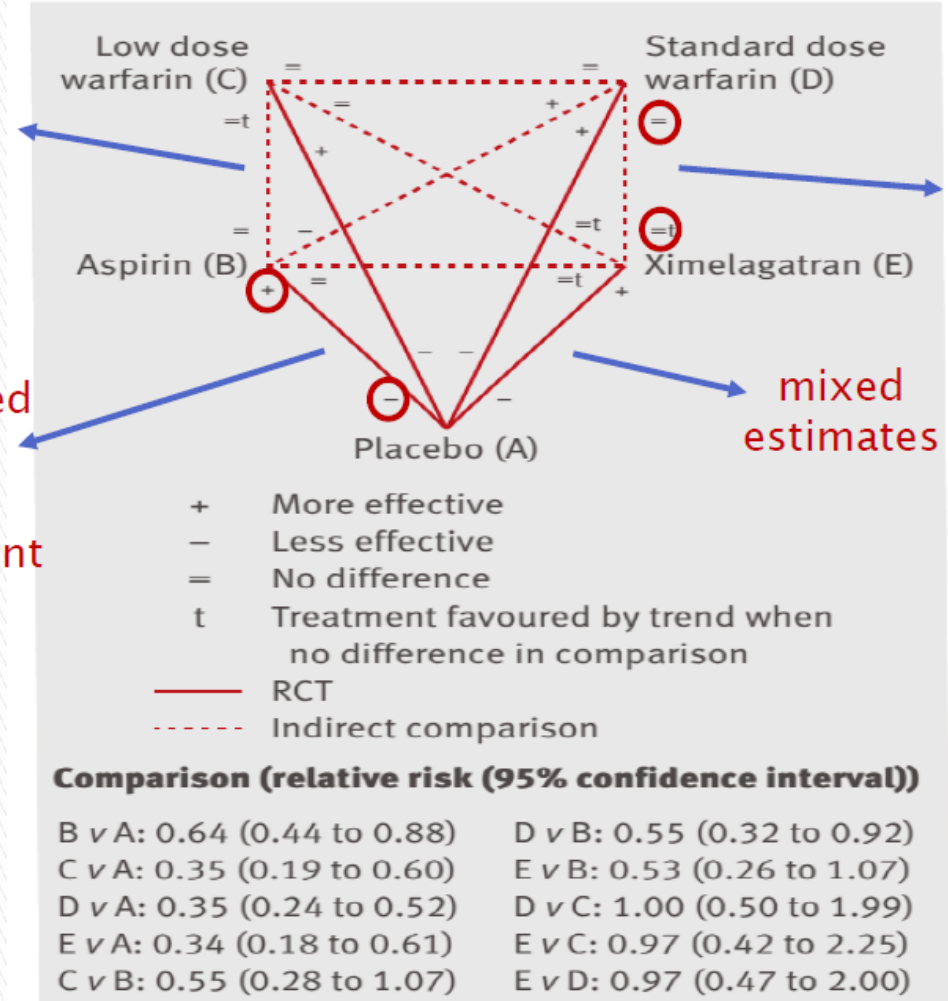


Shade plot showing the p-values of the treatment effects for all pairwise comparisons in the network

[Example in Senn et al. 2013]

indirect estimates

in comparisons with significant relative treatment effects (based on the network estimates):
'+' the favored treatment
'–' the non-favored treatment



mixed estimates

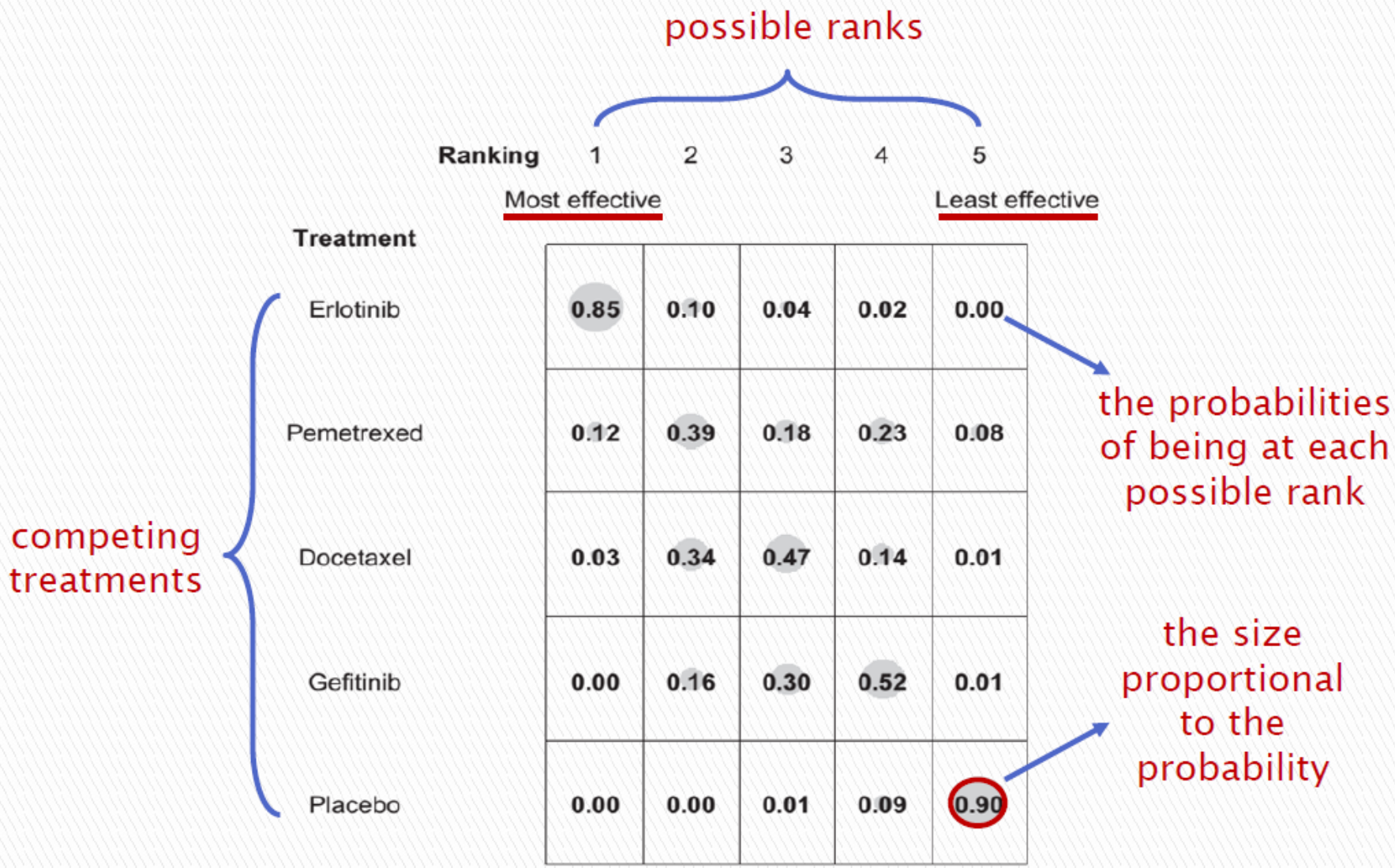
in comparisons with non-significant relative treatment effects (based on the network estimates):
'=t' the favored treatment
'=' the non-favored treatment

Network graph presenting the relative treatment effects for each pairwise comparison

[Example in Fadda et al. 2011]

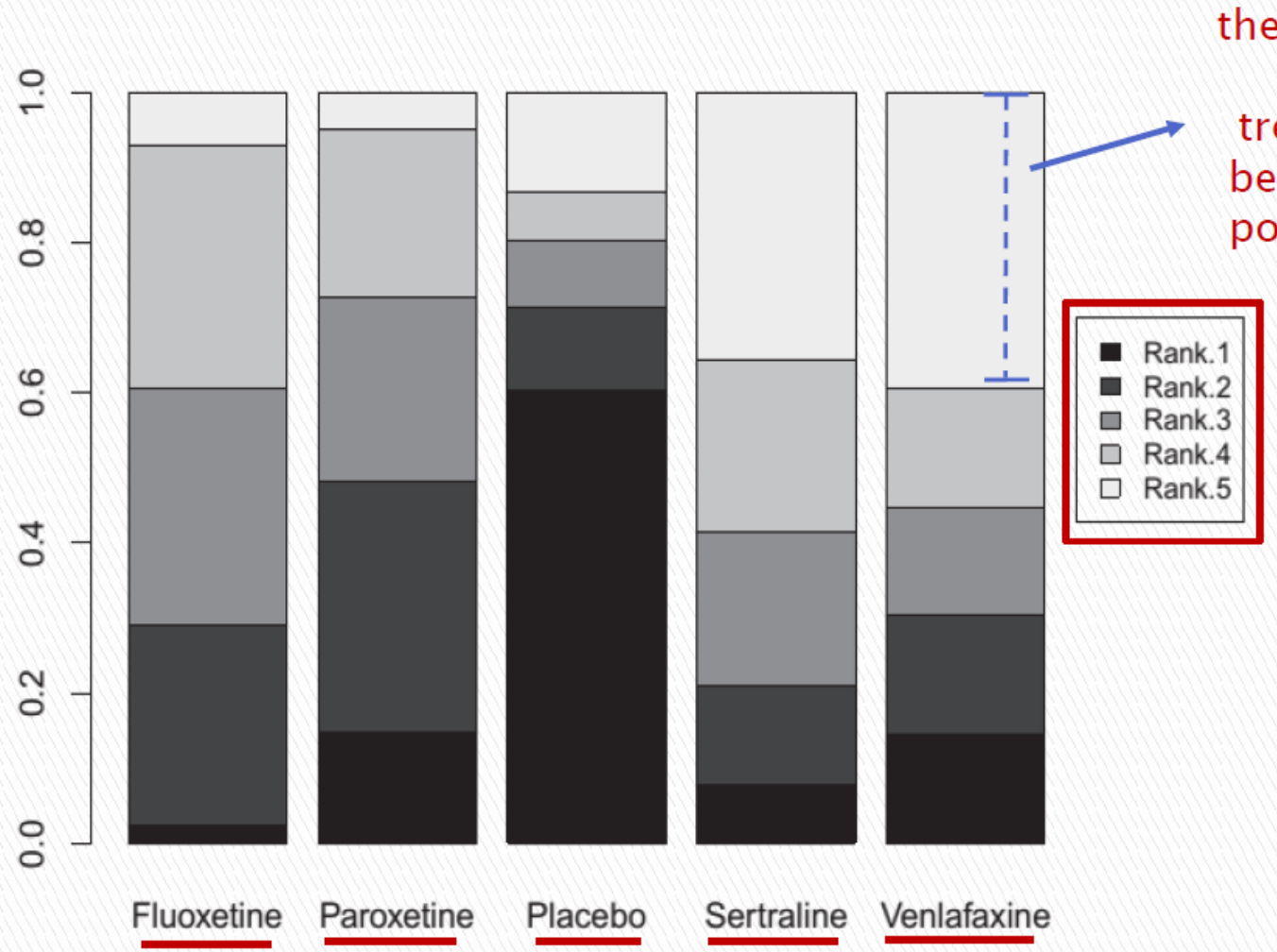
Presenting the results ranking

- Using probability of being the best
- Using probabilities of being at each possible rank
- Using SUCRAS



