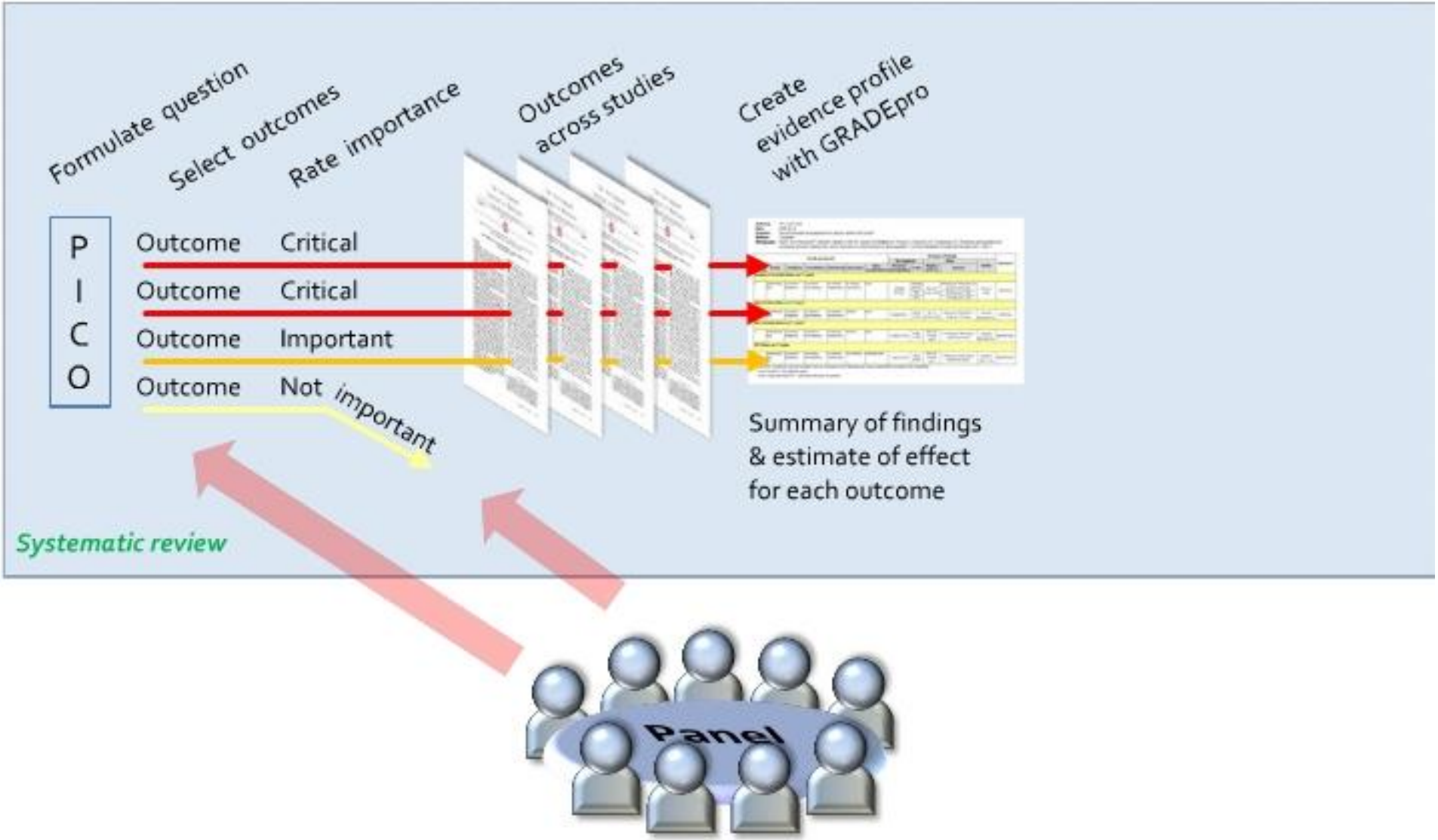
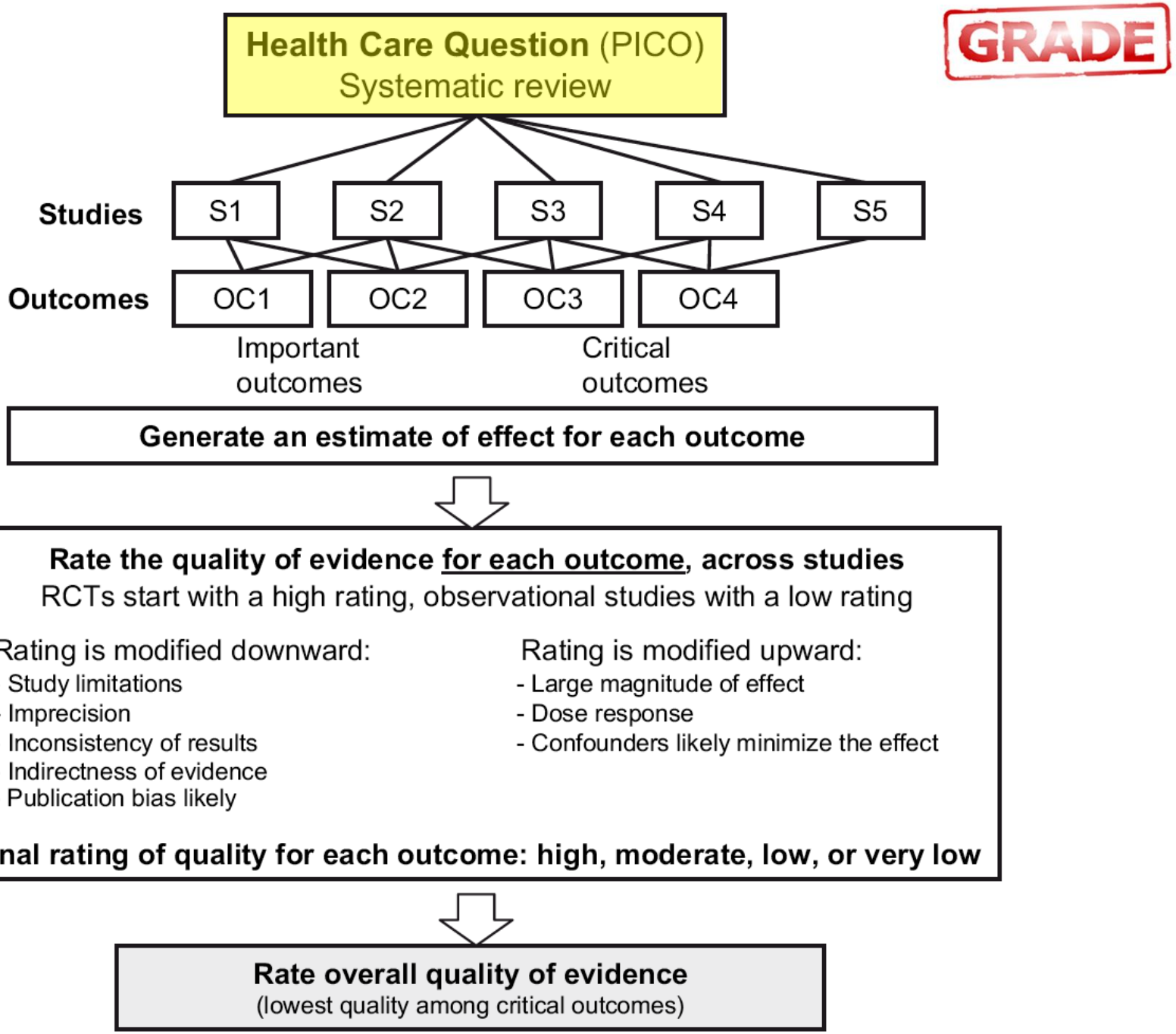
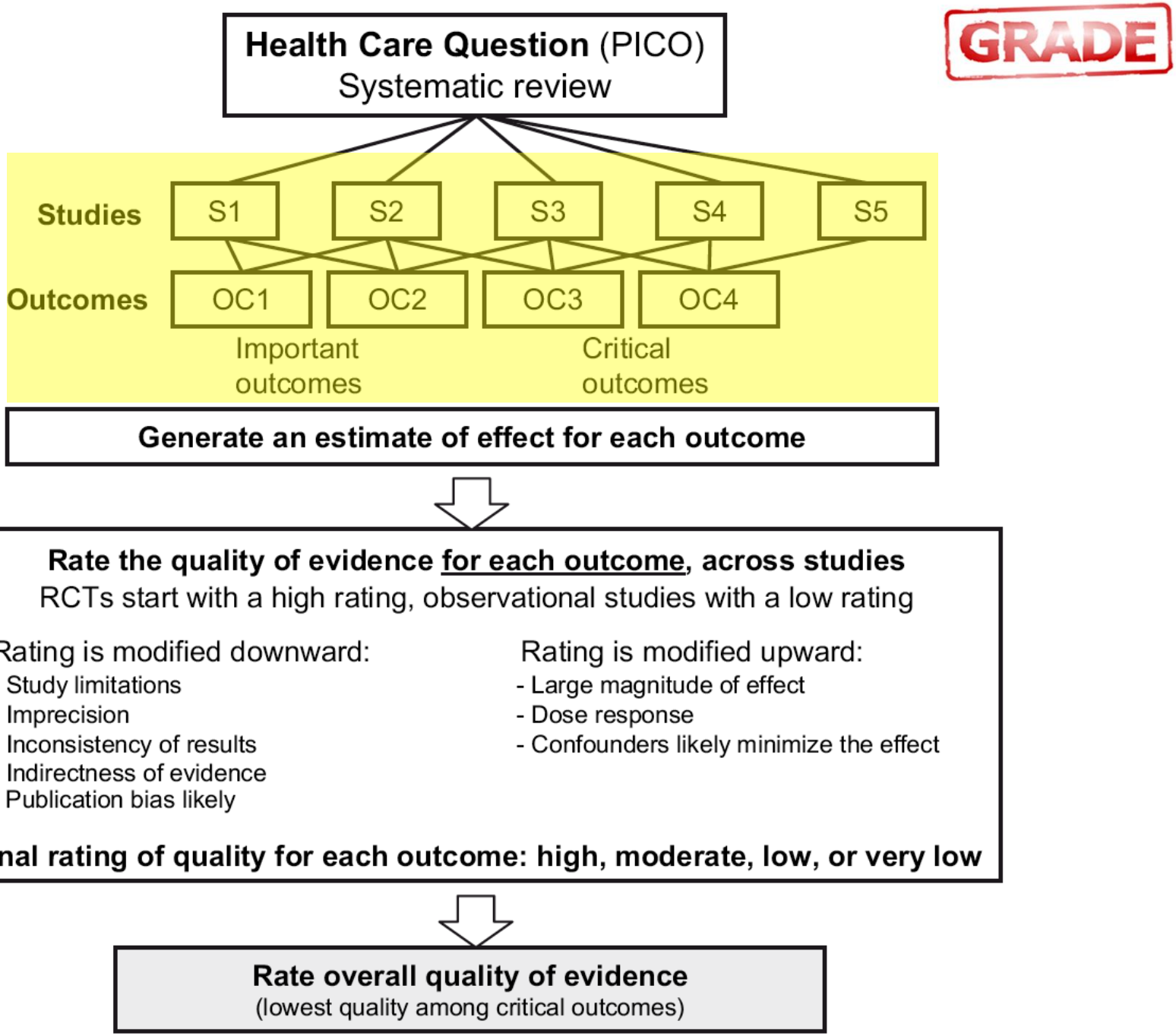
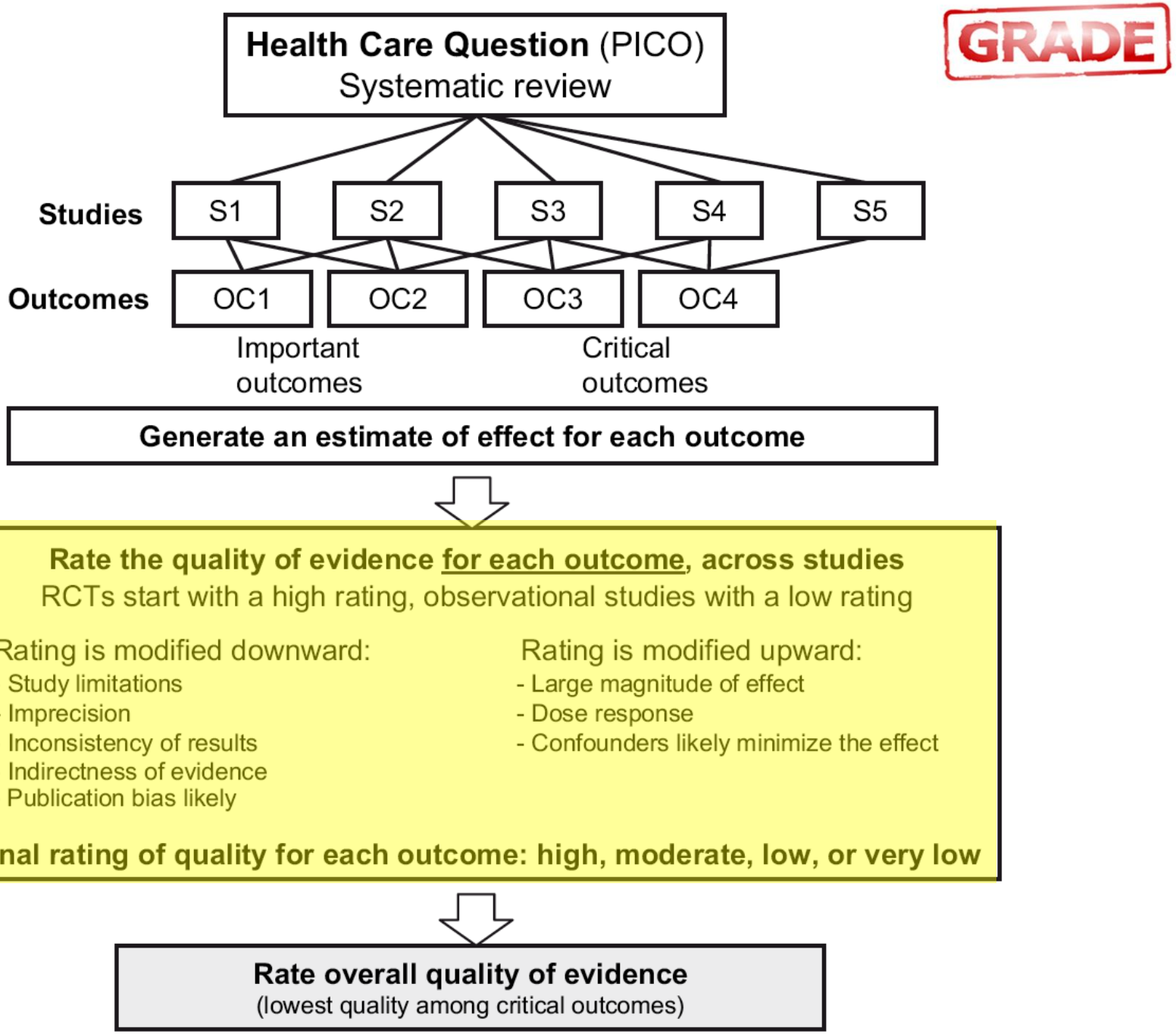

Formulate question
Select outcomes
Rate importance
Outcomes across studies
Create evidence profile with GRADEpro
P
I
C
O
Outcome Critical
Outcome Critical
Outcome Important
Outcome Not important
Summary of findings & estimate of effect for each outcome
Systematic review
Panel


Health Care Question (PICO)
Systematic review
GRADE
Studies
S1
S2
S3
S4
S5
Outcomes
OC1
OC2
OC3
OC4
Important outcomes
Critical outcomes
Generate an estimate of effect for each outcome
Rate the quality of evidence for each outcome, across studies
RCTs start with a high rating, observational studies with a low rating
Rating is modified downward:
- Study limitations
- Imprecision
- Inconsistency of results
- Indirectness of evidence
- Publication bias likely
Rating is modified upward:
- Large magnitude of effect
- Dose response
- Confounders likely minimize the effect
Final rating of quality for each outcome: high, moderate, low, or very low
Rate overall quality of evidence
(lowest quality among critical outcomes)


Health Care Question (PICO)
Systematic review
Studies
S1
S2
S3
S4
S5
Outcomes
OC1
OC2
OC3
OC4
Important outcomes
Critical outcomes
Generate an estimate of effect for each outcome
Rate the quality of evidence for each outcome, across studies
RCTs start with a high rating, observational studies with a low rating
Rating is modified downward:
- Study limitations
- Imprecision
- Inconsistency of results
- Indirectness of evidence
- Publication bias likely
Rating is modified upward:
- Large magnitude of effect
- Dose response
- Confounders likely minimize the effect
Final rating of quality for each outcome: high, moderate, low, or very low
Rate overall quality of evidence
(lowest quality among critical outcomes)


Health Care Question (PICO)
Systematic review
GRADE
Studies
S1
S2
S3
S4
S5
Outcomes
OC1
OC2
OC3
OC4
Important outcomes
Critical outcomes
Generate an estimate of effect for each outcome
Rate the quality of evidence for each outcome, across studies
RCTs start with a high rating, observational studies with a low rating
Rating is modified downward:
- Study limitations
- Imprecision
- Inconsistency of results
- Indirectness of evidence
- Publication bias likely
Rating is modified upward:
- Large magnitude of effect
- Dose response
- Confounders likely minimize the effect
Final rating of quality for each outcome: high, moderate, low, or very low
Rate overall quality of evidence
(lowest quality among critical outcomes)

GRADE Evidence syntheses

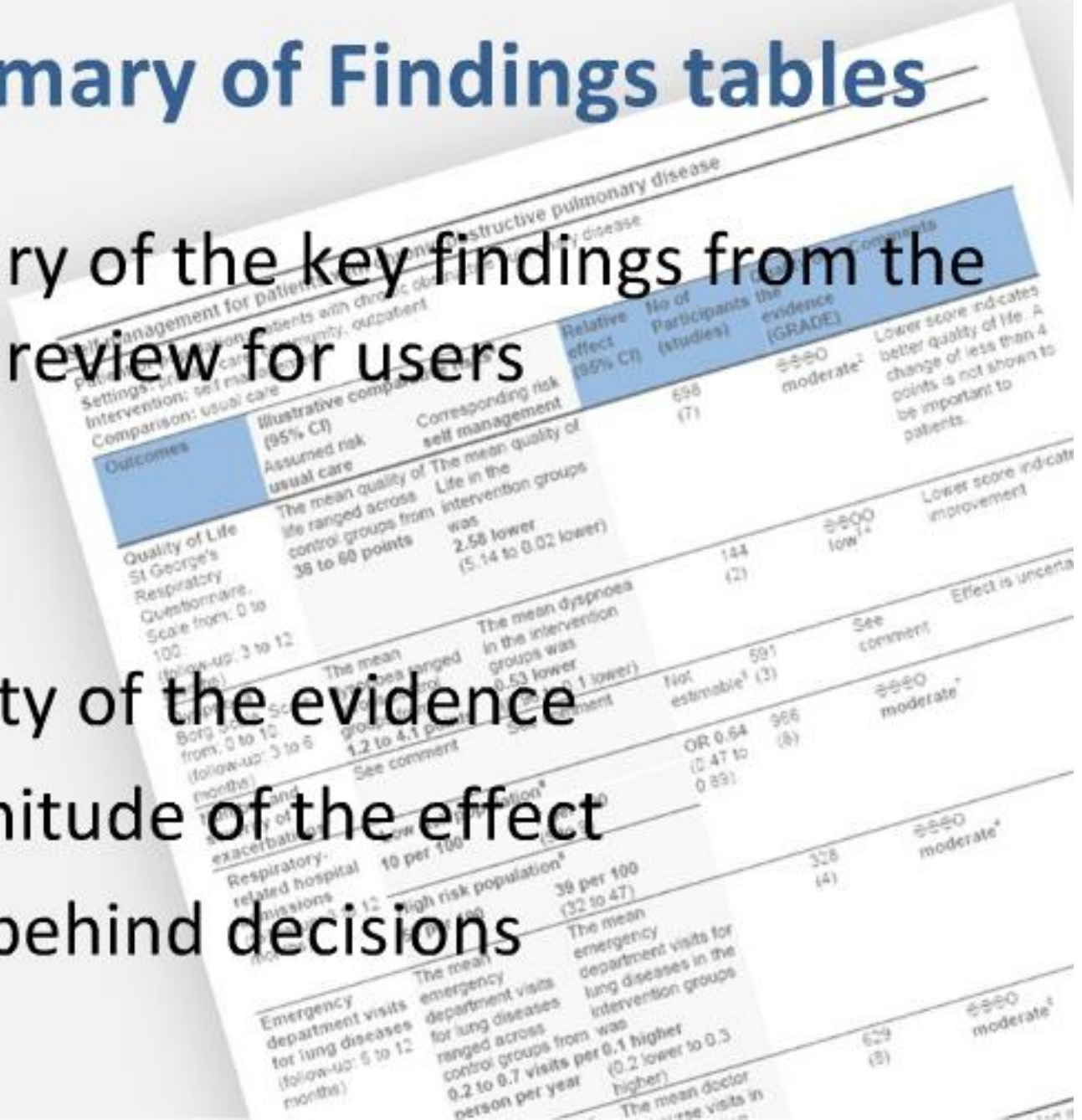
- Is a summary of the key results from the systematic review for guideline panel members
  - Evidence profiles and Summary of Findings Tables

- Presents
  - the quality of the evidence
  - the magnitude of the effect
  - transparent description of judgments about evidence

The Summary of Findings tables

- Is a summary of the key findings from the systematic review for users

- Presents
  - –the quality of the evidence
  - –the magnitude of the effect
  - –reasons behind decisions



Quando e perché fare le SoF

- Da aggiungere ad una SR che volete pubblicare per sintetizzare i risultati e la loro qualità (Summary of Findings)
- Nelle revisioni Cochrane è obbligatorio (Summary of Findings)
- Come strumento di lavoro /materiale di sintesi delle evidenze per la elaborazione di Linee Guida cliniche (Evidence Profile)

Linee guida
TUMORI DEL RENE
Edizione 2013

Author(s): GP MC
Date: 2013-09-30
Question: Should Sorafenib vs Placebo be used for mRCC dopo citochine?
Settings:
Bibliography: Escudier 2007 - NEJM 356:125-34 Escudier 2009 - JCO 20:3312-18

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Sorafenib	Placebo	Relative (95% CI)	Absolute		
Overall survival (follow-up median 6.6 months)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	278/452 (61.5%)	283/482 (58.7%)	HR 0.88 (0.74 to 1.04)	5 fewer per 100 (from 11 fewer to 1 more)	⊕⊕⊕O MODERATE	CRITICAL
Progression-Free Survival (follow-up median 6.6 months)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	187/452 (41.4%)	367/451 (81.4%)	HR 0.51 (0.43 to 0.6)	24 fewer per 100 (from 18 fewer to 30 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL
Quality of life - not measured												
0	-	-	-	-	-	none	-	-	-	-		IMPORTANT
Ipertensione G3/G4 (follow-up median 6.6 months; assessed with: CTC-AE)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	15/452 (3.3%)	0/451 (0%)	RR 30.93 (1.86 to 515.41)	-	⊕⊕⊕⊕ HIGH	CRITICAL
fatigue G3/G4 (follow-up median 6.6 months; assessed with: CTC-AE)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	14/452 (3.1%)	5/451 (1.1%)	RR 2.79 (1.01 to 7.69)	2 more per 100 (from 0 more to 7 more)	⊕⊕⊕⊕ HIGH	CRITICAL
Diarrea G3/G4 (follow-up median 6.6 months; assessed with: CTC-AE)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	14/452 (3.1%)	4/451 (0.89%)	RR 3.49 (1.16 to 10.53)	2 more per 100 (from 0 more to 8 more)	⊕⊕⊕⊕ HIGH	CRITICAL
HFSR G3/G4 (follow-up median 6.6 months; assessed with: CTC-AE)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	29/452 (6.4%)	2/451 (0.44%)	RR 14.47 (3.47 to 6.27)	6 more per 100 (from 1 more to 2 more)	⊕⊕⊕⊕ HIGH	CRITICAL

[1] Crossover da placebo a sorafenib al momento della progressione (48% dei pazienti)

Qualità per singolo outcome considerato

Linee guida
TUMORI DEL RENE
Edizione 2013

Author(s): GP MC
Date: 2013-09-30
Question: Should Sorafenib vs Placebo be used for mRCC dopo citochine?
Settings:
Bibliography: Escudier 2007 - NEJM 356:125-34 Escudier 2009 - JCO 20:3312-18

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Sorafenib	Placebo	Relative (95% CI)	Absolute		
Overall survival (follow-up median 6.6 months)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	278/452 (61.5%)	283/482 (58.7%)	HR 0.88 (0.74 to 1.04)	5 fewer per 100 (from 11 fewer to 1 more)	⊕⊕⊕O MODERATE	CRITICAL
Progression-Free Survival (follow-up median 6.6 months)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	187/452 (41.4%)	367/451 (81.4%)	HR 0.51 (0.43 to 0.6)	24 fewer per 100 (from 18 fewer to 30 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL
Quality of life - not measured												
0	-	-	-	-	-	none	-	-	-	-		IMPORTANT
Ipertensione G3/G4 (follow-up median 6.6 months; assessed with: CTC-AE)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	15/452 (3.3%)	0/451 (0%)	RR 30.93 (1.86 to 515.41)	-	⊕⊕⊕⊕ HIGH	CRITICAL
fatigue G3/G4 (follow-up median 6.6 months; assessed with: CTC-AE)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	14/452 (3.1%)	5/451 (1.1%)	RR 2.79 (1.01 to 7.69)	2 more per 100 (from 0 more to 7 more)	⊕⊕⊕⊕ HIGH	CRITICAL
Diarrea G3/G4 (follow-up median 6.6 months; assessed with: CTC-AE)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	14/452 (3.1%)	4/451 (0.89%)	RR 3.49 (1.16 to 10.53)	2 more per 100 (from 0 more to 8 more)	⊕⊕⊕⊕ HIGH	CRITICAL
HFSR G3/G4 (follow-up median 6.6 months; assessed with: CTC-AE)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	29/452 (6.4%)	2/451 (0.44%)	RR 14.47 (3.47 to 6.27)	6 more per 100 (from 1 more to 2 more)	⊕⊕⊕⊕ HIGH	CRITICAL

[1] Crossover da placebo a sorafenib al momento della progressione (48% dei pazienti)

Qualità globale delle evidenze valutate

Linee guida
TUMORI DEL RENE
Edizione 2013

Author(s): GP MC
Date: 2013-09-30
Question: Should Sorafenib vs Placebo be used for mRCC dopo citochine?
Settings:
Bibliography: Escudier 2007 - NEJM 356:125-34 Escudier 2009 - JCO 20:3312-18

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Sorafenib	Placebo	Relative (95% CI)	Absolute		
Overall survival (follow-up median 6.6 months)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	278/452 (61.5%)	283/482 (58.7%)	HR 0.88 (0.74 to 1.04)	5 fewer per 100 (from 11 fewer to 1 more)	⊕⊕⊕O MODERATE	CRITICAL
Progression-Free Survival (follow-up median 6.6 months)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	187/452 (41.4%)	367/451 (81.4%)	HR 0.51 (0.43 to 0.6)	24 fewer per 100 (from 18 fewer to 30 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL
Quality of life - not measured												
0	-	-	-	-	-	none	-	-	-	-		IMPORTANT
Ipertensione G3/G4 (follow-up median 6.6 months; assessed with: CTC-AE)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	15/452 (3.3%)	0/451 (0%)	RR 30.93 (1.86 to 515.41)	-	⊕⊕⊕⊕ HIGH	CRITICAL
fatigue G3/G4 (follow-up median 6.6 months; assessed with: CTC-AE)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	14/452 (3.1%)	5/451 (1.1%)	RR 2.79 (1.01 to 7.69)	2 more per 100 (from 0 more to 7 more)	⊕⊕⊕⊕ HIGH	CRITICAL
Diarrea G3/G4 (follow-up median 6.6 months; assessed with: CTC-AE)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	14/452 (3.1%)	4/451 (0.89%)	RR 3.49 (1.16 to 10.53)	2 more per 100 (from 0 more to 8 more)	⊕⊕⊕⊕ HIGH	CRITICAL
HFSR G3/G4 (follow-up median 6.6 months; assessed with: CTC-AE)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	29/452 (6.4%)	2/451 (0.44%)	RR 14.47 (3.47 to 6.27)	6 more per 100 (from 1 more to 2 more)	⊕⊕⊕⊕ HIGH	CRITICAL

[1] Crossover da placebo a sorafenib al momento della progressione (48% dei pazienti)

Rilevanza clinica degli effetti

GRADEpro GDT

HOME | GRADEpro GDT OVERVIEW | GUIDELINE RESOURCES | CALENDAR OF EVENTS | GRADE HANDBOOK | CONTACT SUPPORT | LOG IN

GRADE's software for Summary of Findings tables, Health Technology Assessment and Guidelines

LOG IN / SIGN UP

GRADE Quality of Evidence

In the context of making recommendations:

- The quality of evidence reflects the extent of our confidence that the estimates of an effect are adequate to support a particular decision or recommendation.

Quality of the body of evidence

Four levels

⊕⊕⊕⊕ High	We are very confident that the true effect lies close to that of the estimate of the effect
⊕⊕⊕◯ Moderate	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	Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
⊕◯◯◯ Very low	We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Determinants of quality

Study design

RCTs ⊕⊕⊕⊕
observational studies ⊕⊕◯◯

5 factors that can lower quality

1. limitations in detailed design and execution *(risk of bias criteria)*
2. Inconsistency *(or heterogeneity)*
3. Indirectness *(PICO and applicability)*
4. Imprecision *(number of events and confidence intervals)*
5. Publication bias

3 factors can increase quality

1. large magnitude of effect
2. all plausible residual confounding or biases may be working to reduce the demonstrated effect or increase the effect if no effect was observed
3. dose-response gradient

Determinants of quality

Study design

RCTs ⊕⊕⊕⊕
observational studies ⊕⊕◯◯

5 factors that can lower quality

1. limitations in detailed design and execution *(risk of bias criteria)*
2. Inconsistency *(or heterogeneity)*
3. Indirectness *(PICO and applicability)*
4. Imprecision *(number of events and confidence intervals)*
5. Publication bias

3 factors can increase quality

1. large magnitude of effect
2. all plausible residual confounding or biases may be working to reduce the demonstrated effect or increase the effect if no effect was observed
3. dose-response gradient

Determinants of quality

Study design

RCTs ⊕⊕⊕⊕
observational studies ⊕⊕◯◯

5 factors that can lower quality

1. limitations in detailed design and execution *(risk of bias criteria)*
2. Inconsistency *(or heterogeneity)*
3. Indirectness *(PICO and applicability)*
4. Imprecision *(number of events and confidence intervals)*
5. Publication bias

3 factors can increase quality

1. large magnitude of effect
2. all plausible residual confounding or biases may be working to reduce the demonstrated effect or increase the effect if no effect was observed
3. dose-response gradient

Tipo di prove	**Studio controllato e randomizzato = alta** **Studio osservazionale = bassa** **Qualsiasi altro tipo di informazione = molto basso**
B. Aumento della categoria di attribuzione (es. da "bassa" a "moderata")	1. Associazione intervento-*outcome* forte, ovvero con rischio relativo >2 (<0,5), sulla base di prove concordanti provenienti da due o più studi osservazionali, senza alcun fattore di confondimento plausibile (+1 livello) 2. Associazione intervento-*outcome* molto forte, ovvero con rischio relativo >5 (<0,2) (+2 livelli) 3. Presenza di un gradiente dose-risposta (+1 livello) 4. Tutti i possibili fattori di confondimento che avrebbero potuto alterare le stime di effetto avrebbero ridotto l'effetto che si osserva (+1 livello)

Mi posso fidare?

Determinants of quality

5 factors that can lower quality

1. limitations of detailed design and execution *(risk of bias criteria)*
2. Inconsistency *(or heterogeneity)*
3. Indirectness *(PICO and applicability)*
4. Imprecision *(number of events and confidence intervals)*
5. Publication bias

The members of the Grade Working Group

BMJ | 26 APRIL 2008 | VOLUME 336 924

SOURCES OF BIAS IN CLINICAL TRIALS

Type of bias	Description	
Selection bias.	Systematic differences between baseline characteristics of the groups that are compared.	• Sequence generation. • Allocation concealment.
Performance bias.	Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest.	• Blinding of participants and personnel. • Other potential threats to validity.
Detection bias.	Systematic differences between groups in how outcomes are determined.	• Blinding of outcome assessment. • Other potential threats to validity.
Attrition bias.	Systematic differences between groups in withdrawals from a study.	• Incomplete outcome data
Reporting bias.	Systematic differences between reported and unreported findings.	• Selective outcome reporting

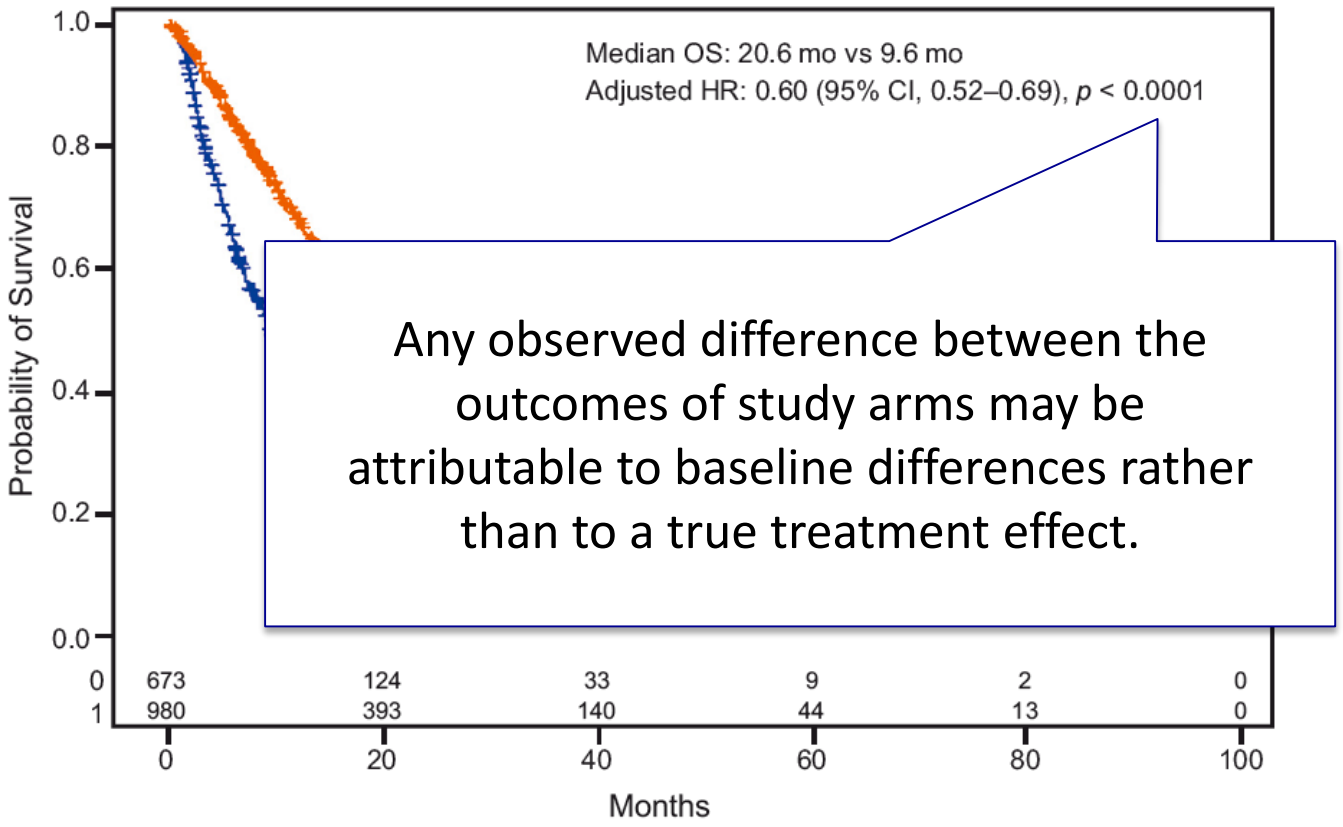
Cochrane Handbook for Systematic Reviews of Interventions

Version 5.1.0

Cytoreductive Nephrectomy in Patients with Synchronous Metastases from Renal Cell Carcinoma: Results from the International Metastatic Renal Cell Carcinoma Database Consortium

Daniel Y.C. Heng[a,,†], J. Connor Wells[a,†], Brian I. Rini[b], Benoit Beuselinck[c], Jae-Lyun Lee[d], Jennifer J. Knox[e], Georg A. Bjarnason[f], Sumanta Kumar Pal[g], Christian K. Kollmannsberger[h], Takeshi Yuasa[i], Sandy Srinivas[j], Frede Donskov[k], Aristotelis Bamias[l], Lori A. Wood[m], D. Scott Ernst[n], Neeraj Agarwal[o], Ulka N. Vaishampayan[p], Sun Young Rha[q], Jenny J. Kim[r], Toni K. Choueiri[s]*

EUROPEAN UROLOGY 66 (2014) 704–710



Any observed difference between the outcomes of study arms may be attributable to baseline differences rather than to a true treatment effect.