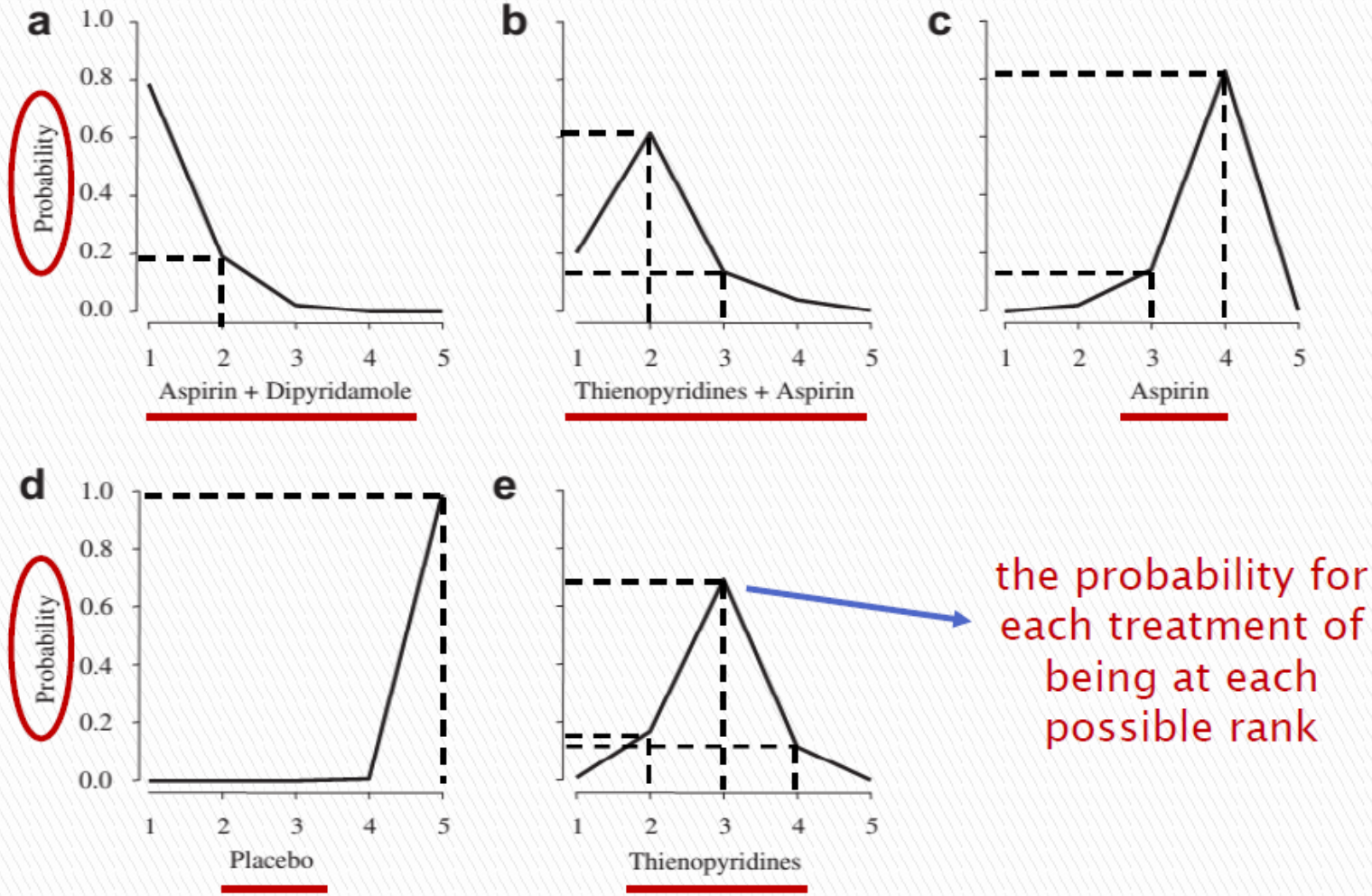
'Bubble-plot' including the ranking probabilities for all treatments

[Example in Hawkins et al. 2009]



Bar plots showing the probability for each treatment of being at a specific rank

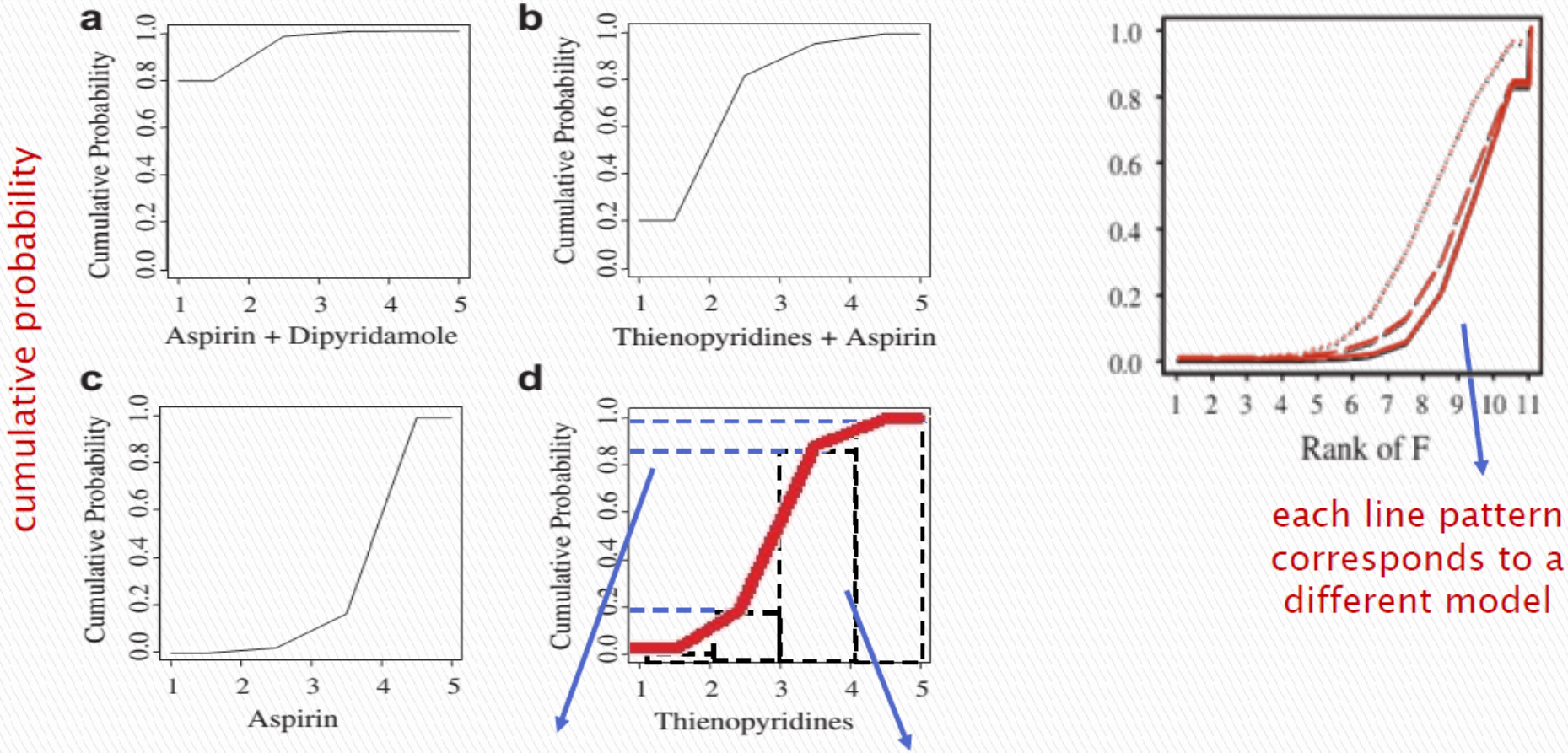
[Example in van Valkenhoef et al. 2012]



the probability for each treatment of being at each possible rank

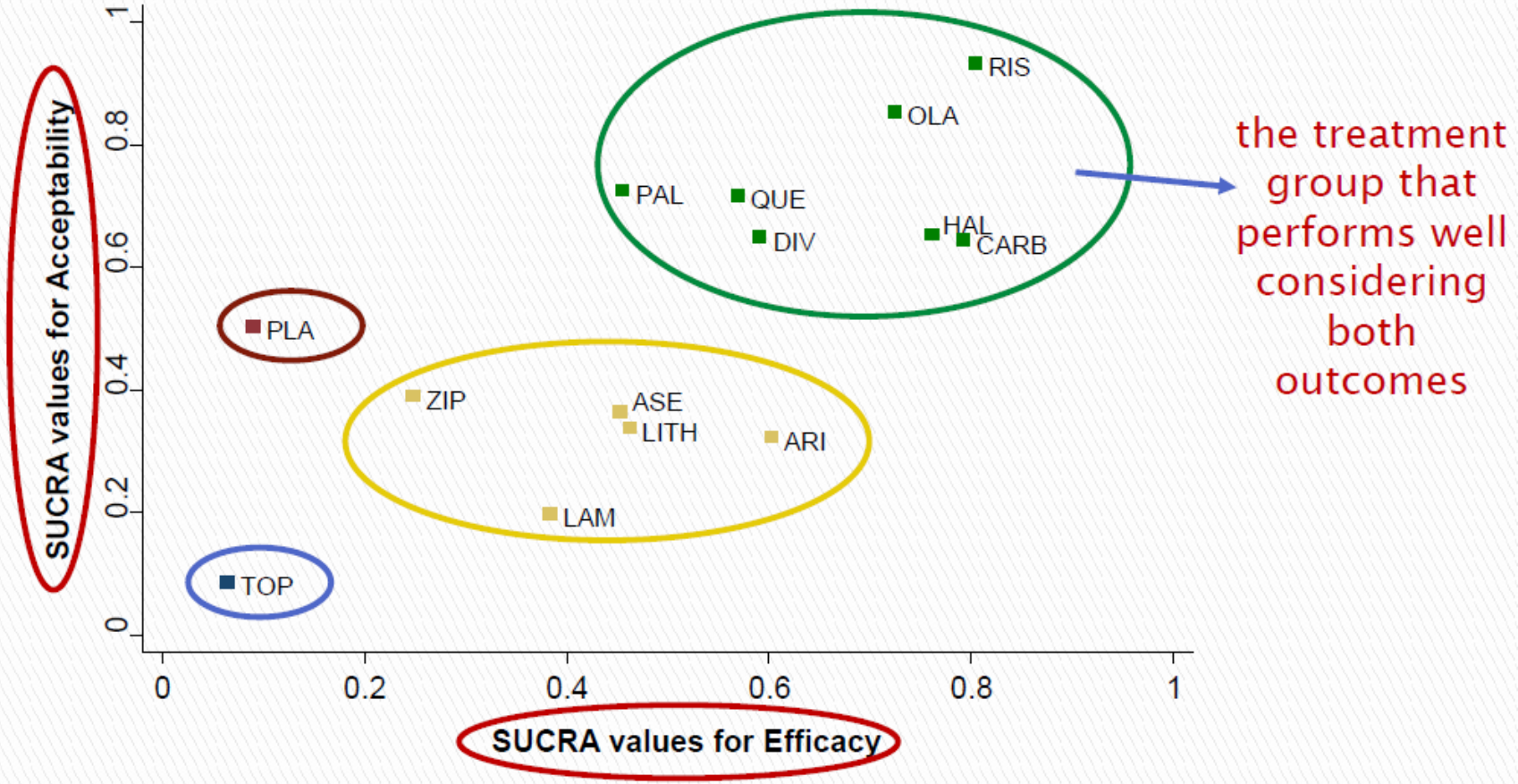
'Rankograms' showing the probability for each treatment of being at a specific rank