SOURCES OF BIAS IN CLINICAL TRIALS

Type of bias	Description	
Selection bias.	Systematic differences between baseline characteristics of the groups that are compared.	• Sequence generation. • Allocation concealment.
Performance bias.	Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest.	• Blinding of participants and personnel. • Other potential threats to validity.
Detection bias.	Systematic differences between groups in how outcomes are determined.	• Blinding of outcome assessment. • Other potential threats to validity.
Attrition bias.	Systematic differences between groups in withdrawals from a study.	• Incomplete outcome data
Reporting bias.	Systematic differences between reported and unreported findings.	• Selective outcome reporting

Cochrane Handbook for Systematic Reviews of Interventions

Version 5.1.0

If no patient blinding was performed…

… were they **unbiased** when filling the QoL questionnaire?

SOURCES OF BIAS IN CLINICAL TRIALS

Type of bias	Description	
Selection bias.	Systematic differences between baseline characteristics of the groups that are compared.	• Sequence generation. • Allocation concealment.
Performance bias.	Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest.	• Blinding of participants and personnel. • Other potential threats to validity.
Detection bias.	Systematic differences between groups in how outcomes are determined.	• Blinding of outcome assessment. • Other potential threats to validity.
Attrition bias.	Systematic differences between groups in withdrawals from a study.	• Incomplete outcome data
Reporting bias.	Systematic differences between reported and unreported findings.	• Selective outcome reporting

Cochrane Handbook for Systematic Reviews of Interventions

Version 5.1.0

If no evaluator blinding was performed…

… was he (totally) **unbiased** when evaluating the scan?

SOURCES OF BIAS IN CLINICAL TRIALS

Type of bias	Description	
Selection bias.	Systematic differences between baseline characteristics of the groups that are compared.	• Sequence generation. • Allocation concealment.
Performance bias.	Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest.	• Blinding of participants and personnel. • Other potential threats to validity.
Detection bias.	Systematic differences between groups in how outcomes are determined.	• Blinding of outcome assessment. • Other potential threats to validity.
Attrition bias.	Systematic differences between groups in withdrawals from a study.	• Incomplete outcome data
Reporting bias.	Systematic differences between reported and unreported findings.	• Selective outcome reporting

Cochrane Handbook for Systematic Reviews of Interventions

Version 5.1.0

Can trial quality be reliably assessed from published reports of cancer trials: evaluation of risk of bias assessments in systematic reviews

Claire L Vale *senior research scientist*, Jayne F Tierney *senior research scientist* , Sarah Burdett *senior research scientist*

BMJ 2013;346:f1798 doi: 10.1136/bmj.f1798 (Published 22 April 2013)

To evaluate attrition bias, on the basis of whether the outcome data were incomplete or not, the authors had to establish a rule of thumb to ensure consistency between assessments. Trials were assessed as low risk of bias if less than 10% of patients were excluded overall and if similar proportions were excluded from both arms. Trials were judged as high risk of bias if there were considerable imbalances between arms or if more than 10% of randomised patients were excluded from the analysis.

CONSORT

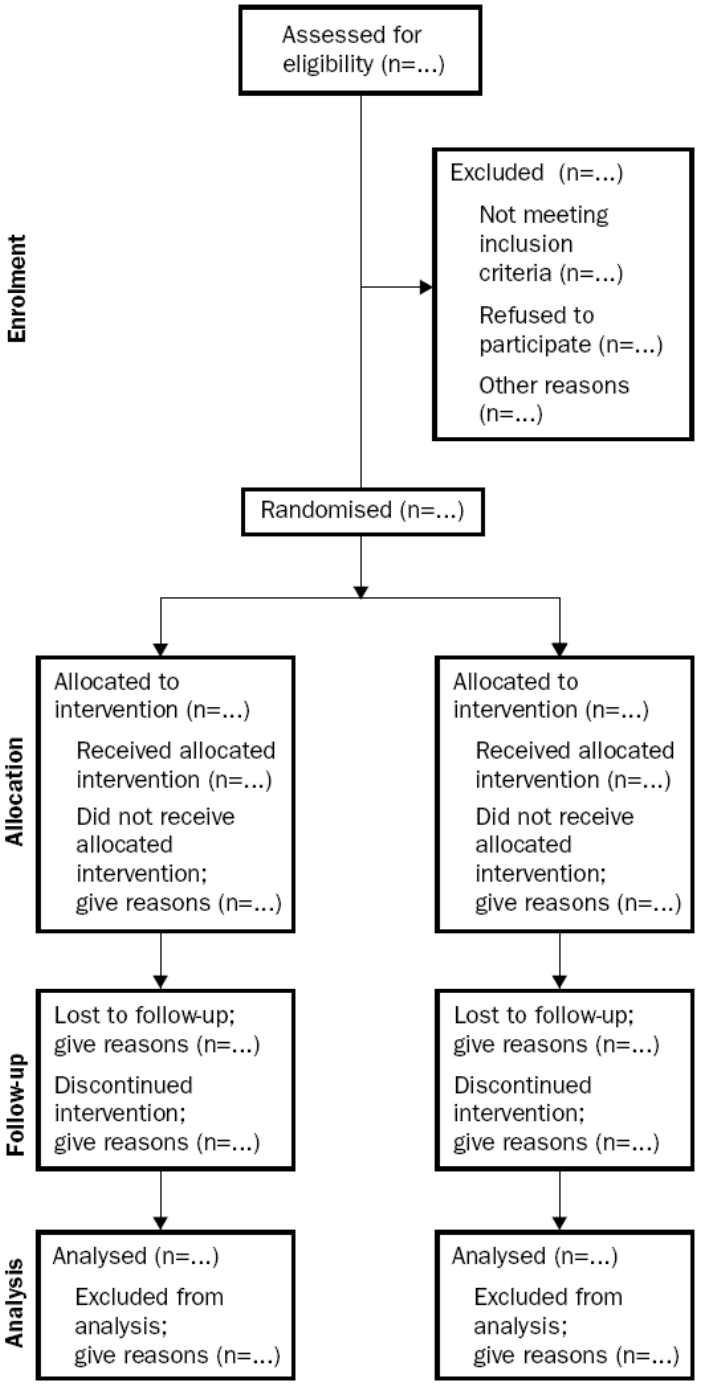
TRANSPARENT REPORTING of TRIALS

The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials

*David Moher, Kenneth F Schulz, Douglas G Altman, for the CONSORT Group**

Lancet 2001; **357:** 1191–94

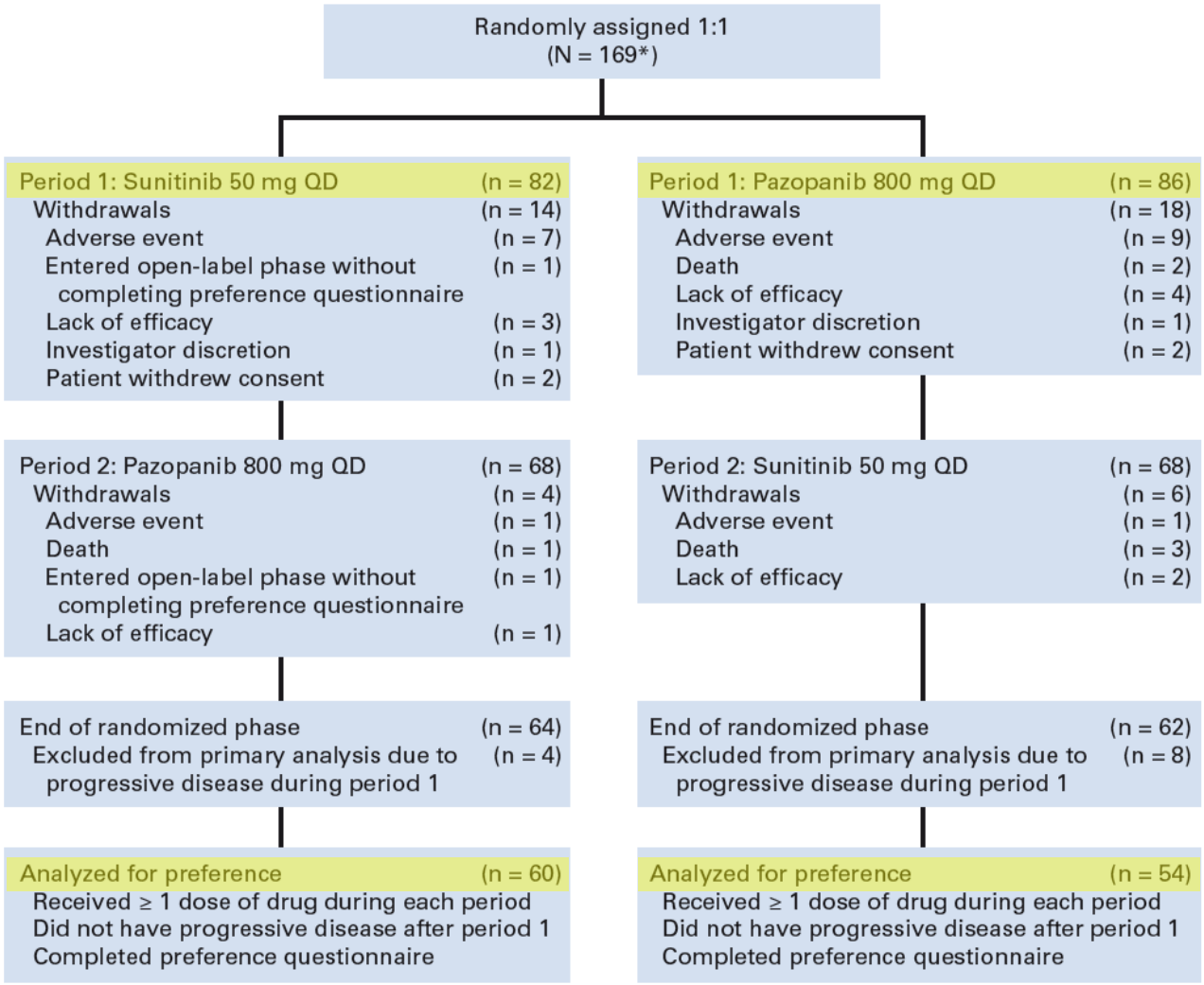
A ciascuno studio è richiesto di dare conto del flusso di pazienti nelle fasi di arruolamento, assegnazione del trattamento, follow-up e analisi



Randomized, Controlled, Double-Blind, Cross-Over Trial Assessing Treatment Preference for Pazopanib Versus Sunitinib in Patients With Metastatic Renal Cell Carcinoma: PISCES Study

Bernard Escudier, Camillo Porta, Petri Bono, Thomas Powles, Tim Eisen, Cora N. Sternberg, Jürgen E. Gschwend, Ugo De Giorgi, Omi Parikh, Robert Hawkins, Emmanuel Sevin, Sylvie Négrier, Sadya Khan, Jose Diaz, Suman Redhu, Faisal Mehmud, and David Cella

J Clin Oncol 32.



SOURCES OF BIAS IN CLINICAL TRIALS

Type of bias	Description	
Selection bias.	Systematic differences between baseline characteristics of the groups that are compared.	• Sequence generation. • Allocation concealment.
Performance bias.	Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest.	• Blinding of participants and personnel. • Other potential threats to validity.
Detection bias.	Systematic differences between groups in how outcomes are determined.	• Blinding of outcome assessment. • Other potential threats to validity.
Attrition bias.	Systematic differences between groups in withdrawals from a study.	• Incomplete outcome data
Reporting bias.	Systematic differences between reported and unreported findings.	• Selective outcome reporting

Cochrane Handbook for Systematic Reviews of Interventions

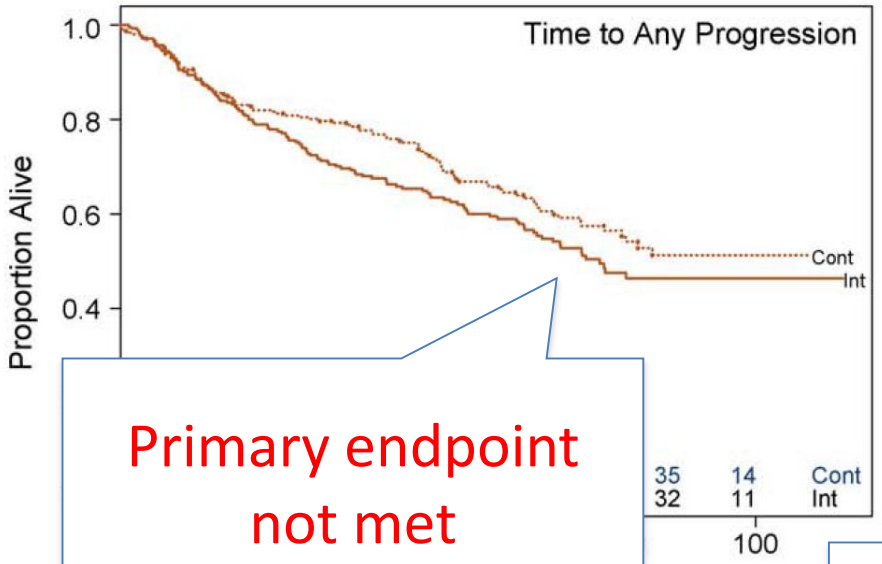
Version 5.1.0

Intermittent Androgen Deprivation for Locally Advanced and Metastatic Prostate Cancer: Results from a Randomised Phase 3 Study of the South European Uroncological Group

Fernando E.C. Calais da Silva [a,], Aldo V. Bono [b], Peter Whelan [c], Maurizio Brausi [d], Anton Marques Queimadelos [e], Jose A. Portillo Martin [f], Ziya Kirkali [g], Fernando M.V. Calais da Silva [h], Chris Robertson [i]*

EUROPEAN UROLOGY 55 (2009) 1269–1277

The study was designed to detect a 30% reduction in median time to progression (objective or subjective) in the intermittent arm compared with the continuous arm.



Primary endpoint not met

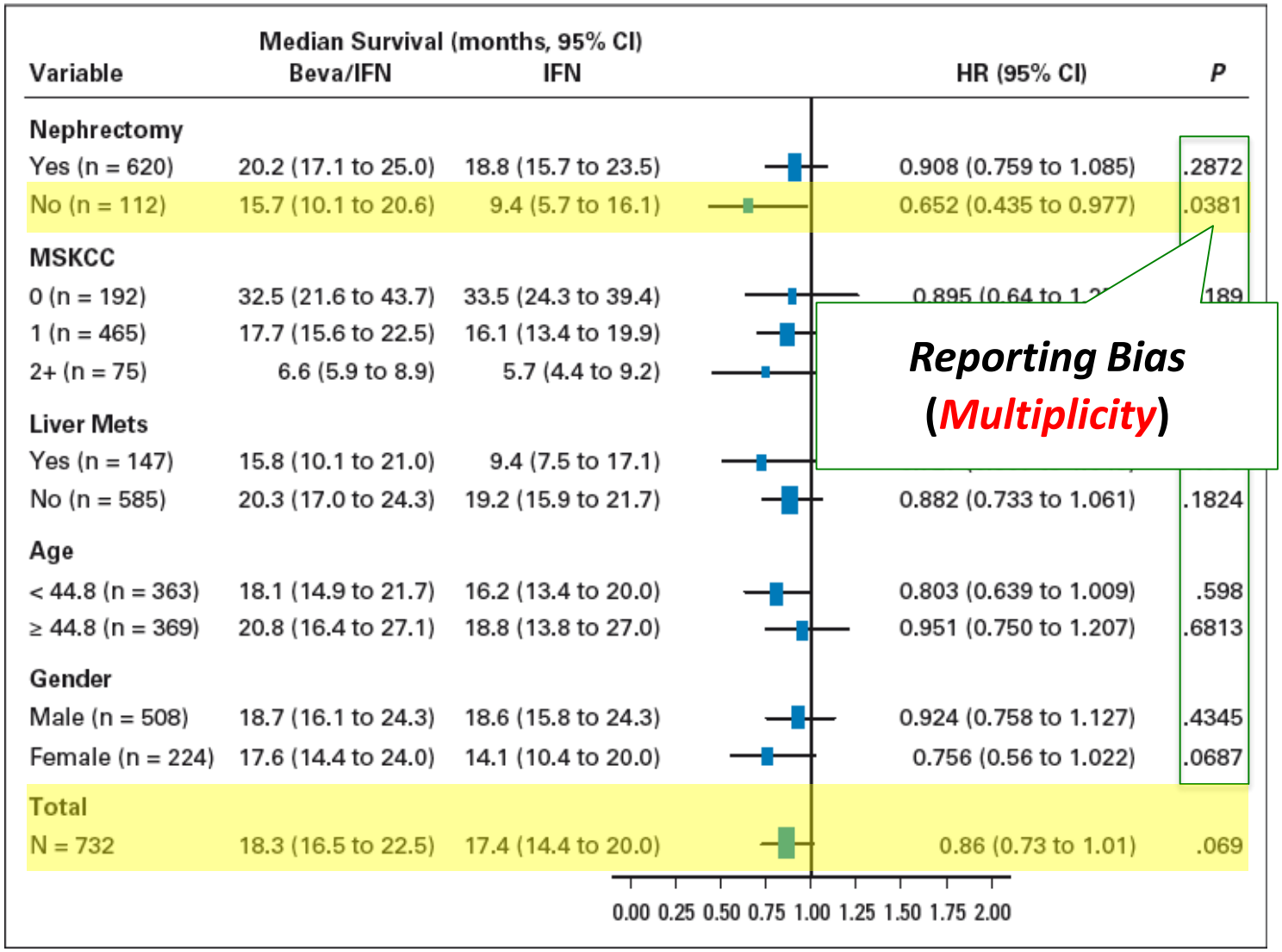
IHT should be considered for use in routine practice because it is associated with no reduction in survival, no clinically meaningful impairment in QoL, and better sexual activity.