[Example in Salanti et al. 2011]



'SUCRA plots' showing the cumulative probability for each treatment of being up to a specific rank

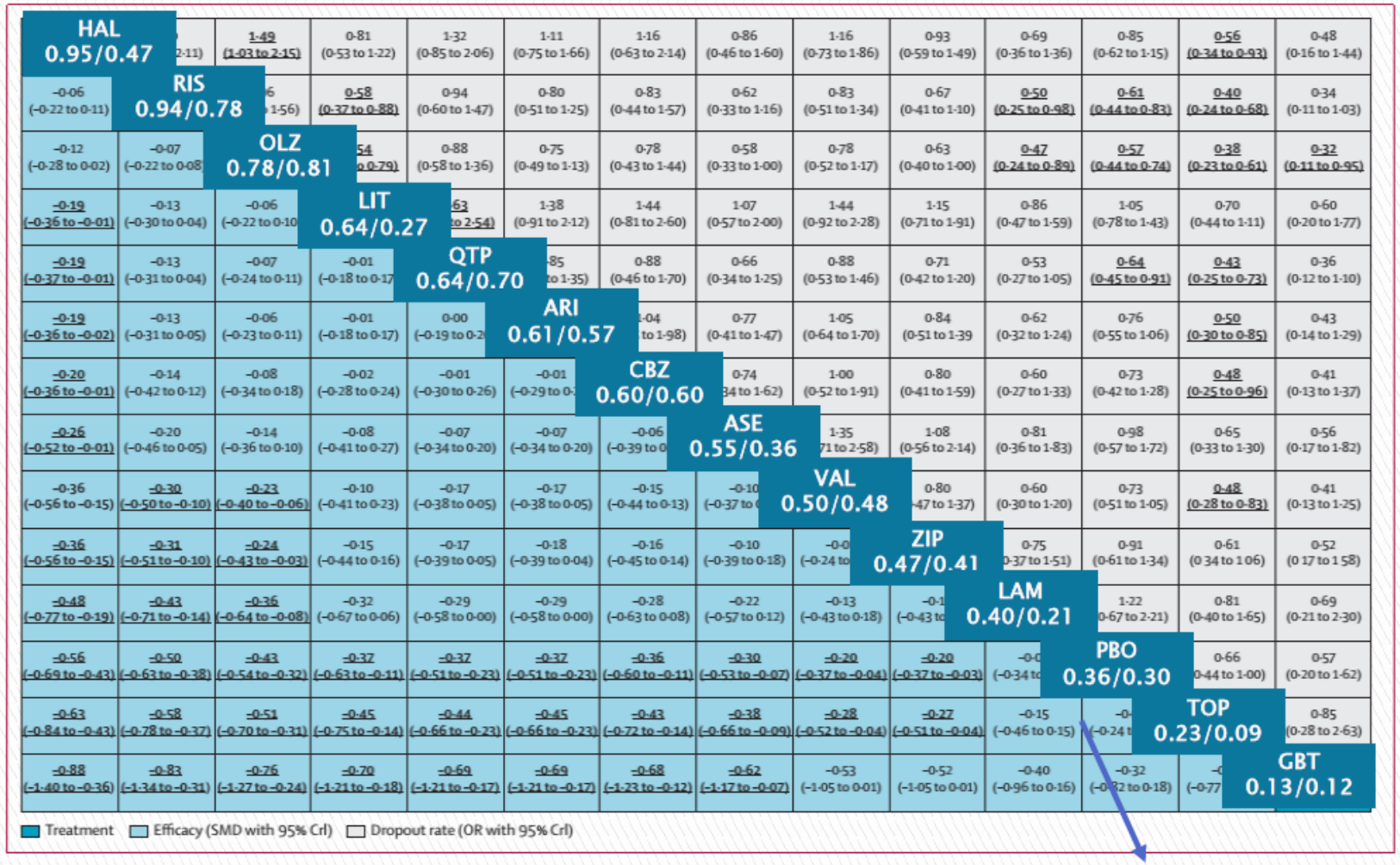
[Examples in Salanti et al. 2011 & Salanti et al. 2010]



the treatment group that performs well considering both outcomes

Scatterplot showing jointly the ranking results for two different outcomes

[Example in Chaimani et al. 2013]



competing treatments ordered according to their relative ranking for efficacy

Table showing all the pairwise relative treatment effects with their 95% CI for one or two outcomes along with the SUCRA values

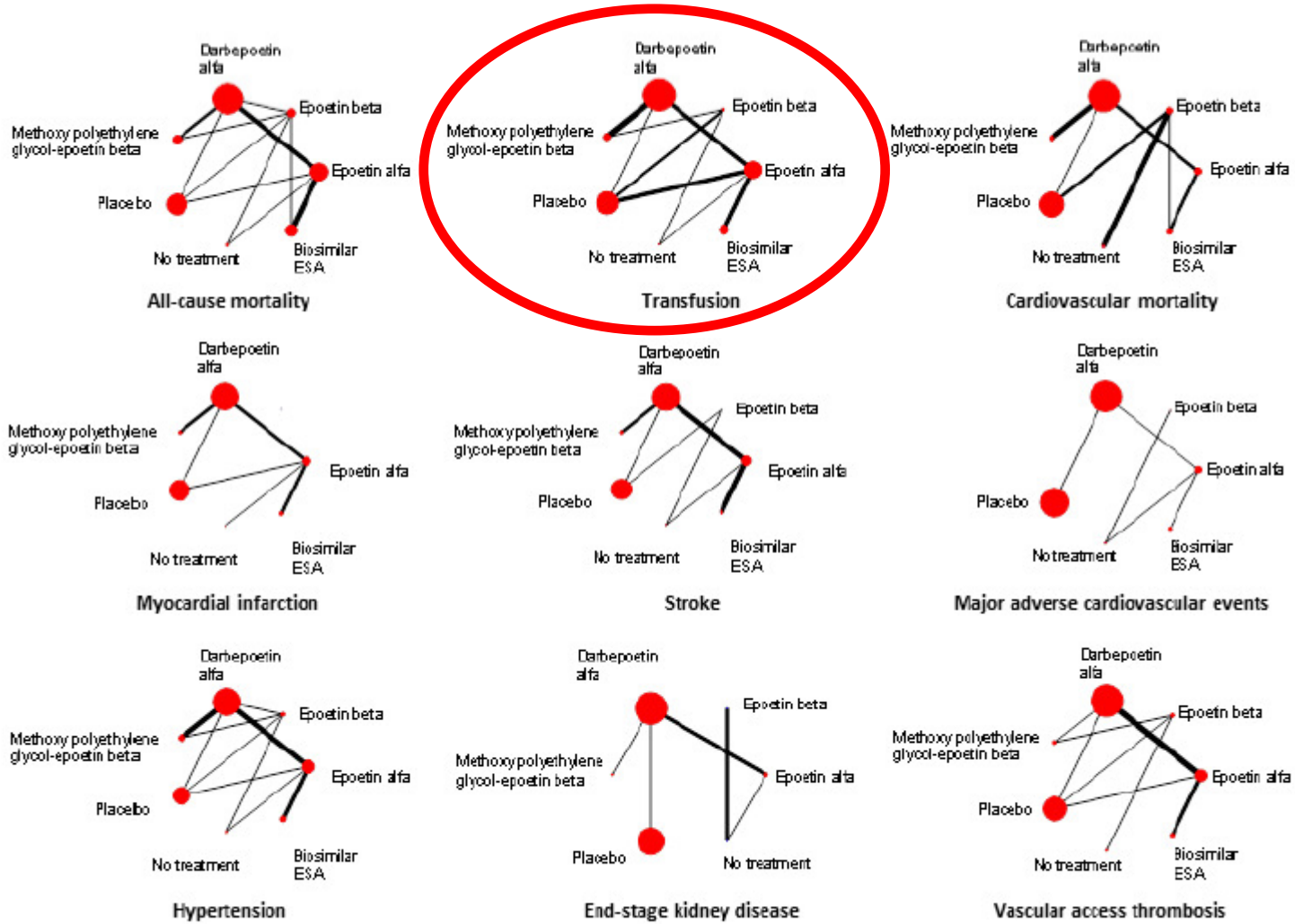
Erythropoiesis-stimulating agents for anaemia in adults with chronic kidney disease: a network meta-analysis (Review)

Palmer SC, Saglimbene V, Mavridis D, Salanti G, Craig JC, Tonelli M, Wiebe N, Strippoli GFM

OBJECTIVES

To compare the efficacy and safety of ESAs (epoetin alfa, epoetin beta, darbepoetin alfa, or methoxy polyethylene glycol-epoetin beta, and biosimilar ESAs, against each other, placebo, or no treatment) to treat anaemia in adults with CKD.

Figure 5. Networks of the treatment efficacy and safety of ESA drugs in the treatment of anaemia in chronic kidney disease. Values lower than 1 favour the active treatment in the comparison



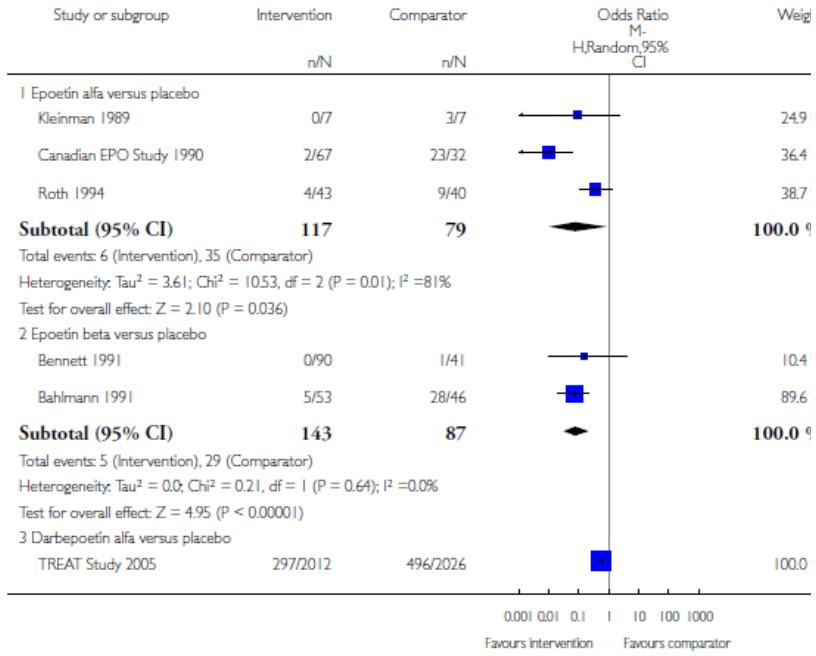
Assessment of clinical and methodological

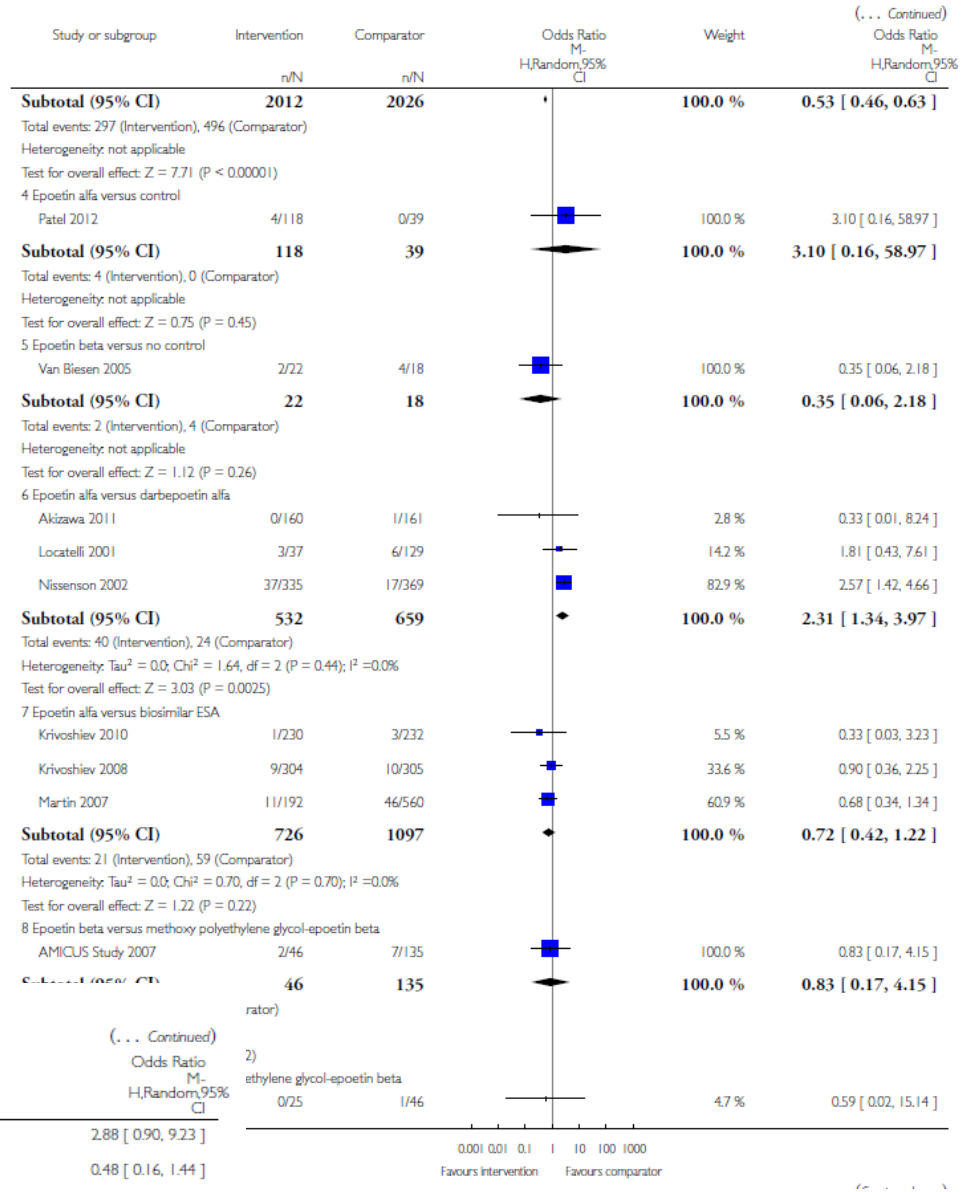
Analysis 1.1. Comparison 1 ESA versus ESA or placebo/no treatment, Outcom

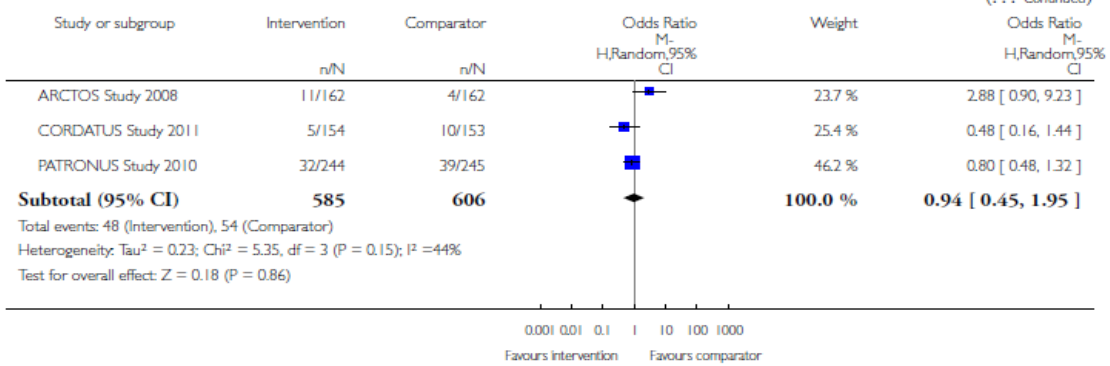
Review: Erythropoiesis-stimulating agents for anaemia in adults with chronic kidney disease: a network meta-analysis

Comparison: 1 ESA versus ESA or placebo/no treatment

Outcome: 1 Blood transfusion







Assessment of similarity (transitivity) across treatment comparisons

Evaluation of the assumption is important and its plausibility determines the validity of the network meta-analysis results.

We inferred about the assumption of transitivity:

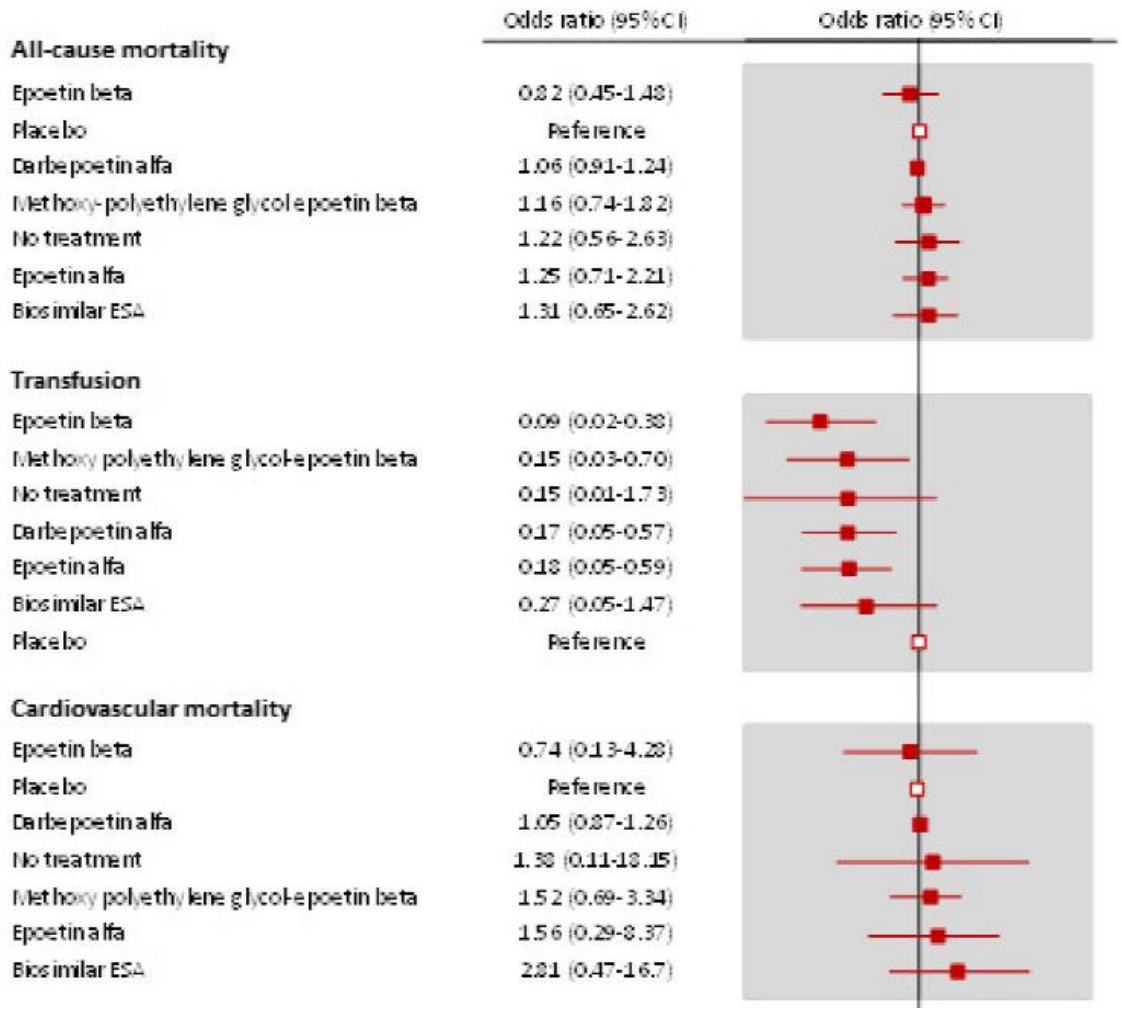
1. We assessed whether the included interventions were similar when they were evaluated in studies with different designs, for example, whether ESAs are administered the same way in studies comparing ESAs to placebo and in those comparing ESAs to other ESAs

2. We compared the distribution of the potential effect modifiers (age, stage of CKD, duration of treatment) across the different pairwise comparisons.