Outcome reporting bias

Phase III Trial of Bevacizumab Plus Interferon Alfa Versus Interferon Alfa Monotherapy in Patients With Metastatic Renal Cell Carcinoma: Final Results of CALGB 90206

Brian I. Rini, Susan Halabi, Jonathan E. Rosenberg, Walter M. Stadler, Daniel A. Vaena, Laura Archer, James N. Atkins, Joel Picus, Piotr Czaykowski, Janice Dutcher, and Eric J. Small

J Clin Oncol 28:2137-2143. © 2010 by American Society of Clinical Oncology



Multiplicity

Multiple comparisons, **multiplicity** or multiple testing problem occurs when one considers a set of statistical inferences simultaneously or infers a subset of parameters selected based on the observed values

Probability of at least one false significant result

Number of tests	Probability
1	0.050
2	0.098
5	0.226
10	0.401
50	0.923

Several statistical techniques have been developed to prevent this from happening, allowing significance levels for single and multiple comparisons to be directly compared. These techniques generally require a higher significance threshold for individual comparisons, so as to compensate for the number of inferences being made.

Multiplicity is everywhere…

- In **subgroup** analyses (also when pre-specified)
- In **multiple endpoints** (this is why the primary endpoint must be pre-specified)
- In **interim** analyses
- In **reanalysis** of the same study (spanish intermittent ADT study)
- In **model building** (prognostic models, gene-signatures …)

Lacosamide (LCM) compared to placebo for partial-onset seizures

Participants (studies)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							Risk with placebo	Risk with Lacosamide (LCM)		Risk with placebo	Risk difference with Lacosamide (LCM)
Nausea (assessed with: all dosage arms pooled)											
1105 (3 RCTs)	not serious	not serious	serious [c]	not serious	none	⨁⨁⨁◯ MODERATE	16/364 (4.4%)	73/741 (9.9%)	**RR 2.20** (1.05 to 4.60)	4 per 100	**5 more per 100** (from 0 fewer to 16 more)
Nausea (assessed with: LCM at 200mg)											
530 (2 RCTs)	serious [a]	not serious	not serious	serious [d]	none	⨁⨁◯◯ LOW	11/260 (4.2%)	20/270 (7.4%)	**RR 1.93** (0.49 to 7.56)	4 per 100	**4 more per 100** (from 2 fewer to 28 more)
Nausea (assessed with: LCM at 400mg)											
835	serious [a]	not serious	not serious	not serious	none	⨁⨁⨁◯	16/364 (4.4%)	53/471	**RR 2.43**	4 per 100	**6 more per 100**

Filter by active cell

Explanations | References

a. unplanned subgroup analysis

b. not downgraded for imprecision because the low number of events

c. all dosage arms pooled

d. 95% CIs consistent with conflicting recommendations

Multiplicity

Multiple comparisons, **multiplicity** or multiple testing problem occurs when one considers a set of statistical inferences simultaneously or infers a subset of parameters selected based on the observed values

Probability of at least one false significant result

Number of tests	Probability
1	0.050
2	0.098
5	0.226
10	0.401
50	0.923

Several statistical techniques have been developed to prevent this from happening, allowing significance levels for single and multiple comparisons to be directly compared. These techniques generally require a higher significance threshold for individual comparisons, so as to compensate for the number of inferences being made.

Multiplicity in randomised trials II: subgroup and interim analyses

Kenneth F Schulz, David A Grimes

Lancet 2005; 365: 1657–61

Number of planned interim analyses	Interim analysis	Pocock	Peto	O'Brien-Fleming
2	1	0·029	0·001	0·005
	2 (final)	0·029	0·05	0·048
3	1	0·022	0·001	0·0005
	2	0·022	0·001	0·014
	3 (final)	0·022	0·05	0·045
4	1	0·018	0·001	0·0001
	2	0·018	0·001	0·004
	3	0·018	0·001	0·019
	4 (final)	0·018	0·05	0·043
5	1	0·016	0·001	0·00001
	2	0·016	0·001	0·0013
	3	0·016	0·001	0·008
	4	0·016	0·001	0·023
	5 (final)	0·016	0·05	0·041

Overall α=0·05.

Table 2: Interim stopping levels (p values) for different numbers of planned interim analyses by group sequential design[14,15]

The European Agency for the Evaluation of Medicinal Products
Evaluation of Medicines for Human Use

London, 19 September 2002
CPMP/EWP/908/99

COMMITTEE FOR PROPRIETARY MEDICINAL PRODUCTS (CPMP)

POINTS TO CONSIDER ON MULTIPLICITY ISSUES IN CLINICAL TRIALS

2. ADJUSTMENT FOR MULTIPLICITY – WHEN IS IT NECESSARY AND WHEN IS IT NOT?

Sometimes a series of related objectives is pursued in the same trial, where one objective is of greatest importance but convincing results in others would clearly add to the value of the treatment.

In these situations, there is no intention or opportunity to select the most favourable result and, consequently, the individual type I error levels are set equal to the overall type I error level α, i.e. no reduction is necessary.

In such cases the hypotheses may be tested (and confidence intervals may be provided) according to a hierarchical strategy. The hierarchical order may be a natural one (e.g. hypotheses are ordered in time or with respect to the seriousness of the considered variables) or may result from the particular interests of the investigator. Again, no reduction or splitting of α is necessary. The hierarchical order for testing null hypotheses, however, has to be pre-specified in the study protocol.

Effect of a monoclonal antibody to PCSK9, REGN727/SAR236553, to reduce low-density lipoprotein cholesterol in patients with heterozygous familial hypercholesterolaemia on stable statin dose with or without ezetimibe therapy: a phase 2 randomised controlled trial

Evan A Stein, Dan Gipe, Jean Bergeron, Daniel Gaudet, Robert Weiss, Robert Dufour, Richard Wu, Robert Pordy
Lancet 2012; 380: 29–36

To address the multiple comparisons of the four treatment groups compared with placebo for the primary efficacy endpoint analysis, we applied a hierarchical testing procedure to ensure strong control of the overall type-1 error rate at the two-sided 0·05 significance level.

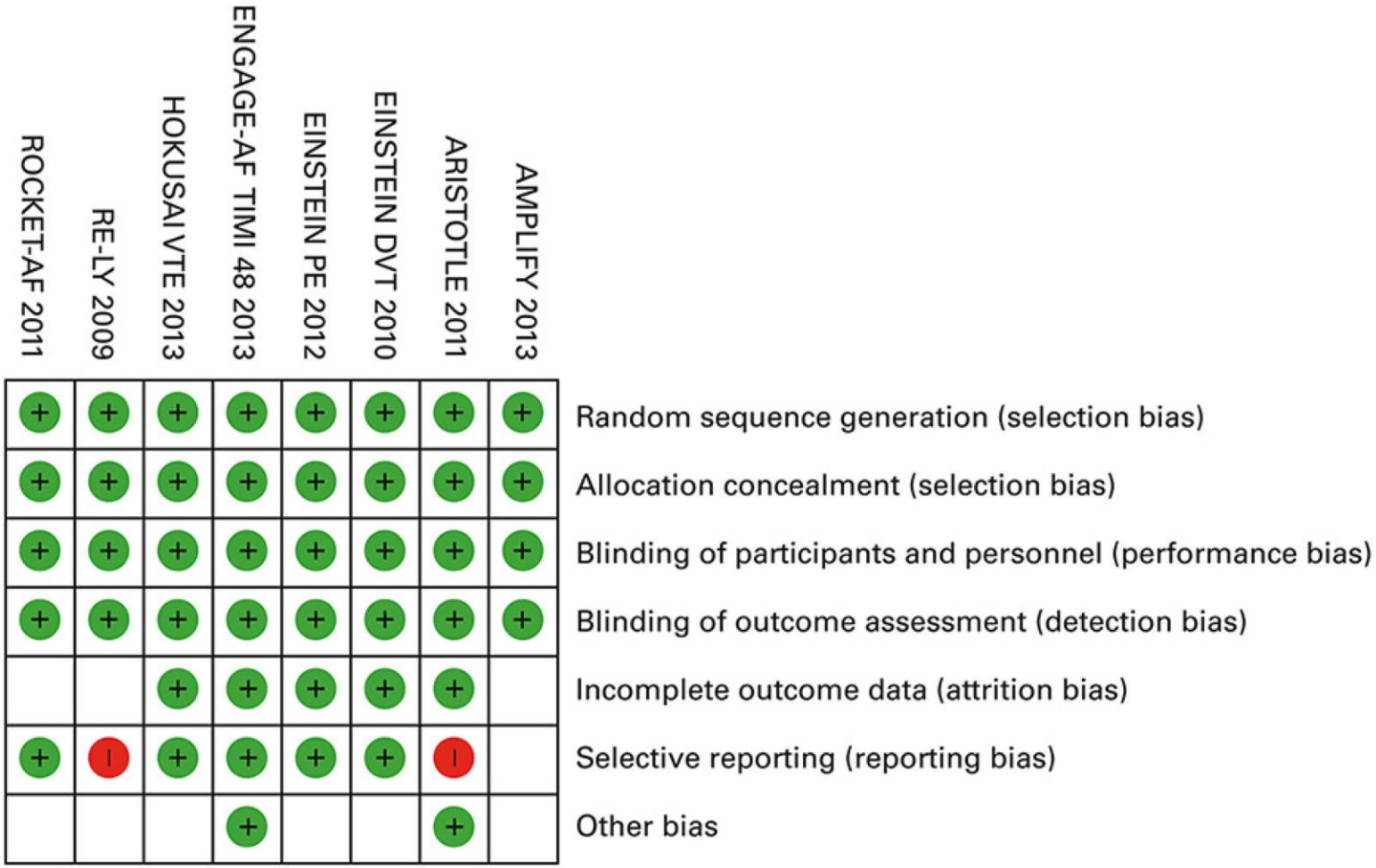
The order used was REGN727 150 mg every 2 weeks versus placebo first; REGN727 300 mg every 4 weeks versus placebo second; REGN727 200 mg every 4 weeks versus placebo third; and finally, REGN727 150 mg every 4 weeks versus placebo.

The hierarchical testing sequence continued only when the higher-order test was statistically significant at the two-sided 5% significant level.

Direct oral anticoagulants in the elderly: systematic review and meta-analysis of evidence, current and future directions

Angélique H. Sadlon, Dimitrios A. Tsakiris

Swiss Med Wkly. 2016;146:w14356



Mi posso fidare?

Determinants of quality

5 factors that can lower quality

1. limitations of detailed design and execution *(risk of bias criteria)*
2. Inconsistency *(or heterogeneity)*
3. Indirectness *(PICO and applicability)*
4. Imprecision *(number of events and confidence intervals)*
5. Publication bias

The members of the Grade Working Group

BMJ | 26 APRIL 2008 | VOLUME 336 924

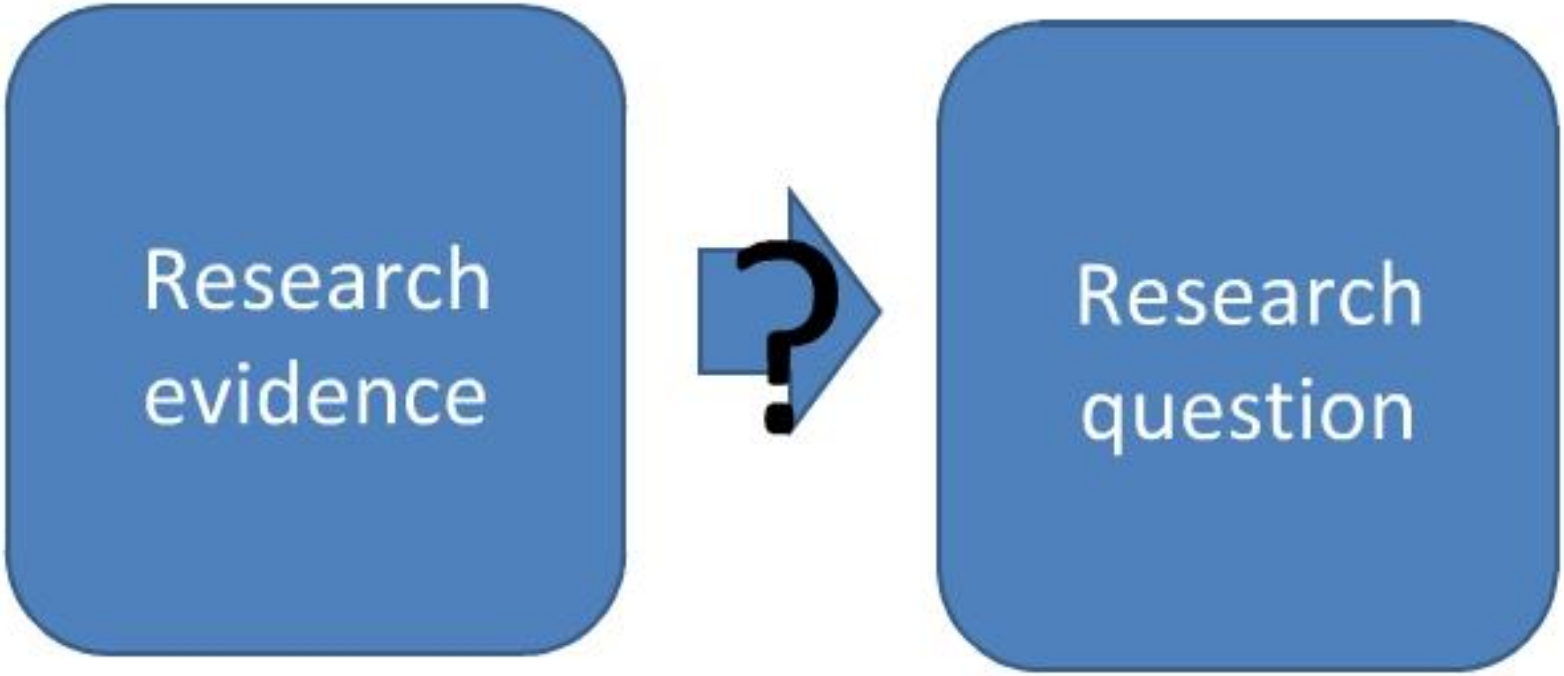
Direct evidence…

…comes from research that:

- is conducted in the **Population** that we are providing answers for;
- includes the **Intervention** that we are interested in…
- …and compares these interventions with the appropriate **Alternatives**;
- measures the **Outcomes** in which we are interested

Directness of Evidence

generalizability, transferability, applicability, external validity



Directness of Evidence

generalizability, transferability, applicability, external validity

- differences in
  - populations/patients (high income countries – low/middle income countries, patients with HIV – all patients)
  - interventions (new antibiotics in a class - old)
  - comparator appropriate (old antibiotics, no or other class)
  - outcomes (important – surrogate; signs and symptoms – mortality)

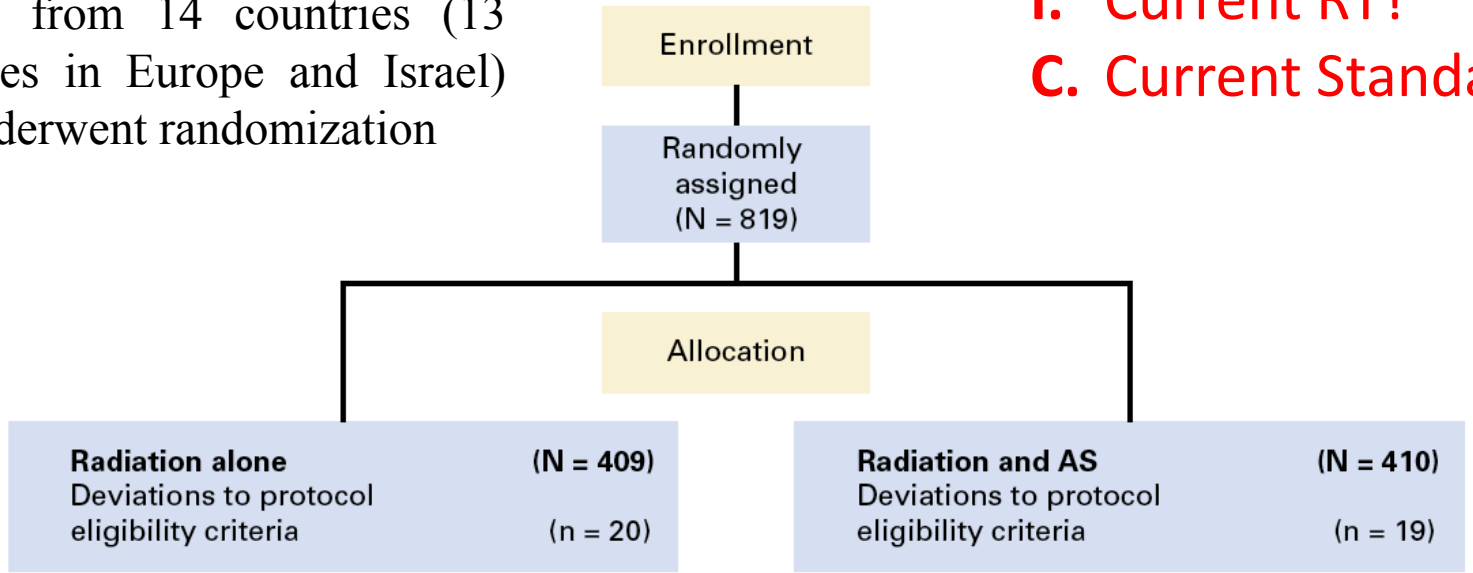
Short Androgen Suppression and Radiation Dose Escalation for Intermediate- and High-Risk Localized Prostate Cancer: Results of EORTC Trial 22991