The inconsistency factor is the absolute difference in the log odds ratio estimated from indirect and direct treatment comparisons and is reported together with the 95% confidence interval. A 95% confidence interval that includes zero indicates that the result is compatible with zero inconsistency between effect estimates using indirect (networkmeta-analysis) and direct (conventional pairwise meta-analysis) treatment comparisons.

Transfusion		
Epoetin alfa - epoetin beta - placebo – no treatment	2.09	0.00-6.91
Epoetin alfa - darbepoetin alfa - placebo	1.97	0.00-4.20
Epoetin beta - darbepoetin alfa – methoxy polyethylene glycol-epoetin beta - placebo	1.26	0.00-3.39

Figure 6. Forest plots for results from network meta-analyses comparing ESAs versus placebo



Annals of Internal Medicine RESEARCH AND REPORTING METHODS

The PRISMA Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-analyses of Health Care Interventions: Checklist and Explanations

Brian Hutton, PhD, MSc; Georgia Salanti, PhD; Deborah M. Caldwell, PhD, MA, BA; Anna Chaimani, PhD; Christopher H. Schmid, PhD; Chris Cameron, MSc; John P.A. Ioannidis, MD, DSc; Sharon Straus, MD, MSc; Kristian Thorlund, PhD; Jeroen P. Jansen, PhD; Cynthia Mulrow, MD, MSc; Ferrán Catalá-López, PhD, MPH, PharmD; Peter C. Gøtzsche, MD, MSc; Kay Dickersin, PhD, MA; Isabelle Boutron, MD, PhD; Douglas G. Altman, DSc; and David Moher, PhD

The PRISMA statement is a reporting guideline designed to improve the completeness of reporting of systematic reviews and meta-analyses. Authors have used this guideline worldwide to prepare their reviews for publication. In the past, these reports typically compared 2 treatment alternatives. With the evolution of systematic reviews that compare multiple treatments, some of them only indirectly, authors face novel challenges for conducting and reporting their reviews. This extension of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) statement was developed specifically to improve the reporting of systematic reviews incorporating network meta-analyses.

A group of experts participated in a systematic review, Delphi survey, and face-to-face discussion and consensus meeting to establish new checklist items for this extension statement. Current PRISMA items were also clarified. A modified, 32-item PRISMA extension checklist was developed to address what the group considered to be immediately relevant to the reporting of network meta-analyses.

This document presents the extension and provides examples of good reporting, as well as elaborations regarding the rationale for new checklist items and the modification of previously existing items from the PRISMA statement. It also highlights educational information related to key considerations in the practice of network meta-analysis. The target audience includes authors and readers of network meta-analyses, as well as journal editors and peer reviewers.

Ann Intern Med. 2015;162:777-784. doi:10.7326/M14-2385 **www.annals.org**
For author affiliations, see end of text.

Table. Checklist of Items to Include When Reporting a Systen

Section/Topic	Item # *	Checklist Item†
TITLE		
Title	1	Identify the report as a systemati *meta-analysis).*
ABSTRACT		
Structured summary	2	Provide a structured summary in Background: main objectives Methods: data sources; study and *synthesis methods, such* Results: number of studies anc confidence/credible interval *to summarize pairwise comp brevity.* Discussion/Conclusions: limita Other: primary source of fundi
INTRODUCTION		
Rationale	3	Describe the rationale for the re *why a network meta-analysis*
Objectives	4	Provide an explicit statement of interventions, comparisons,
METHODS		
Protocol and registration	5	Indicate whether a review protoc and, if available, provide reg
Eligibility criteria	6	Specify study characteristics (e.g years considered, language, *Clearly describe eligible trea have been clustered or merg*
Information sources	7	Describe all information sources authors to identify additiona
Search	8	Present full electronic search str it could be repeated.
Study selection	9	State the process for selecting st and, if applicable, included
Data collection process	10	Describe method of data extract and any processes for obtai
Data items	11	List and define all variables for w assumptions and simplificati
Geometry of the network	S1	Describe methods used to explo potential biases related to it summarized for presentatio the evidence base to reader
Risk of bias within individual studies	12	Describe methods used for asse whether this was done at the in any data synthesis.
Summary measures	13	State the principal summary mea *of additional summary meas cumulative ranking curve (SU summary findings from meta*
Planned methods of analysis	14	Describe the methods of handlin meta-analysis. This should in *Handling of multigroup trials; Selection of variance structure; Selection of prior distributions Assessment of model fit.*
Assessment of inconsistency	S2	Describe the statistical methods the treatment network(s) stu
Risk of bias across studies	15	Specify any assessment of risk of bias, selective reporting with
Additional analyses	16	Describe methods of additional include, but not be limited t Sensitivity or subgroup analys Meta-regression analyses; *Alternative formulations of the Use of alternative prior distributions for Bayesian analyses (if applicable).*

RESEARCH AND REPORTING METHODS

PRISMA Extension for Network Meta-analysis

Table–Continued

Section/Topic	Item # *	Checklist Item†	Reported on Page #
RESULTS‡			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	
Presentation of network structure	S3	Provide a network graph of the included studies to enable visualization of the geometry of the treatment network.	
Summary of network geometry	S4	Provide a brief overview of characteristics of the treatment network. This may include commentary on the abundance of trials and randomized patients for the different interventions and pairwise comparisons in the network, gaps of evidence in the treatment network, and potential biases reflected by the network structure.	
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment.	
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: 1) simple summary data for each intervention group, and 2) effect estimates and confidence intervals. *Modified approaches may be needed to deal with information from larger networks.*	
Synthesis of results	21	Present results of each meta-analysis done, including confidence/credible intervals. *In larger networks, authors may focus on comparisons versus a particular comparator (e.g., placebo or standard care), with full findings presented in an appendix. League tables and forest plots may be considered to summarize pairwise comparisons.* If additional summary measures were explored (such as treatment rankings), these should also be presented.	
Exploration for inconsistency	S5	Describe results from investigations of inconsistency. This may include such information as measures of model fit to compare consistency and inconsistency models, *P* values from statistical tests, or summary of inconsistency estimates from different parts of the treatment network.	
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies for the evidence base being studied.	
Results of additional analyses	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression analyses, *alternative network geometries studied, alternative choice of prior distributions for Bayesian analyses,* and so forth).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings, including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., health care providers, researchers, and policymakers).	
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review level (e.g., incomplete retrieval of identified research, reporting bias). *Comment on the validity of the assumptions, such as transitivity and consistency. Comment on any concerns regarding network geometry (e.g., avoidance of certain comparisons).*	
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review. This should also include information regarding whether funding has been received from manufacturers of treatments in the network and/or whether some of the authors are content experts with professional conflicts of interest that could affect use of treatments in the network.	

(Continued on following page)

La valutazione della certezza delle prove

(M. Cinquini)

The main consideration for study limitations in a network meta-analysis is to ensure that the relative contributions of different sources of direct evidence (which may have different study limitations) are accounted for appropriately

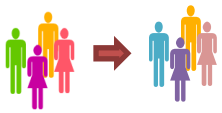
Salanti G. et al. Plos One 2014; 9: e99682

Determinants of certainty in a body of evidence

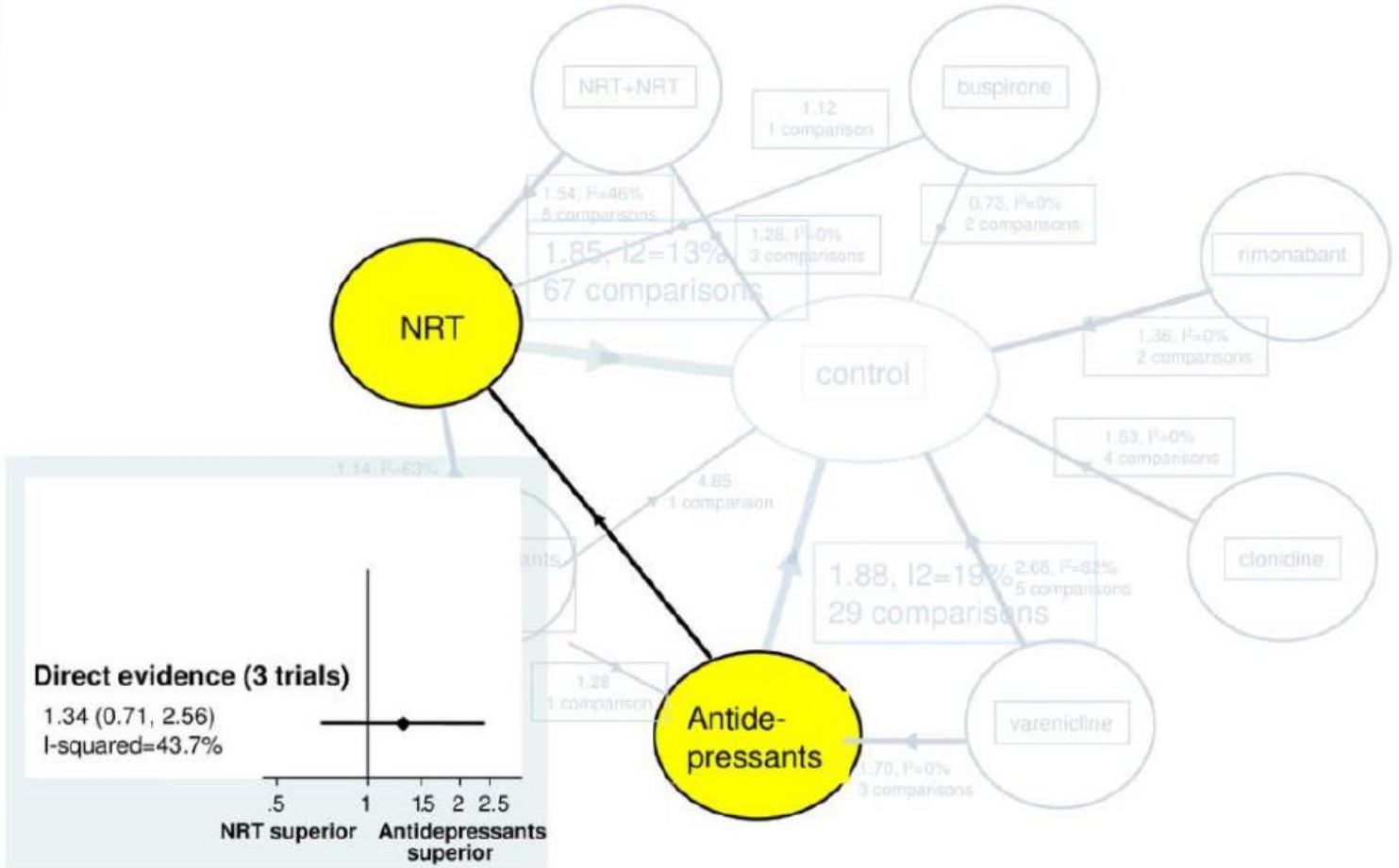
GRADE

- A body of evidence starts as: high | ⨁⨁⨁⨁
- 5 factors that can lower quality
  1. Risk of bias criteria
  2. Inconsistency (or heterogeneity)
  3. Indirectness (PICO and applicability)
  4. Imprecision
  5. Publication bias