Michel Bolla, Philippe Maingon, Christian Carrie, Salvador Villa, Petros Kitsios, Philip M.P. Poortmans, Santhanam Sundar, Elzbieta M. van der Steen-Banasik, John Armstrong, Jean-François Bosset, Fernanda G. Herrera, Bradley Pieters, Annerie Slot, Amit Bahl, Rahamim Ben-Yosef, Dirk Boehmer, Christopher Scrase, Laurette Renard, Emad Shash, Corneel Coens, Alphonsus C.M. van den Bergh, and Laurence Collette

J Clin Oncol 34:1748-1756. © 2016 by American Society of Clinical Oncology

Between September 21, 2001 and April 24, 2008, a total of 819 patients were recruited by 37 centers from 14 countries (13 countries in Europe and Israel) and underwent randomization

P. Today's patients?
I. Current RT?
C. Current Standard?



Lacosamide (LCM) compared to placebo for partial-onset seizures

Participants (studies)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							Risk with placebo	Risk with Lacosamide (LCM)		Risk with placebo	Risk difference with Lacosamide (LCM)
Seizure-free (assessed with: monitoring during the treatment period)											
1105 (3 RCTs)	not serious	not serious	not serious	not serious [b]	none	⨁⨁⨁⨁ HIGH	3/364 (0.8%)	18/741 (2.4%)	**RR 2.01** (0.66 to 6.05)	1 per 100	**1 more per 100** (from 0 fewer to 4 more)
Discontinuation due to AEs (assessed with: all dosage arms pooled)											
1105 (3 RCTs)	not serious	not serious	serious [c]	not serious	none	⨁⨁⨁◯ MODERATE	18/364 (4.9%)	102/741 (13.8%)	**RR 2.73** (1.68 to 4.44)	5 per 100	**9 more per 100** (from 3 more to 17 more)
Discontinuation due to AEs (assessed with: LCM at 200mg)											

Filter by active cell

Explanations | References

a. unplanned subgroup analysis

b. not downgraded for imprecision because the low number of events

c. all dosage arms pooled

d. 95% CIs consistent with conflicting recommendations

Pertuzumab plus Trastuzumab plus Docetaxel for Metastatic Breast Cancer

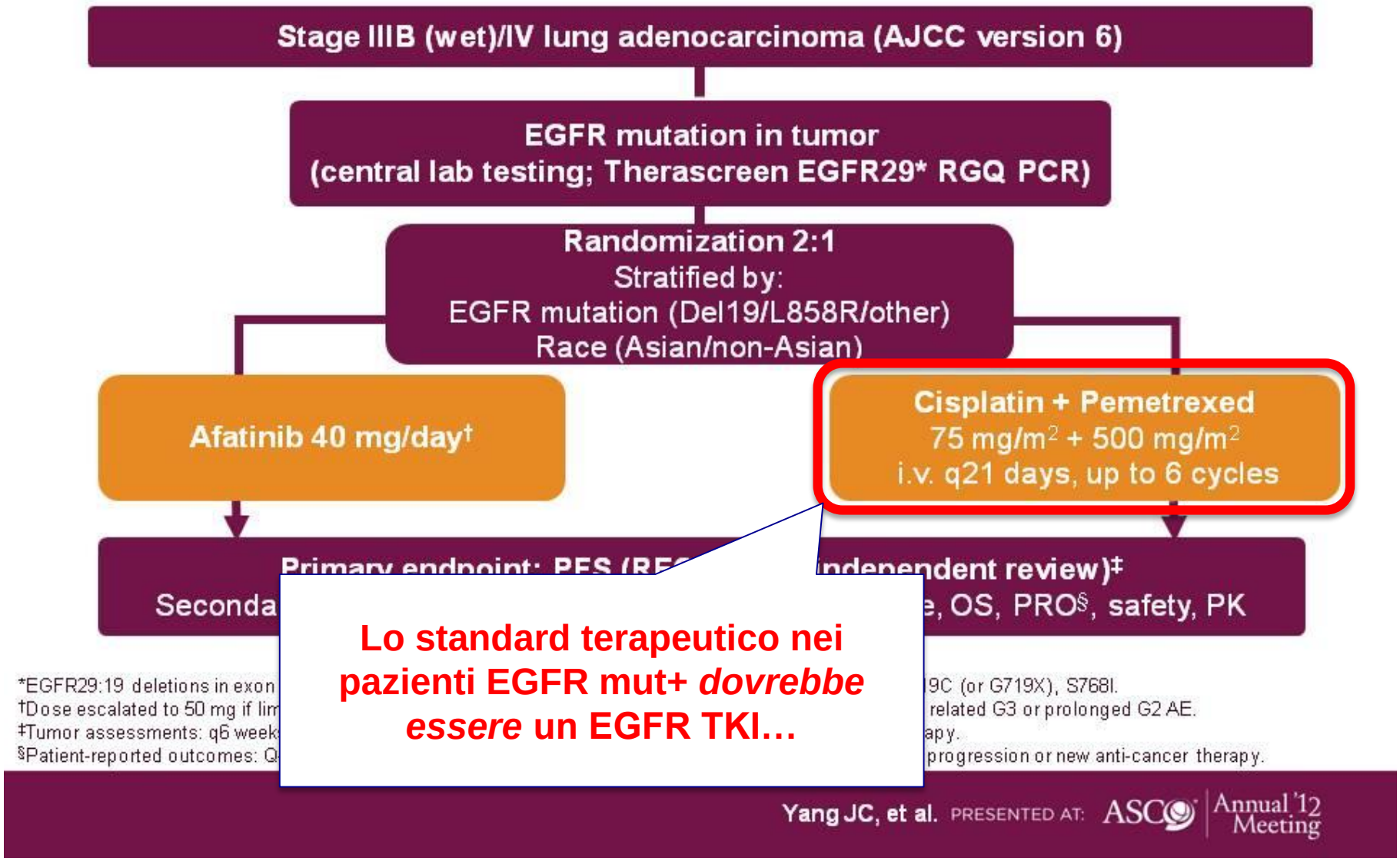
José Baselga, M.D., Ph.D., Javier Cortés, M.D., Sung-Bae Kim, M.D., Seock-Ah Im, M.D., Roberto Hegg, M.D., Young-Hyuck Im, M.D., Laslo Roman, M.D., José Luiz Pedrini, M.D., Tadeusz Pienkowski, M.D., Adam Knott, Ph.D., Emma Clark, M.Sc., Mark C. Benyunes, M.D., Graham Ross, F.F.P.M., and Sandra M. Swain, M.D., for the CLEOPATRA Study Group*

N Engl J Med 2012;366:109-19.

Characteristic	Placebo plus Trastuzumab plus Docetaxel (N=406)	Pertuzumab plus Trastuzumab plus Docetaxel (N=402)
Prior adjuvant or neoadjuvant chemotherapy — no. (%)		
No	214 (52.7)	218 (54.2)
Yes§	192 (47.3)	184 (45.8)
Anthracycline	164 (40.4)	150 (37.3)
Hormone	97 (23.9)	106 (26.4)
Taxane	94 (23.2)	91 (22.6)
Trastuzumab	41 (10.1)	47 (11.7)

Non rappresentativo della pratica clinica corrente

LUX-Lung 3 Study Design



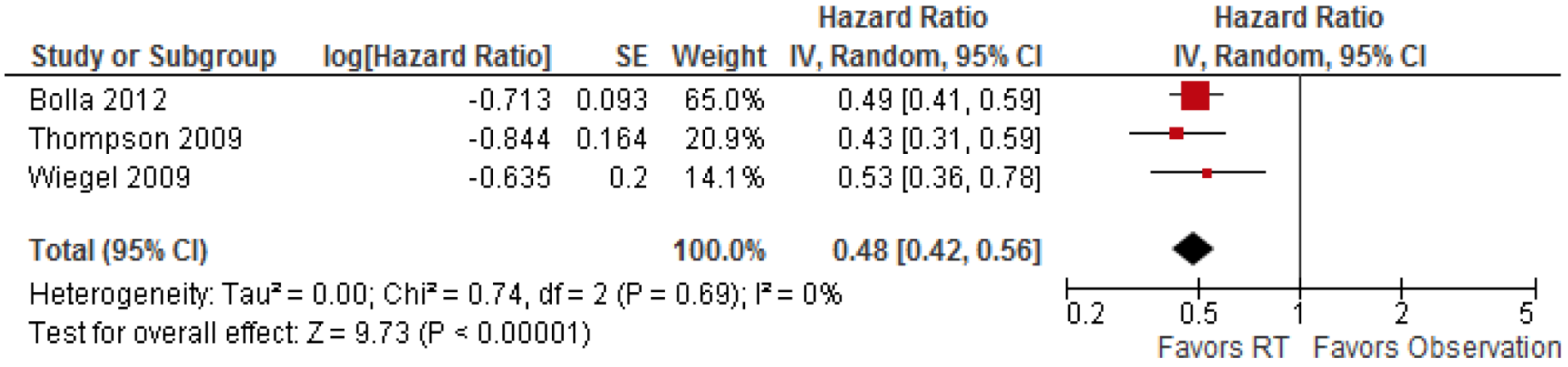
Lo standard terapeutico nei pazienti EGFR mut+ *dovrebbe essere* un EGFR TKI…

Presented By James Chih-Hsin Yang, MD, PhD at 2012 Annual Meeting

American Urological Association (AUA) Guideline

ADJUVANT AND SALVAGE RADIOTHERAPY AFTER PROSTATECTOMY: ASTRO/AUA GUIDELINE

Ian Murchie Thompson,* Richard Valicenti,* Peter C. Albertsen, Brian Davis, S. Larry Goldenberg, Carol A. Hahn, Eric A. Klein, Jeff Michalski, Mack Roach III, Oliver Sartor, J. Stuart Wolf Jr. and Martha M. Faraday



Meta-analysis of biochemical recurrence data from SWOG 879427, EORTC 2291125, and ARO 96-0226

EUROPEAN JOURNAL OF CANCER 42 (2006) 1344–1350

Prostate-specific antigen (PSA) alone is not an appropriate surrogate marker of long-term therapeutic benefit in prostate cancer trials

Laurence Collette[a,*], *Tomasz Burzykowski*[b], *Fritz H. Schröder*[c]

We review the published literature pertaining to the validation of PSA endpoints as surrogate in all disease stages.

We discuss the limitations of these studies and conclude that so far, PSA is not a validated surrogate endpoint in any of the disease settings and treatment conditions considered.

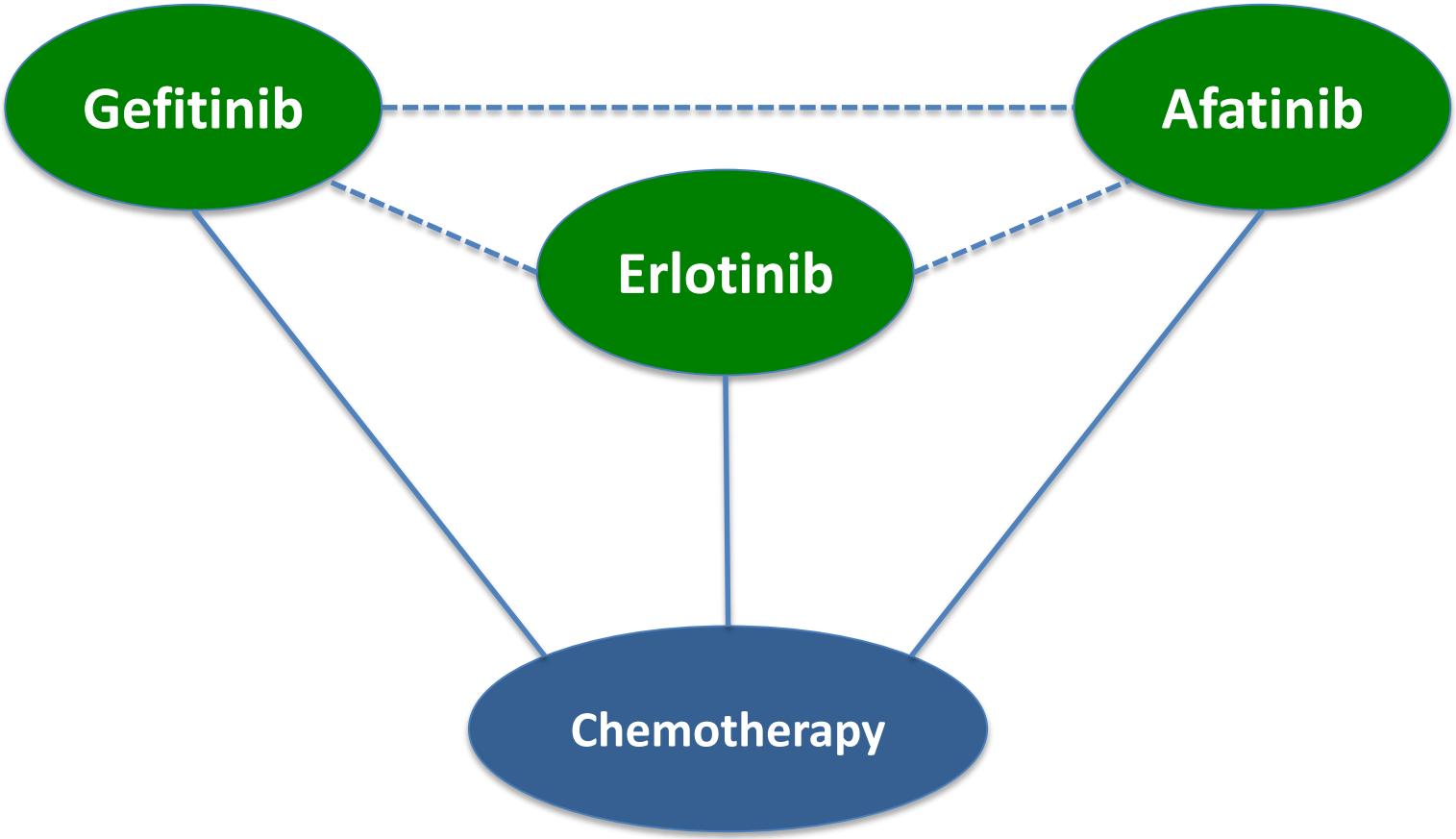
Directness of Evidence

generalizability, transferability, applicability, external validity

- differences in
  - populations/patients (high income countries – low/middle income countries, patients with HIV – all patients)
  - interventions (new antibiotics in a class - old)
  - comparator appropriate (old antibiotics, no or other class)
  - outcomes (important – surrogate; signs and symptoms – mortality)
- indirect comparisons
  - interested in A versus B,

    but have A versus control and B versus control

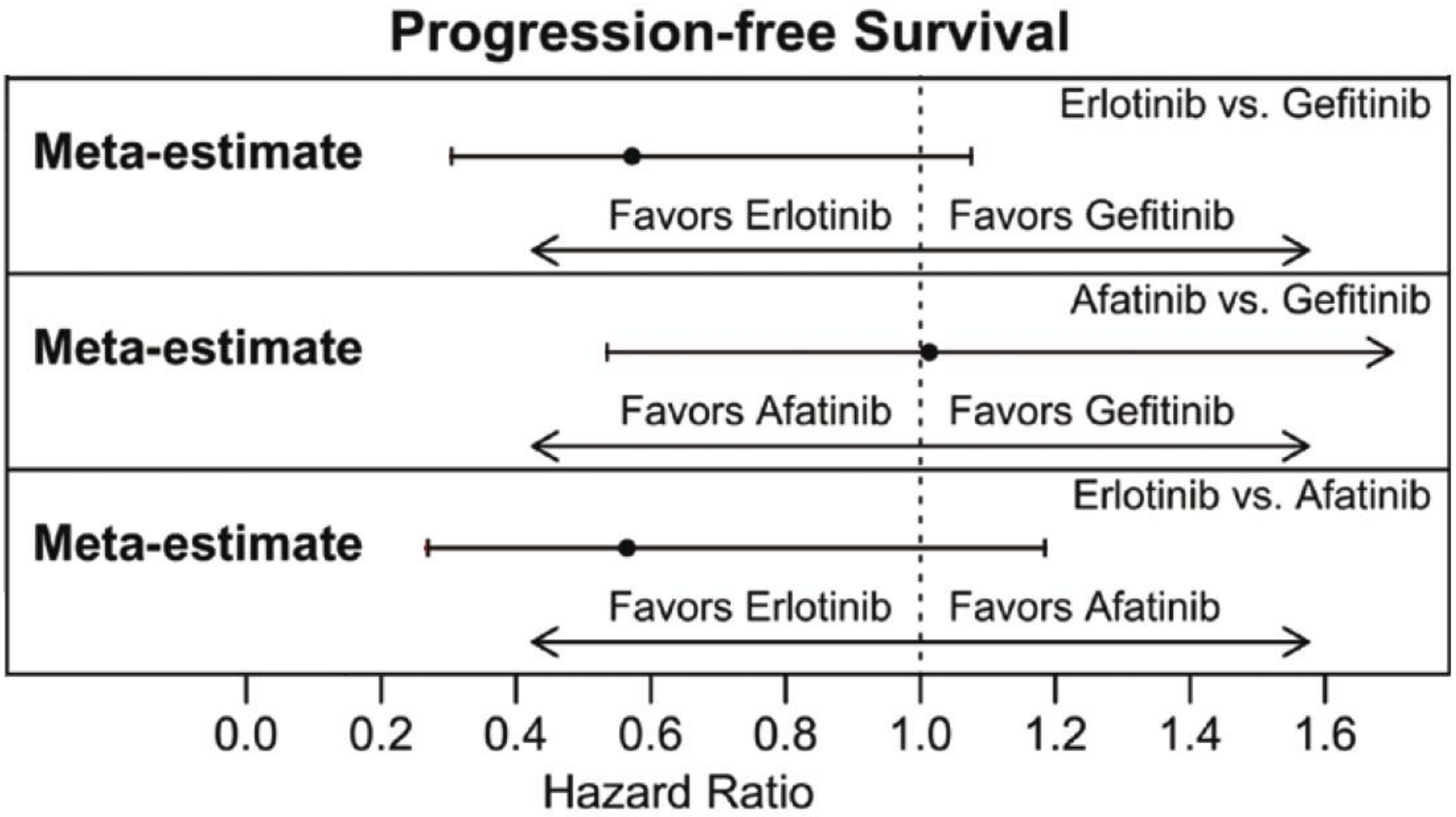
Unmet medical need
ma assenza di confronti diretti…