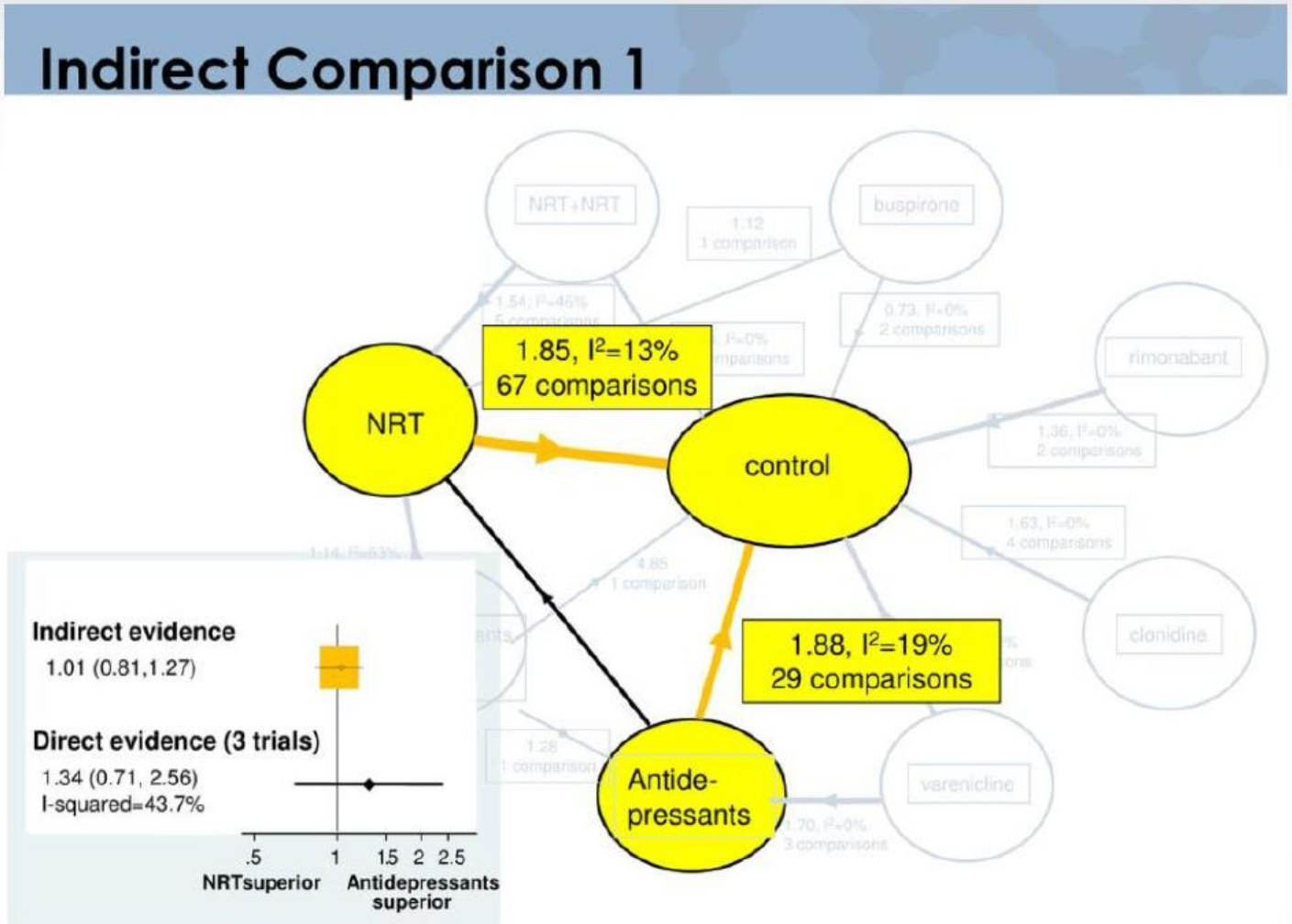


Direct Comparison

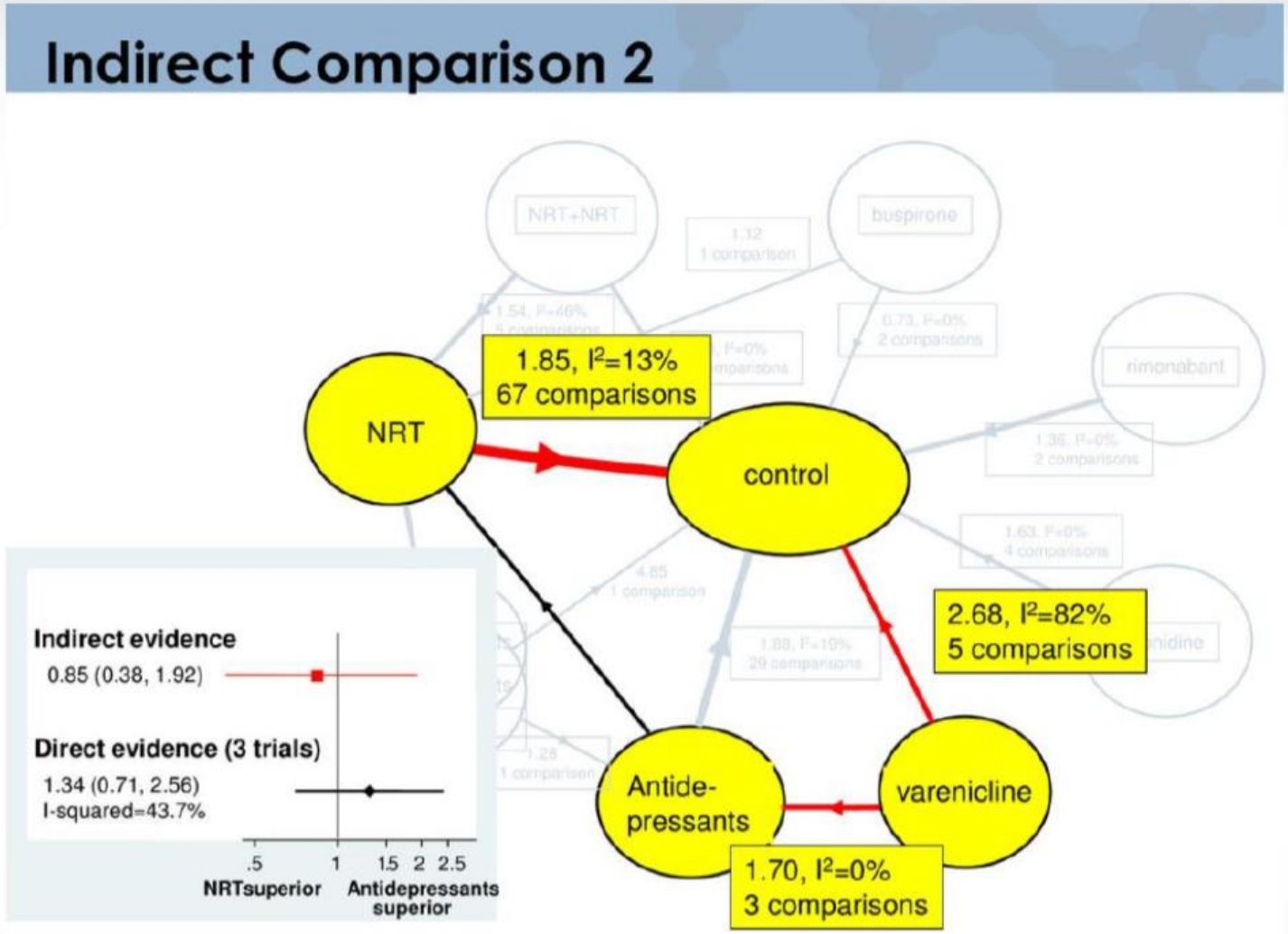


Indirect Comparison 1



1° order loop

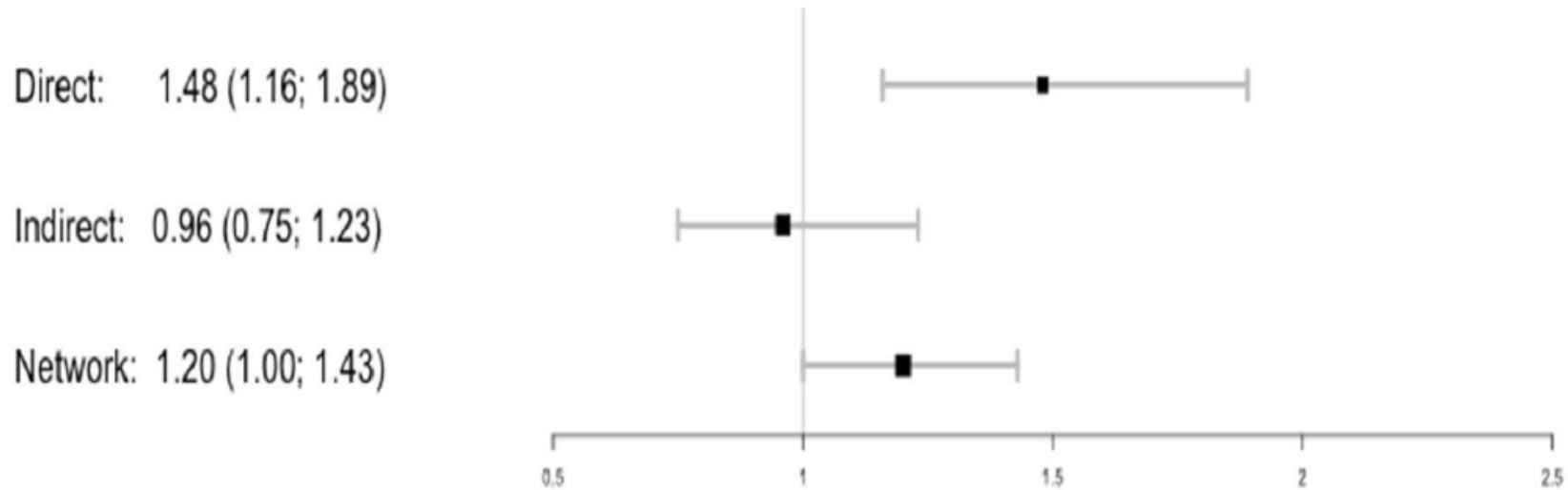
Indirect Comparison 2



2° order loop

Steps for assessing certainty in NMA

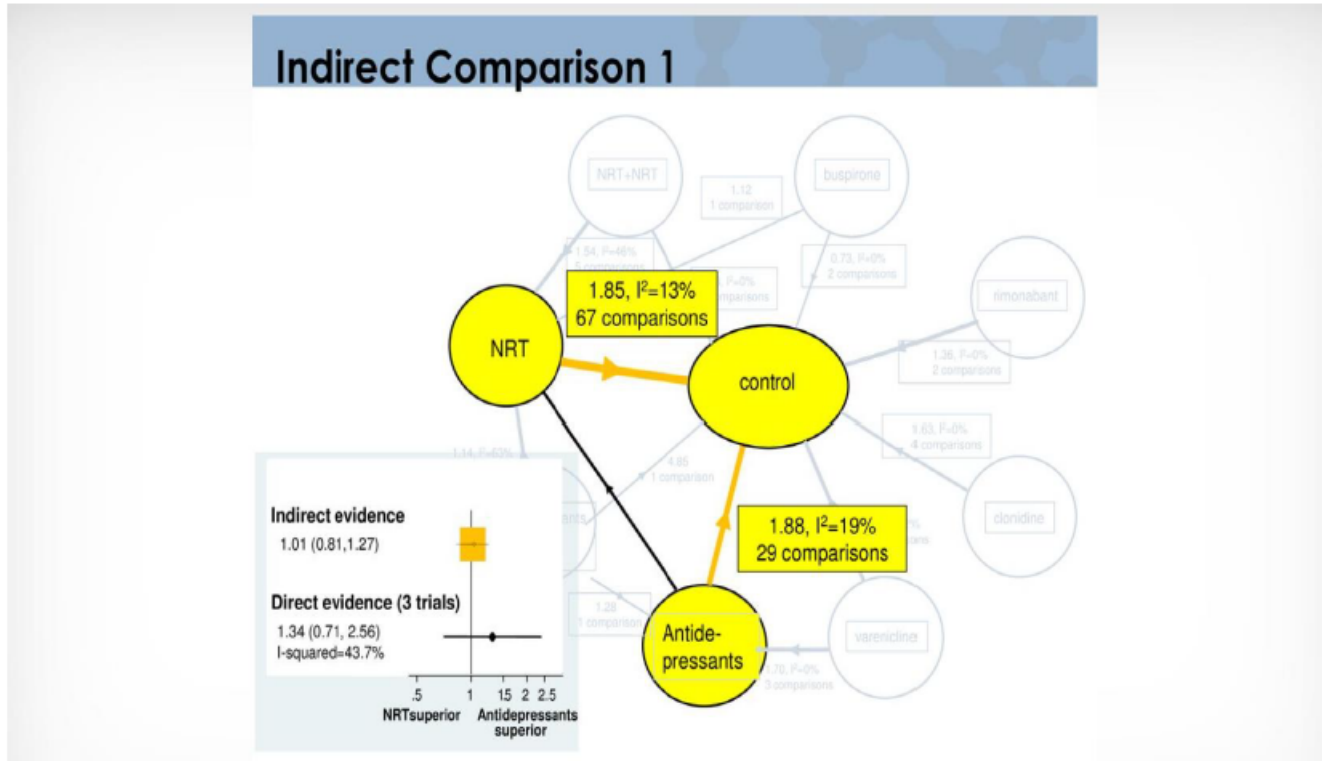
- **Step 1: Presenting direct and indirect effect estimates and 95% CI**
- Making valid inferences on the basis of a NMA (valid judgment about certailty) requires understanding how much direct and indirect evidence contribute to the NMA effect estimates
- **Node splitting** approach which separates evidence on a particular comparison (a "node") into direct and indirect estimates of treatment effect



Steps for assessing certainty in NMA

- **Step 2: Rating of quality of direct and indirect effect estimates for all domains except imprecision**
- To keep the quality rating of the indirect evidence manageable,

we suggest a **focus on first order loops**, which usually contribute most information to the indirect estimate.



Steps for assessing certainty in NMA

Presenting and rating of quality of NMA effect estimates

- **Steps 3 :** Use the certainty of evidence of direct or indirect estimate evidence on the basis of which contribute the most to NMA evidence (node splitting) for all domains except imprecision
- **Step 4**: Assess imprecision of the NMA estimate

Sintesi percorso per valutare certezza evidenza NMA

Se c'è solo evidenza diretta

1. Valutare certezza evidenza diretta (dalle MA pairwise per tutti i domini tranne imprecision)
2. Valutare imprecisione della stima NMA , non pairwise (Il contributo della evidenza che deriva da tutto il network può comunque aumentare la precisione della stima. Questa è una delle ragioni per cui si fa una NMA)

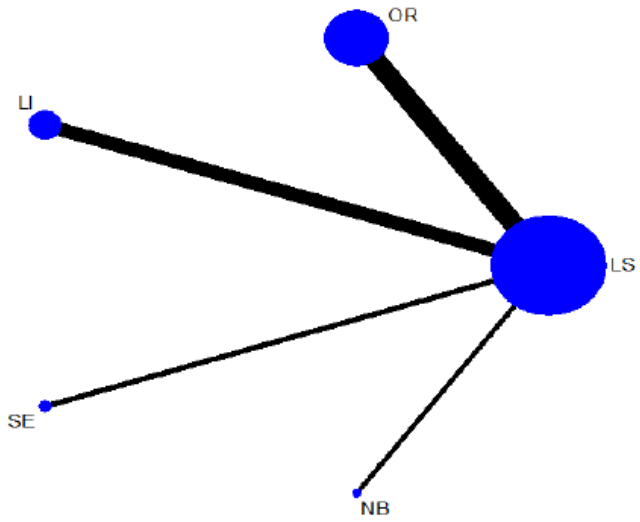
Approccio non contestualizzato: si abbassa per imprecisione se i CI crossano la linea di non effetto

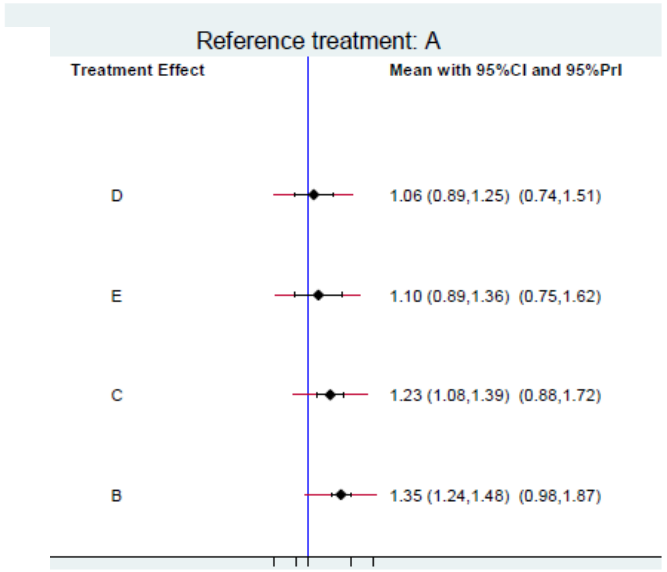
Approccio parzialmente contestualizzato : gli autori della RS devono stabilire a priori le soglie per effetto trivial, piccolo, modesto, grande. Si contano il numero di soglie che vengono attraversate dai CI;

se crossano una soglia si abbassa di un livello,

se crossano due soglie si abbassa di due livelli

se crossano 3 o + soglie si abbassa di 3 livelli





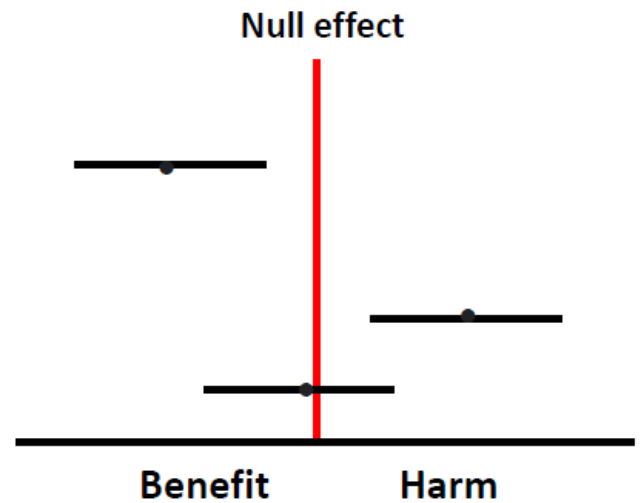
Imprecisione della stima NMA

Non è possibile calcolare OIS perché non si conosce il sample size della NMA

Approccio non contestualizzato:

Non interessa la dimensione dell'effetto; solo la direzione

si abbassa per imprecision se i CI crossano la linea di non effetto



Imprecisione della stima NMA

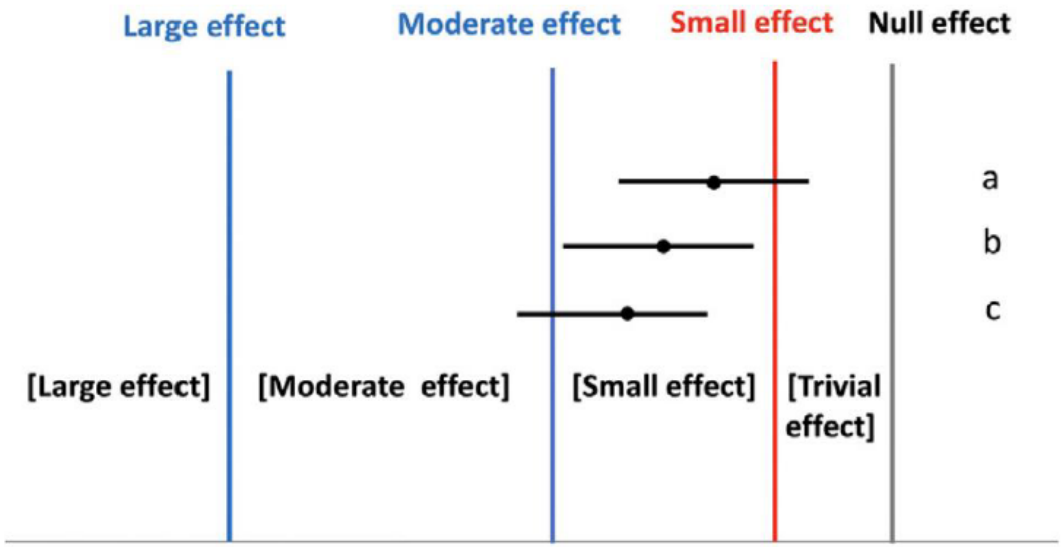
Approccio Contestualizzato: interessa la dimensione dell'effetto: irrilevante, piccola, moderata, grande