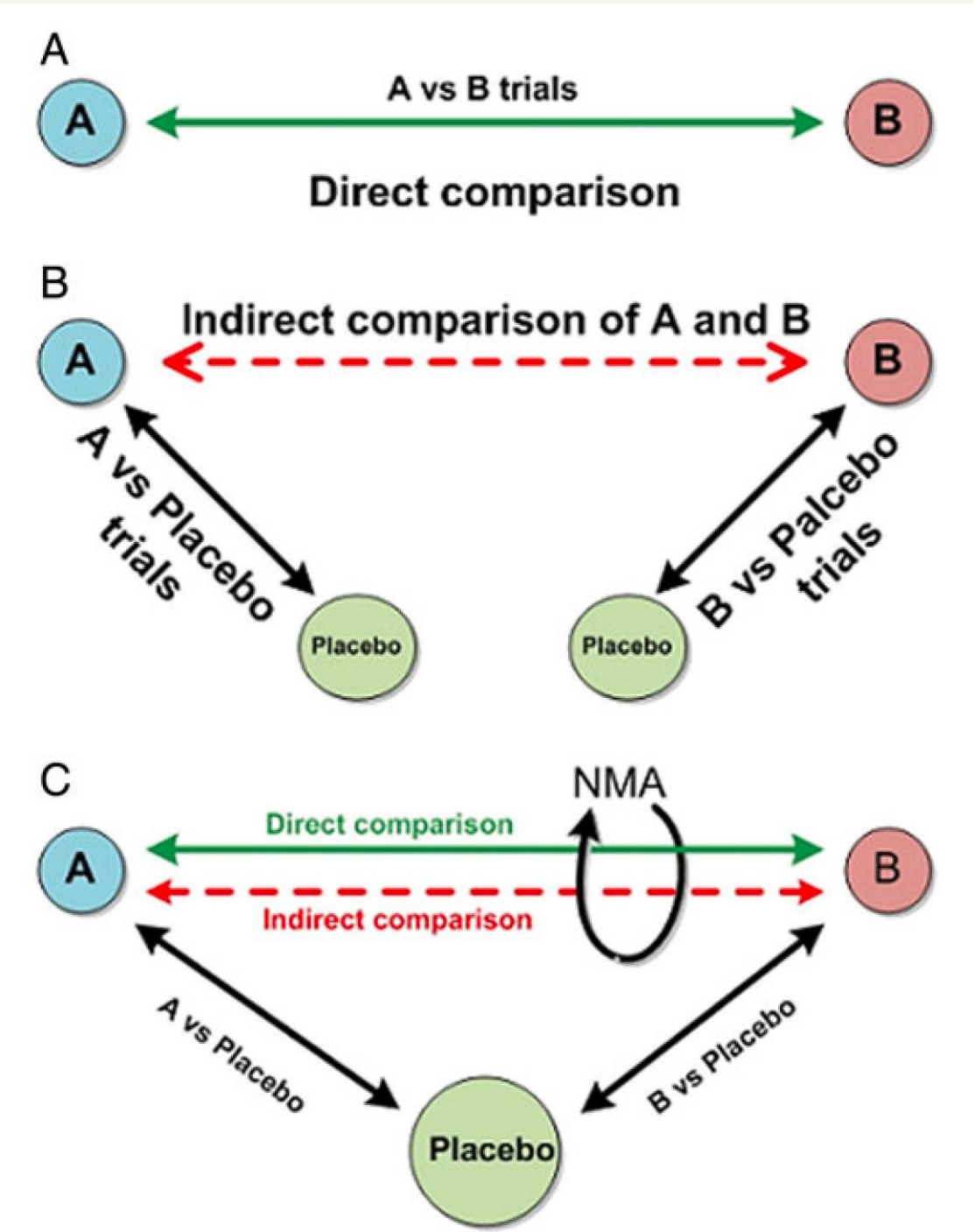


Meta-Analysis of First-Line Therapies in Advanced Non–Small-Cell Lung Cancer Harboring *EGFR*-Activating Mutations

Benjamin Haaland, PhD,† Pui San Tan, MPharm,‡ Gilberto de Castro, Jr, MD, PhD,§ and Gilberto Lopes, MD, MBA, FAMS‖¶*

(*J Thorac Oncol.* 2014;9: 805–811)





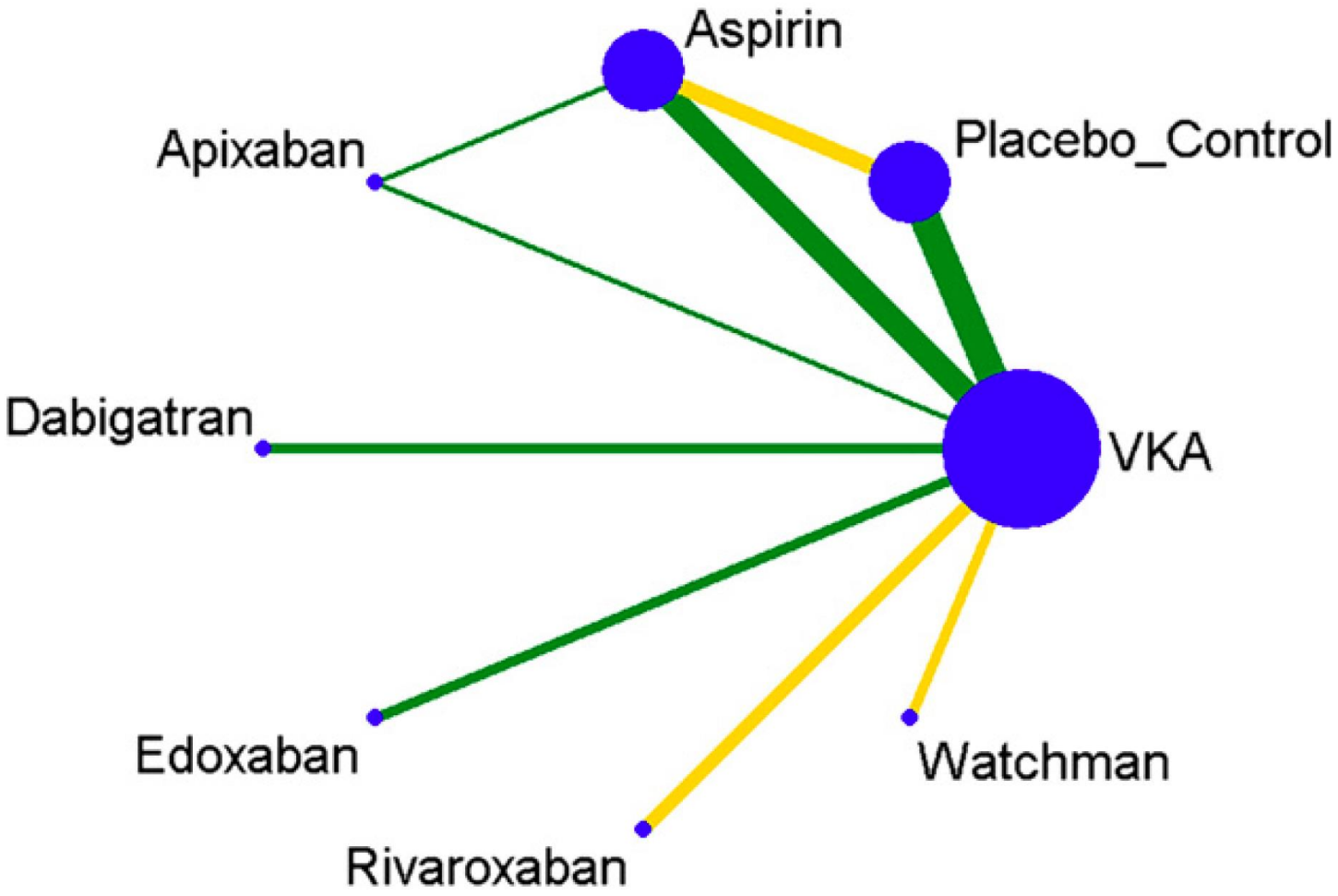
Similarity

Consistency

Comparative Effectiveness of Interventions for Stroke Prevention in Atrial Fibrillation: A Network Meta-Analysis

Larisa G. Tereshchenko, MD, PhD, FHRS; Charles A. Henrikson, MD, MPH, FHRS; Joaquin Cigarroa, MD; Jonathan S. Steinberg, MD, FHRS

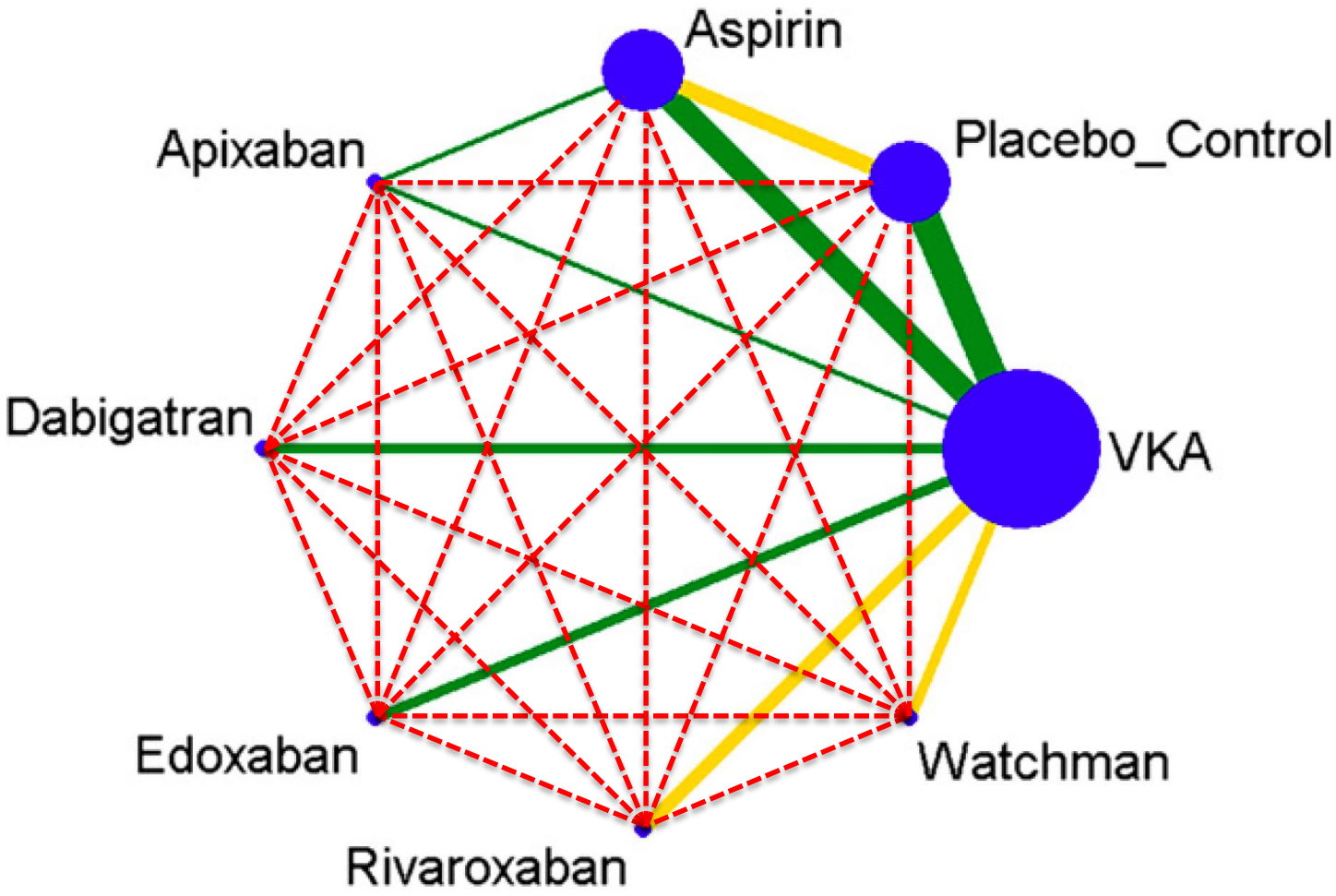
(*J Am Heart Assoc*. 2016;5:e003206 doi: 10.1161/JAHA.116.003206)



Comparative Effectiveness of Interventions for Stroke Prevention in Atrial Fibrillation: A Network Meta-Analysis

Larisa G. Tereshchenko, MD, PhD, FHRS; Charles A. Henrikson, MD, MPH, FHRS; Joaquin Cigarroa, MD; Jonathan S. Steinberg, MD, FHRS

(*J Am Heart Assoc*. 2016;5:e003206 doi: 10.1161/JAHA.116.003206)



Demystifying trial networks and network meta-analysis

Edward J Mills

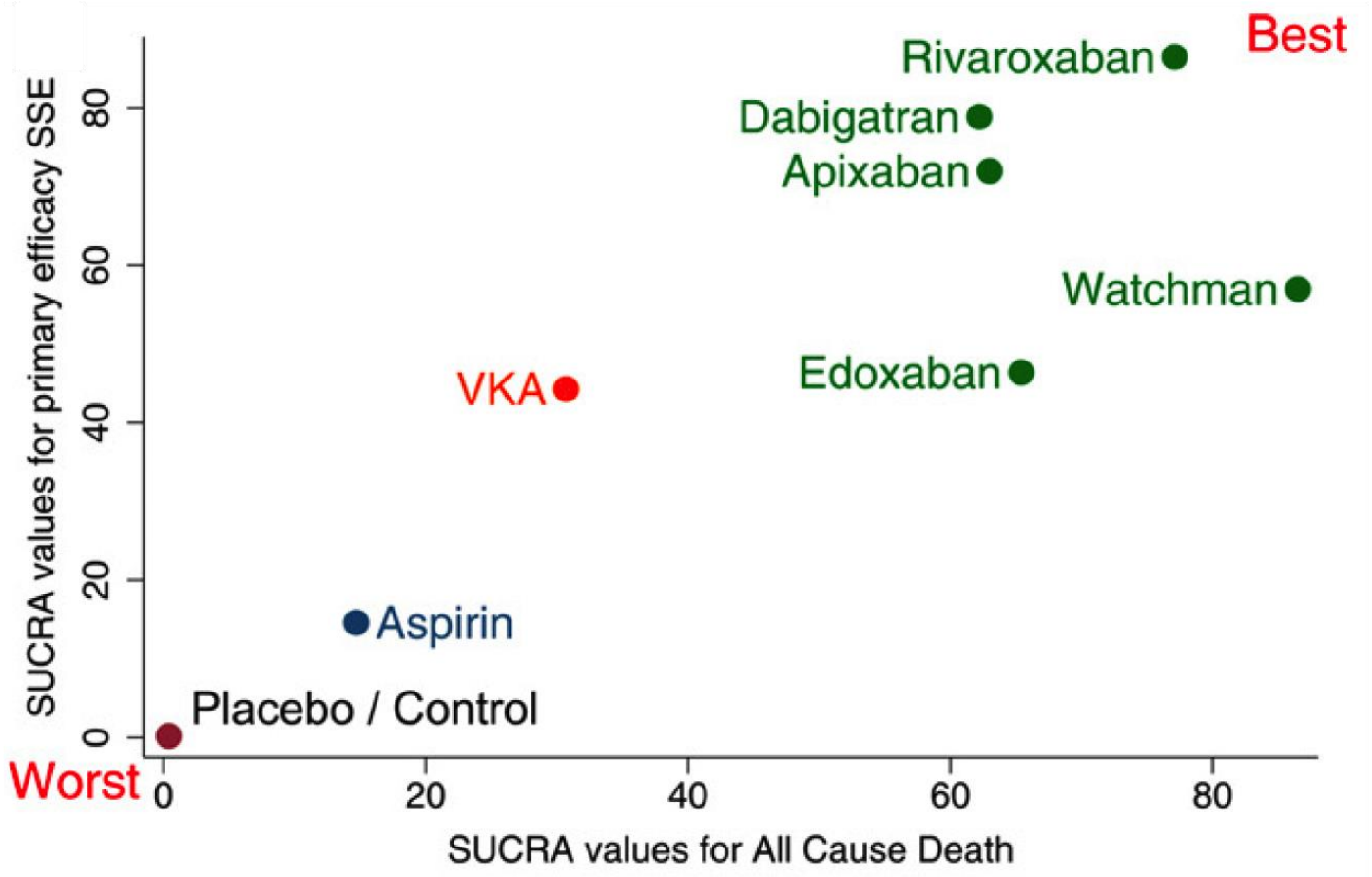
BMJ 2013;346:f2914 doi: 10.1136/bmj.f2914 (Published 14 May 2013)

One of the most appealing but misunderstood elements of network meta-analysis is the reporting of probabilities of which treatment is the best, followed by next best, and so on.

Comparative Effectiveness of Interventions for Stroke Prevention in Atrial Fibrillation: A Network Meta-Analysis

Larisa G. Tereshchenko, MD, PhD, FHRS; Charles A. Henrikson, MD, MPH, FHRS; Joaquin Cigarroa, MD; Jonathan S. Steinberg, MD, FHRS

(*J Am Heart Assoc*. 2016;5:e003206 doi: 10.1161/JAHA.116.003206)



Cluster analysis of surface under the cumulative ranking curves (SUCRA) values

Comparative Effectiveness of Interventions for Stroke Prevention in Atrial Fibrillation: A Network Meta-Analysis

Larisa G. Tereshchenko, MD, PhD, FHRS; Charles A. Henrikson, MD, MPH, FHRS; Joaquin Cigarroa, MD; Jonathan S. Steinberg, MD, FHRS

(*J Am Heart Assoc*. 2016;5:e003206 doi: 10.1161/JAHA.116.003206)

Ranking of the Antithrombotic Interventions

	1° Efficacy: Stroke or Systemic Embolism						1° Safety: Major Bleedings					
	SUCRA		Pr. Best		Rank		SUCRA		Pr. Best		Rank	
Treatment	U	A	U	A	U	A	U	A	U	A	U	A
VKA	44.3	47.5	0	0	4.9	4.7	27	21.9	0	0	6.1	6.5
Placebo/control	0.2	2.9	0	0	8	7.8	96.4	90.8	81.4	72.2	1.2	1.6
Aspirin	14.6	16.3	0	0	7	6.9	61.2	57.3	0.5	2.4	3.7	4
Apixaban	72	72.5	13.2	13.9	3	2.9	57.4	63.1	3.6	3.9	4	3.6
Dabigatran	78.9	75.7	21.1	19.5	2.5	2.7	44.3	46.1	1.1	0.7	4.9	4.8
Edoxaban	46.4	49.5	0.6	2	4.8	4.5	74.7	78.1	12.6	16.8	2.8	2.5
Rivaroxaban	86.5	77	46.1	30.4	1.9	2.6	36.5	23.7	0.8	0.4	5.4	6.3
Watchman	57	58.6	19	34.2	4	3.9	2.5	19	0	3.6	7.8	6.7

A indicates adjusted; Pr. Best, probability of being the best; SUCRA, the surface under the cumulative ranking curve; U, unadjusted

Demystifying trial networks and network meta-analysis

Edward J Mills

BMJ 2013;346:f2914 doi: 10.1136/bmj.f2914 (Published 14 May 2013)

One of the most appealing but misunderstood elements of network meta-analysis is the reporting of probabilities of which treatment is the best, followed by next best, and so on.