I membri del panel devono stabilire le soglie (valori dell'esito di interesse) per effetto irrilevante, piccolo, moderato, grande, (possibilmente basandosi su dati della letteratura). Si contano il numero di soglie che vengono attraversate dai CI:

se crossano una soglia si abbassa di un livello

se crossano due soglie si abbassa di 2 livelli

se crossano 3 o + soglie si abbassa di 3 livelli



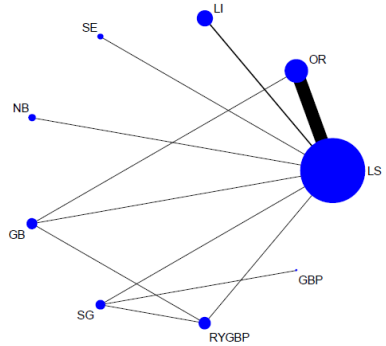
- **Se c’è solo evidenza indiretta**

Si considerano solo le due comparison del primo loop (se sono interessato ad B vs C, considero le pairwise di A vs B e di A vs C

Si valuta certainty delle due comparison (pairwise) indirette del primo loop per tutte le dimensioni tranne imprecision.

Si considera la certezza più bassa tra le due

Si valuta imprecisione della stima della NMA come sopra



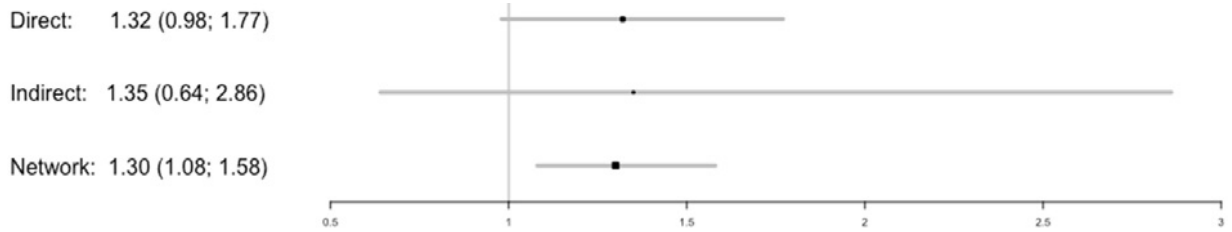
- **Se c’è evidenza mista**

Devo vedere quale delle due certezze contribuisce di più alla stima network

- **Se una stima (diretta o indiretta) contribuisce di più alla stima network**

Valuto la certezza per tutte le dimensioni tranne imprecision della evidenza che contribuisce di più seguendo gli approcci descritti sopra

Si valuta imprecisione della stima della NMA come sopra



- **Se le due stime contribuiscono in egual misura**

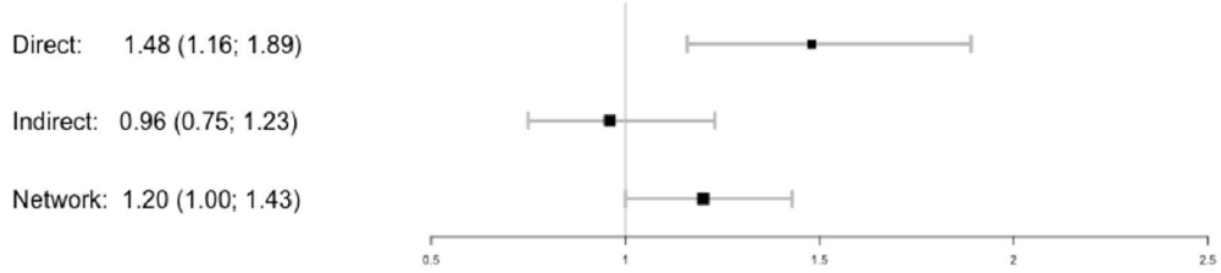
devo vedere se sono coerenti

- **se sono coerenti**:

valuto certezza di entrambe per tutte le dimensioni tranne imprecisione

considero quella con certezza più alta

Si valuta imprecisione della stima della NMA come sopra



- **Se non sono coerenti**

Procedo come sopra ma abbasso ulteriormente per incoherence

NMA-SoF table example 2

Estimates of effects, credible intervals, and certainty of the evidence for chemoprevention of colorectal cancer in individuals with previous colorectal neoplasia

Bayesian NMA-SoF table

BENEFITS

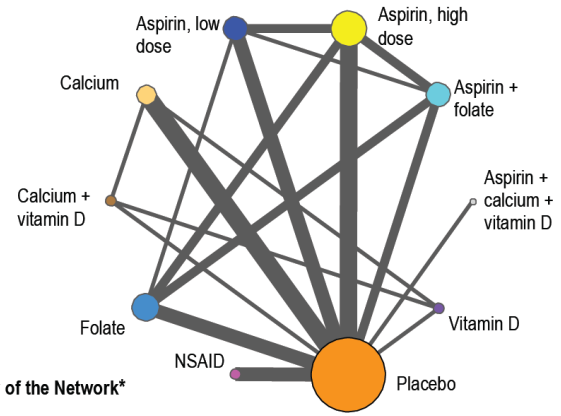
Patient or population: Individuals with previous colorectal neoplasia

Interventions: Low and high dose aspirin, nonaspirin non-steroidal anti-inflammatory drugs (NSAIDs), calcium, vitamin D, folic acid

Comparator (reference): Placebo

Outcome: Prevention of advanced neoplasia; range of follow up between three to five years

Setting: Outpatient



Geometry of the Network*

Total studies: 21 RCT **Total Participants**: 12088		**Relative effect** (95% CrI)	**Anticipated absolute effect*** (95% CrI)			**Certainty of evidence**	**Ranking**** (95% CrI)	**Interpretation of Findings**
			***Without* intervention**	***With* intervention**	**Difference**			
●	Aspirin + calcium + vitamin D (1 RCT; 427 participants)	**0.71** (0.18 to 2.49) Network estimate	74 per 1000[1]	53 per 1000	21 fewer per 1000 (61 fewer to 110 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 5]	3 (1 to 10)	Probably inferior
●	Calcium + vitamin D (1 RCT; 1028 participants)	**0.91** (0.52 to 1.63) Network estimate	74 per 1000[1]	67 per 1000	7 fewer per 1000 (36 fewer to 47 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 5]	6 (1 to 10)	Probably inferior
●	Aspirin + folate (2 RCT; 916 participants)	**0.73** (0.43 to 1.19) Network estimate	74 per 1000[1]	54 per 1000	20 fewer per 1000 (42 fewer to 14 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 5]	4 (2 to 8)	Probably inferior
●	Aspirin, high dose (3 RCT; 917 participants)	**0.81** (0.50 to 1.28) Network estimate	74 per 1000[1]	60 per 1000	14 fewer per 1000 (37 fewer to 21 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 5]	5 (2 to 9)	Probably inferior

NMA-SoF table example 2

●	Aspirin, low dose (3 RCT; 823 participants)	**0.71** (0.41 to 1.23) Network estimate	74 per 1000[1]	53 per 1000	21 fewer per 1000 (44 fewer to 17 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 5]	3 (2 to 9)	Probably inferior
●	Nonaspirin NSAIDs (4 RCT; 3486 participants)	**0.37** (0.24 to 0.53) Network estimate	74 per 1000[1]	27 per 1000	47 fewer per 1000 (56 fewer to 35 fewer)	⊕⊕⊕⊕ **High**[5]	1 (1 to 2)	Definitely superior
●	Vitamin D (1 RCT; 764 participants)	**1.19** (0.65 to 2.15) Network estimate	74 per 1000[1]	88 per 1000	14 more per 1000 (26 fewer to 85 more)	⊕⊕◯◯ **Low** Due to Imprecision[3, 5]	9 (3 to 10)	Probably inferior
●	Calcium (3 RCT; 2503 participants)	**1.00** (0.66 to1.52) Network estimate	74 per 1000[1]	74 per 1000	0 fewer per 1000 (25 fewer to 38 more)	⊕⊕◯◯ **Low** Due to Imprecision[4, 5]	7 (3 to 10)	Probably inferior
●	Folate (3 RCT; 1224 participants)	**1.32** (0.85 to 2.00) Network estimate	74 per 1000[1]	51 per 1000	23 more per 1000 (11 fewer to 74 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 5]	9 (5 to 10)	Probably inferior
●	Placebo	Reference comparator	No estimable	No estimable	No estimable	Reference comparator	7 (4 to 9)	Reference comparator

NMA-SoF table definitions

* Lines represent direct comparisons

** Estimates are reported as odds ratio. CrI: credible interval. Results are expressed in credible intervals as opposed to the confidence intervals (CI) since a Bayesian analysis has been conducted.

*** Anticipated absolute effect. Anticipated absolute effect compares two risks by calculating the difference between the risks of the intervention group with the risk of the control group.

**** Surface under the cumulative (SUCRA) ranking and credible intervals for efficacy are presented. Rank statistics is defined as the probabilities that a treatment out of *n* treatments in a network meta-analysis is the best, the second, the third and so on until the least effective treatment.

GRADE Working Group grades of evidence (or certainty in the evidence)

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanatory Footnotes

[1] Baseline risks (assumed control risk) obtained from the National Cancer Institute pooling project

[2] Very serious imprecision since 95% CrI crosses unity, and with wide credible intervals suggesting high possibility of harm.

[3] Very serious imprecision since RR>1 (suggesting greater likelihood of harm than benefit), and with wide credible intervals).

[4] Very serious imprecision since RR is one (suggesting no evidence of benefit) and wide credible intervals suggesting high possibility of harm.

[5] Conceptually, there was no significant intransitivity, with comparable distribution of plausible effect modifiers across trials of different chemopreventive agents.

Buon lavoro e….

Tanti auguri ☺