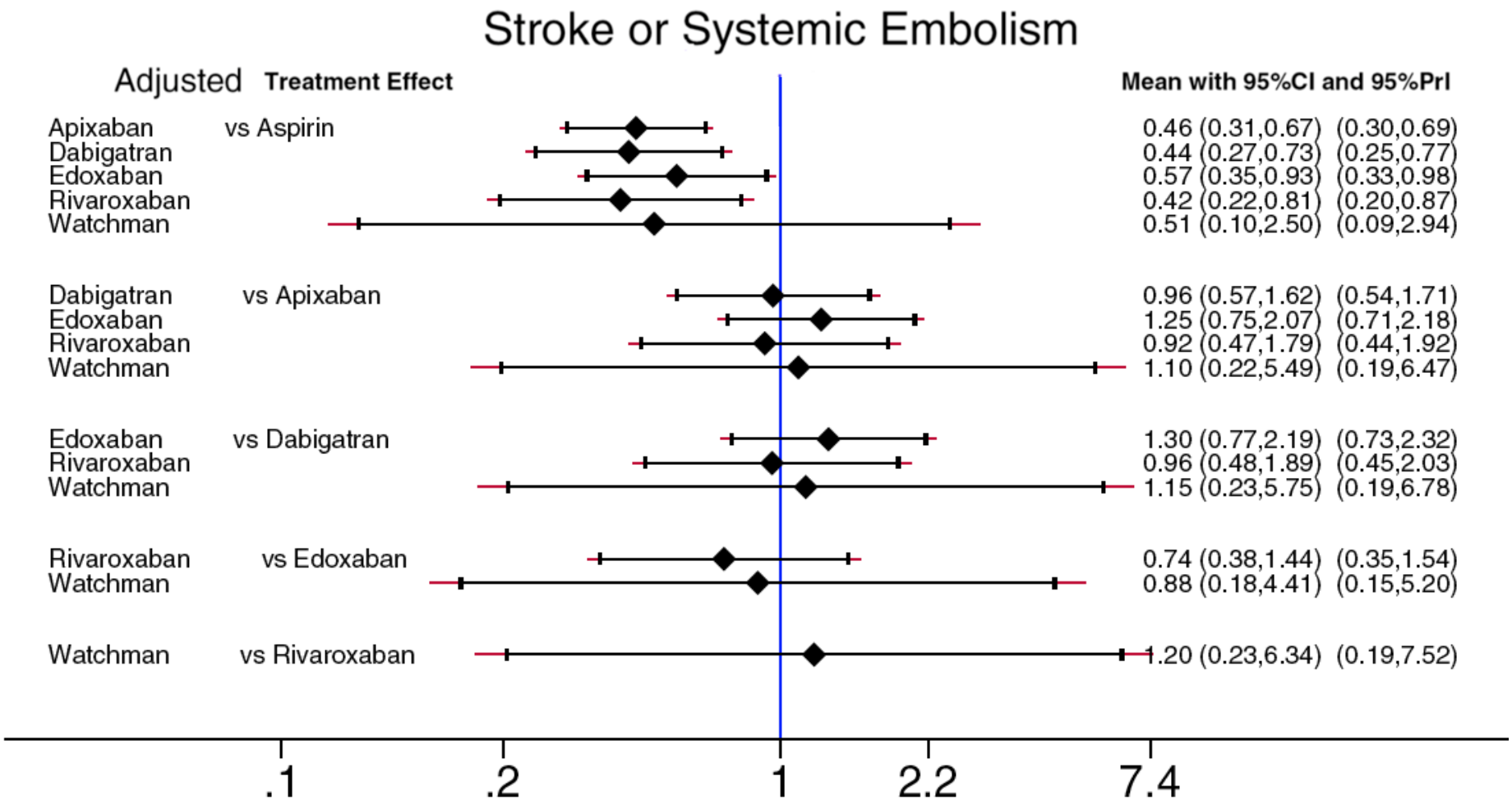
A risk exists that one may incorrectly emphasize the probabilities as being clinically useful when the treatment effects are, in fact, not different from the null beyond chance.

For that reason, authors should place less emphasis on the probabilities of a network meta-analysis output and greater emphasis on the treatment effects and their uncertainty.

Comparative Effectiveness of Interventions for Stroke Prevention in Atrial Fibrillation: A Network Meta-Analysis

Larisa G. Tereshchenko, MD, PhD, FHRS; Charles A. Henrikson, MD, MPH, FHRS; Joaquin Cigarroa, MD; Jonathan S. Steinberg, MD, FHRS

(*J Am Heart Assoc*. 2016;5:e003206 doi: 10.1161/JAHA.116.003206)



Adjusted predictive interval plot for the primary efficacy outcome stroke and systemic embolism

Demystifying trial networks and network meta-analysis

Edward J Mills

BMJ 2013;346:f2914 doi: 10.1136/bmj.f2914 (Published 14 May 2013)

The problem with network analysis in regards to a meta-analysis, is that a network meta-analysis is more likely to be valid when analyzing very similar studies for very similar patient populations.

Since network meta-analysis extends the number and type of studies being combined, there is even more potential for combining studies that are not adequately similar.

Mi posso fidare?

Determinants of quality

5 factors that can lower quality

1. limitations of detailed design and execution *(risk of bias criteria)*
2. Inconsistency *(or heterogeneity)*
3. Indirectness *(PICO and applicability)*
4. Imprecision *(number of events and confidence intervals)*
5. Publication bias

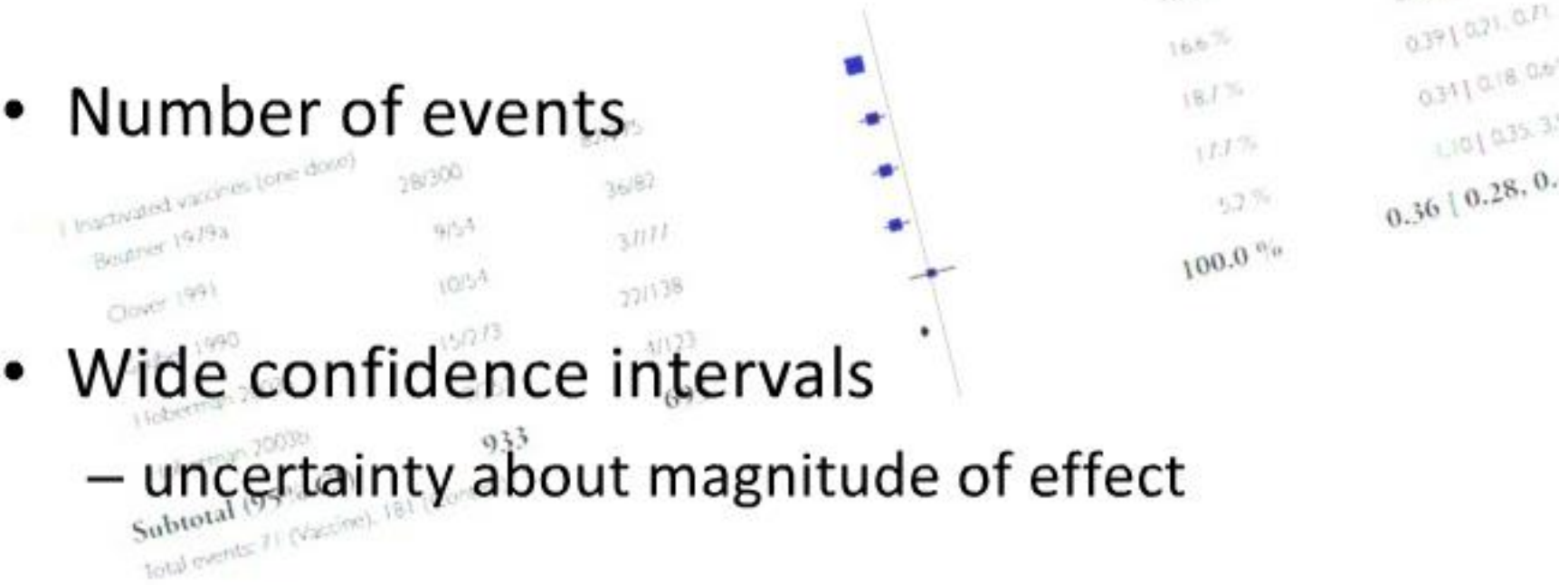
The members of the Grade Working Group

BMJ | 26 APRIL 2008 | VOLUME 336 924

When are results precise enough?

Consider

- Small sample size
  - (Optimal Information Size, OIS)
- Number of events
- Wide confidence intervals
  - uncertainty about magnitude of effect

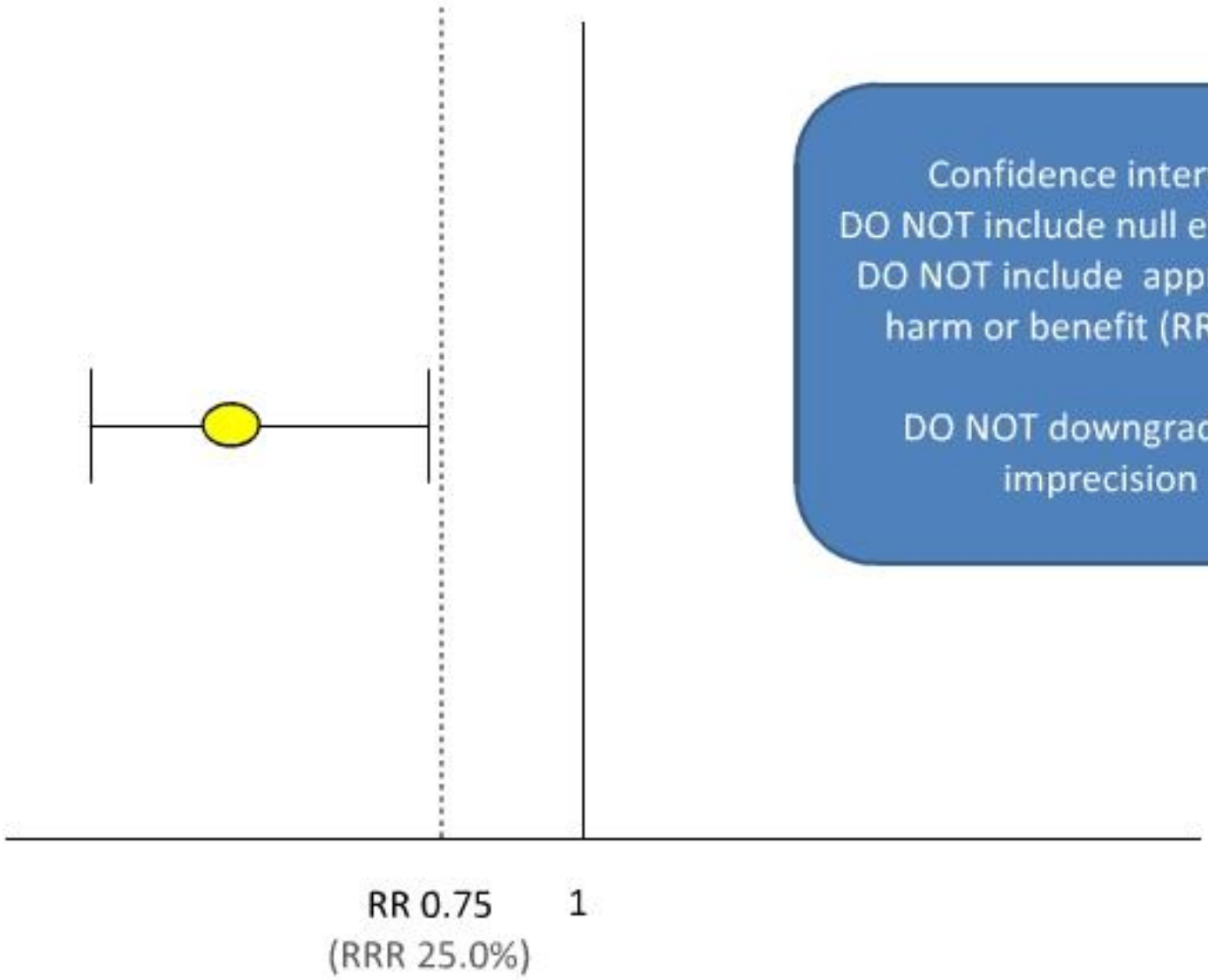


Optimal information size

- We suggest the following:

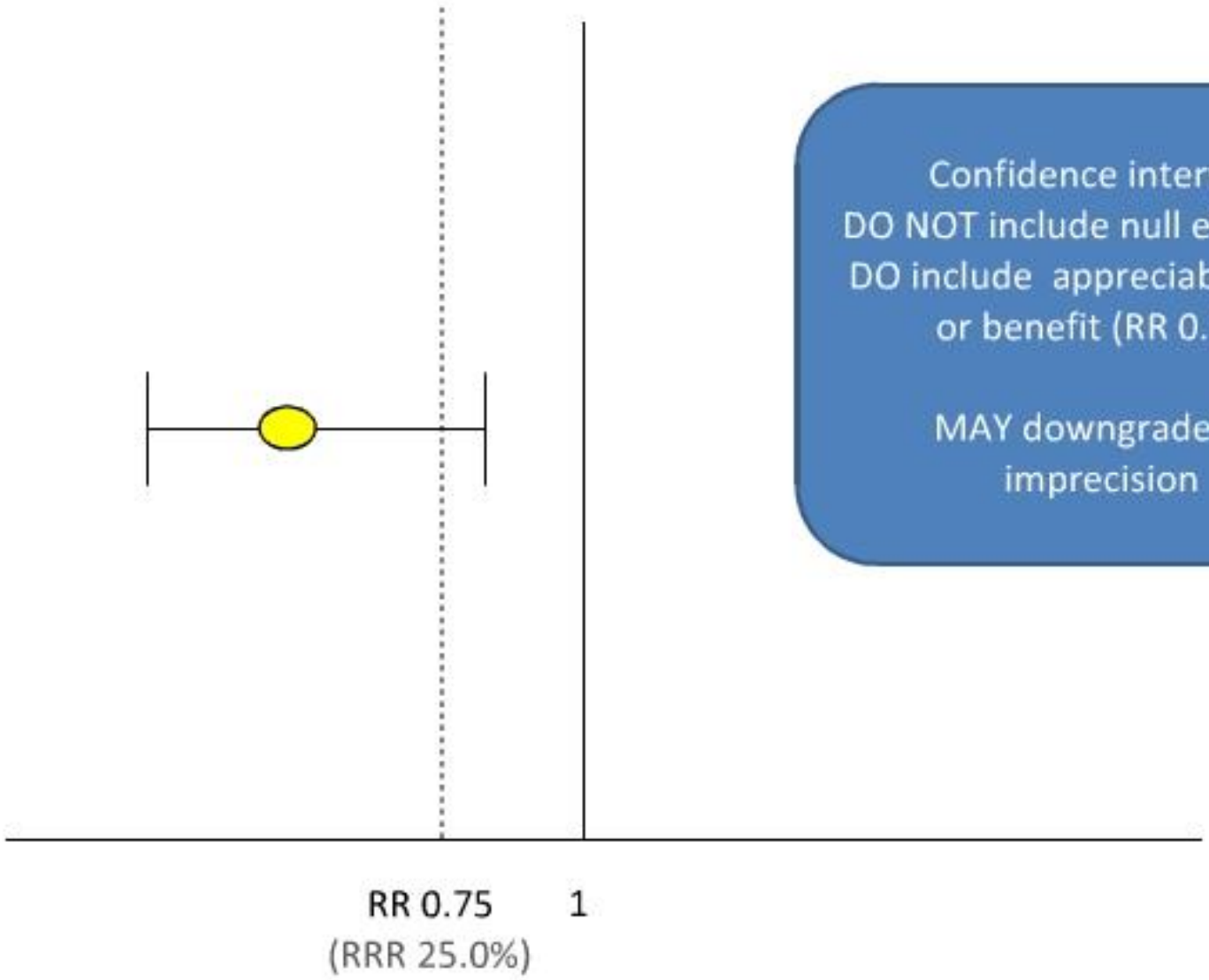
  if the total number of patients included in a systematic review is less than the number of patients generated by a conventional sample size calculation for a single adequately powered trial, consider rating down for imprecision.

  Authors have referred to this threshold as the "optimal information size" (OIS)



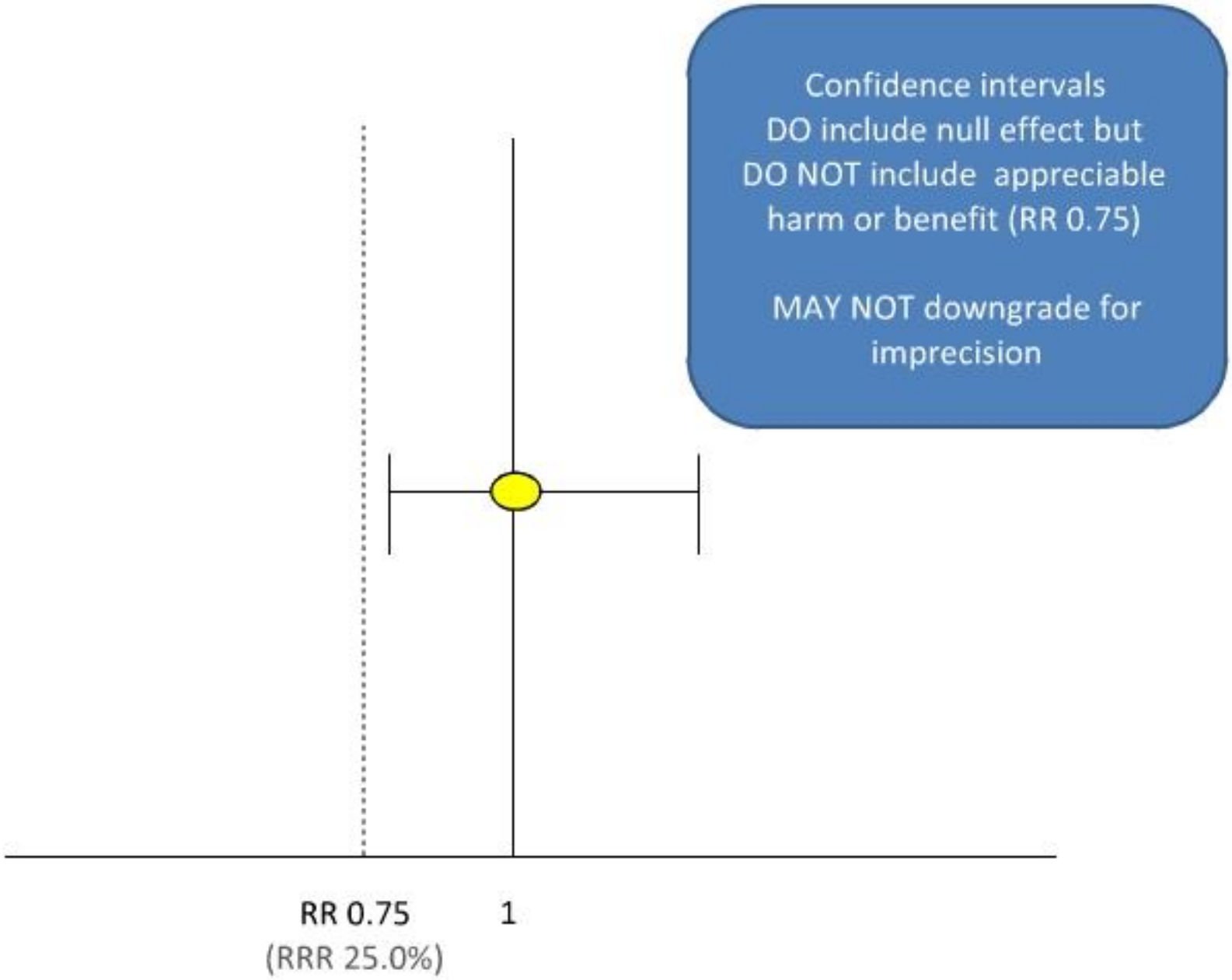
Confidence intervals
DO NOT include null effect and
DO NOT include appreciable
harm or benefit (RR 0.75)

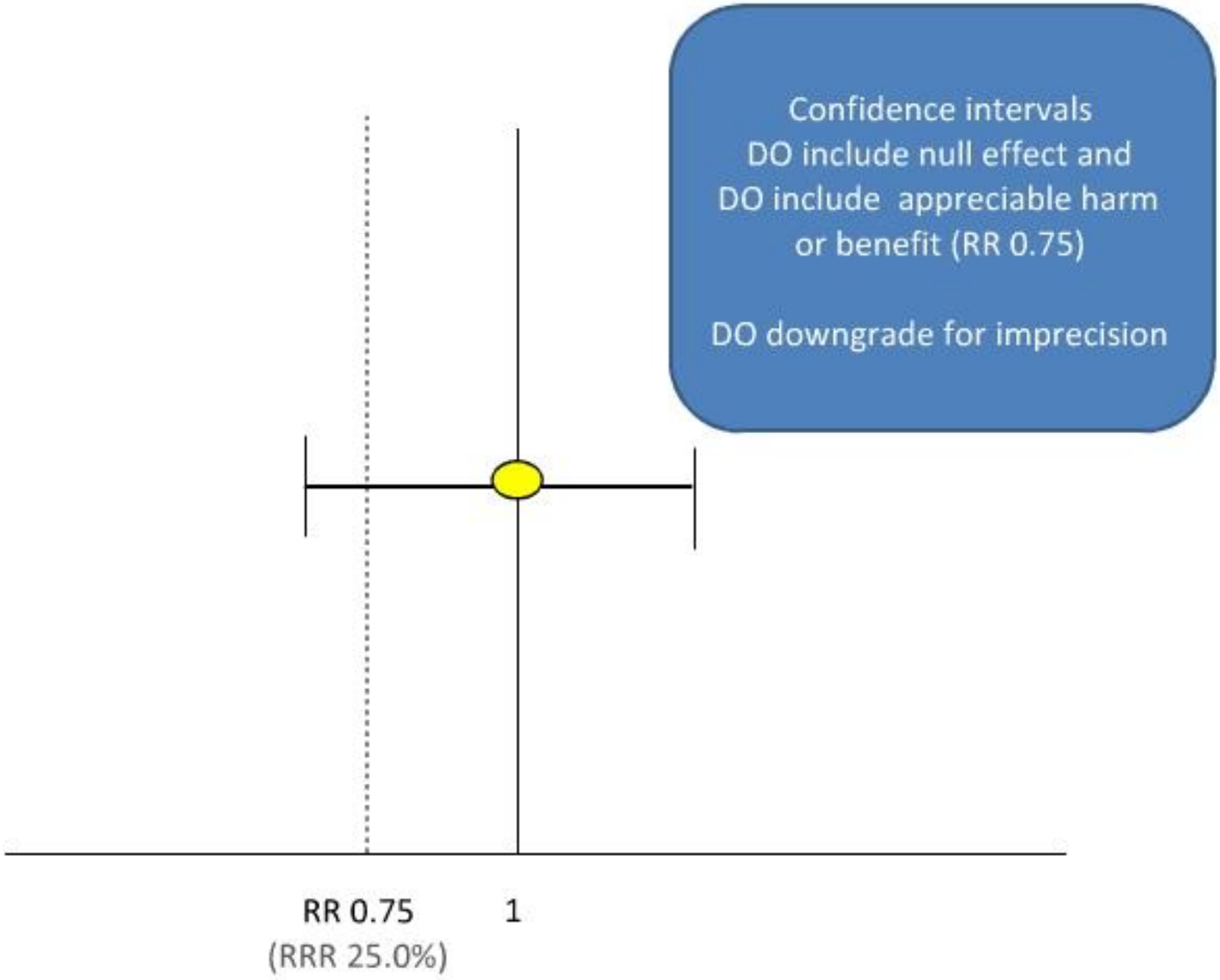
DO NOT downgrade for
imprecision



Confidence intervals
DO NOT include null effect but
DO include appreciable harm
or benefit (RR 0.75)

MAY downgrade for
imprecision


Confidence intervals
DO include null effect but
DO NOT include appreciable
harm or benefit (RR 0.75)
MAY NOT downgrade for
imprecision
RR 0.75
(RRR 25.0%)
1

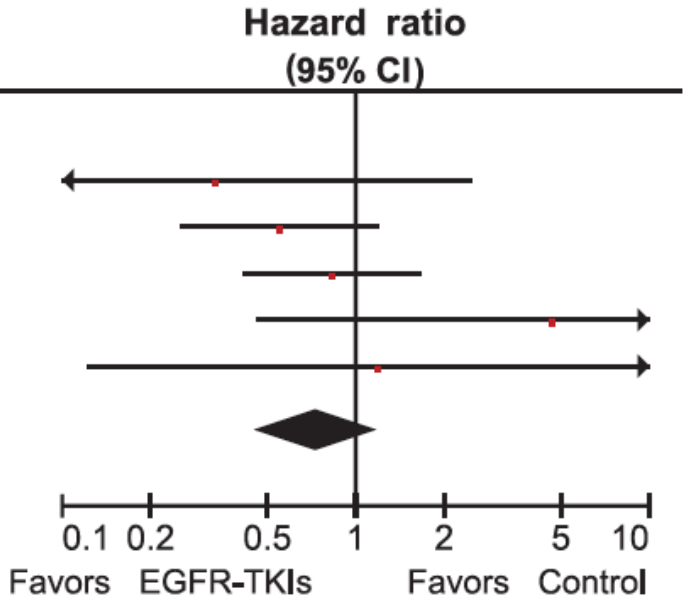

Confidence intervals
DO include null effect and
DO include appreciable harm
or benefit (RR 0.75)
DO downgrade for imprecision
RR 0.75
(RRR 25.0%)
1

Impact of EGFR Inhibitor in Non–Small Cell Lung Cancer on Progression-Free and Overall Survival: A Meta-Analysis

Chee Khoon Lee, Chris Brown, Richard J. Gralla, Vera Hirsh, Sumitra Thongprasert, Chun-Ming Tsai, Eng Huat Tan, James Chung-Man Ho, Da Tong Chu, Adel Zaatar, Jemela Anne Osorio Sanchez, Vu Van Vu, Joseph Siu Kie Au, Akira Inoue, Siow Ming Lee, Val Gebski, James Chih-Hsin Yang

J Natl Cancer Inst;2013;105:595–605

Study	Hazard ratio (95% CI)
EGFRmut$^+$ (second-line/subsequent therapy)	
ISEL	0.33 (0.04 to 2.48)
BR21	0.55 (0.25 to 1.20)
INTEREST	0.83 (0.41 to 1.68)
V-15-32	4.66 (0.46 to 47.40)
TITAN	1.19 (0.12 to 11.64)
Subtotal (95% CI)	**0.74 (0.45 to 1.19)**



L'intervallo comprende sia la rilevanza clinica a favore del braccio sperimentale sia la linea di non-effetto

Lacosamide (LCM) compared to placebo for partial-onset seizures

Participants (studies)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							Risk with placebo	Risk with Lacosamide (LCM)		Risk with placebo	Risk difference with Lacosamide (LCM)
Nausea (assessed with: all dosage arms pooled)											
1105 (3 RCTs)	not serious	not serious	serious [c]	not serious	none	⨁⨁⨁◯ MODERATE	16/364 (4.4%)	73/741 (9.9%)	**RR 2.20** (1.05 to 4.60)	4 per 100	**5 more per 100** (from 0 fewer to 16 more)
Nausea (assessed with: LCM at 200mg)											
530 (2 RCTs)	serious [a]	not serious	not serious	serious [d]	none	⨁⨁◯◯ LOW	11/260 (4.2%)	20/270 (7.4%)	**RR 1.93** (0.49 to 7.56)	4 per 100	**4 more per 100** (from 2 fewer to 28 more)
Nausea (assessed with: LCM at 400mg)											
835	serious [a]	not serious	not serious	not serious	none	⨁⨁⨁◯	16/364 (4.4%)	53/471	**RR 2.43**	4 per 100	**6 more per 100**

Filter by active cell

Explanations | References

a. unplanned subgroup analysis

b. not downgraded for imprecision because the low number of events

c. all dosage arms pooled

d. 95% CIs consistent with conflicting recommendations

Lacosamide (LCM) compared to placebo for partial-onset seizures

Participants (studies)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)		Relative effect (95% CI)	Anticipated absolute effects	
							Risk with placebo	Risk with Lacosamide (LCM)		Risk with placebo	Risk difference with Lacosamide (LCM)
Seizure-free (assessed with: monitoring during the treatment period)											
1105 (3 RCTs)	not serious	not serious	not serious	not serious [b]	none	⨁⨁⨁⨁ HIGH	3/364 (0.8%)	18/741 (2.4%)	**RR 2.01** (0.66 to 6.05)	1 per 100	**1 more per 100** (from 0 fewer to 4 more)
Discontinuation due to AEs (assessed with: all dosage arms pooled)											
1105 (3 RCTs)	not serious	not serious	serious [c]	not serious	none	⨁⨁⨁◯ MODERATE	18/364 (4.9%)	102/741 (13.8%)	**RR 2.73** (1.68 to 4.44)	5 per 100	**9 more per 100** (from 3 more to 17 more)
Discontinuation due to AEs (assessed with: LCM at 200mg)											

Filter by active cell

Explanations | References

a. unplanned subgroup analysis

b. not downgraded for imprecision because the low number of events

c. all dosage arms pooled

d. 95% CIs consistent with conflicting recommendations

Determinants of quality

5 factors that can lower quality

1. limitations of detailed design and execution *(risk of bias criteria)*
2. Inconsistency *(or heterogeneity)*
3. Indirectness *(PICO and applicability)*
4. Imprecision *(number of events and confidence intervals)*
5. Publication bias

The members of the Grade Working Group

BMJ | 26 APRIL 2008 | VOLUME 336 **924**

What is Heterogeneity?

- **Any kind of variability among studies in a systematic review may be termed heterogeneity.**

Eterogeneità delle stime di effetto tra gli studi che *non trova spiegazione logica* (diversità nel tipo di intervento o nella composizione delle popolazioni studiate)

What is Heterogeneity?

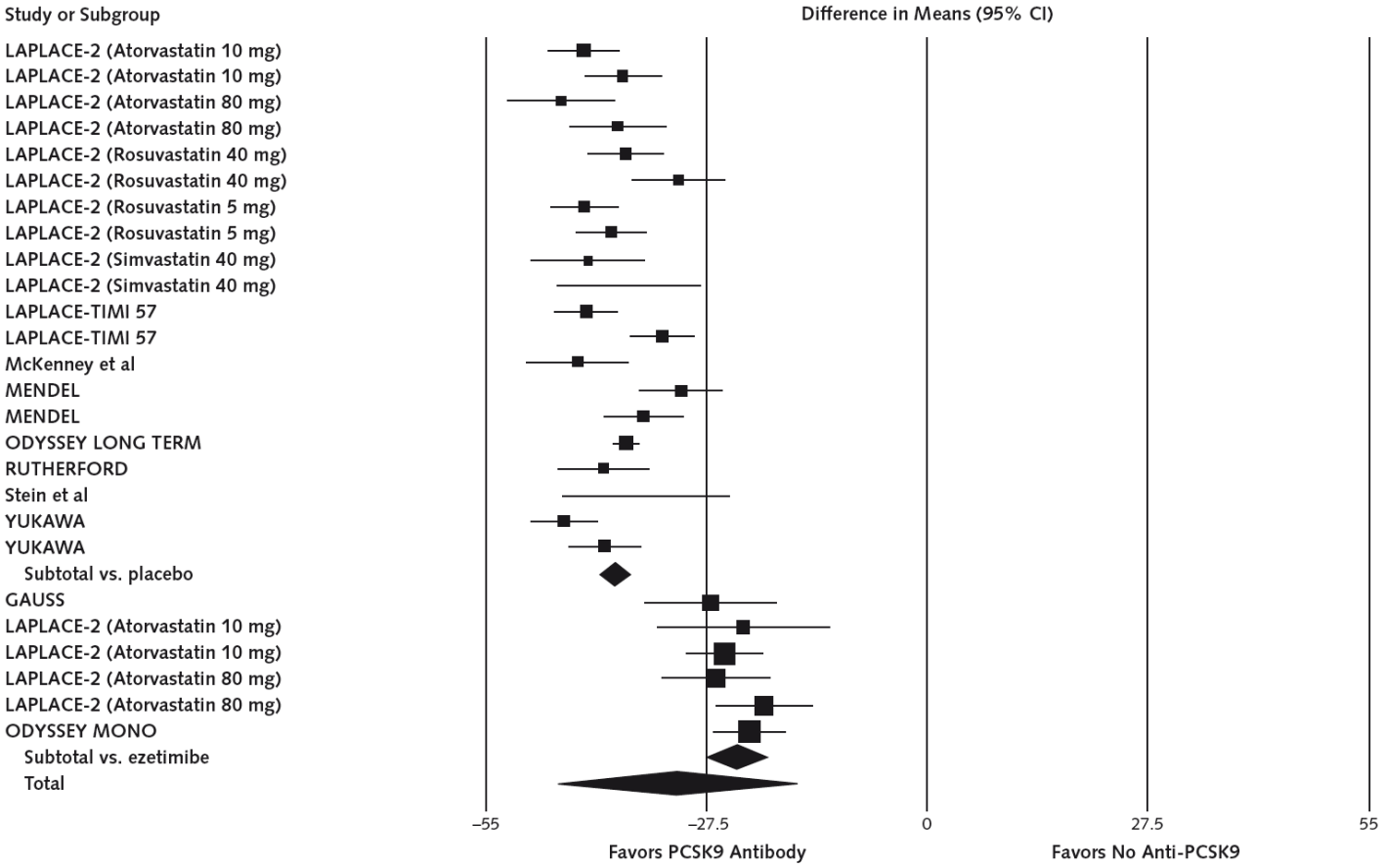
- **Any kind of variability among studies in a systematic review may be termed heterogeneity.**
- **I-squared (I^2)** $I^2 = \left(\frac{Q - df}{Q}\right) \times 100\%$, where Q is the chi-squared statistic and df is its degrees of freedom
  - ✓ describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance).
  - ✓ thresholds for the interpretation of I^2:
    - 0% to 40%: might not be important;
    - 30% to 60%: may represent moderate heterogeneity;
    - **50% to 90%: may represent substantial heterogeneity;**
    - 75% to 100%: considerable heterogeneity.

Effects of Proprotein Convertase Subtilisin/Kexin Type 9 Antibodies in Adults With Hypercholesterolemia

A Systematic Review and Meta-analysis

Eliano Pio Navarese, MD, PhD; Michalina Kołodziejczak, MD; Volker Schulze, MD; Paul A. Gurbel, MD; Udaya Tantry, PhD; Yingfeng Lin, MD; Maximilian Brockmeyer, MD; David E. Kandzari, MD; Julia M. Kubica, MD; Ralph B. D'Agostino Sr., PhD; Jacek Kubica, MD, PhD; Massimo Volpe, MD; Stefan Agewall, MD; Dean J. Kereiakes, MD; and Malte Kelm, MD

Ann Intern Med. 2015;163:40-51.



Group	Effect Size (95% CI)						Test of Null (2-Tail)		Heterogeneity		
	Number of Studies	Point Estimate	SE	Variance	Lower Limit	Upper Limit	*Z* Value	*P* Value	*Q* Value	*P* Value	I^2
Random-effects analysis											
Overall	26	−31.492	7.580	57.455	−46.348	−16.635	−4.155	0.000	187.788	0.000	86.687

Inconsistency: simple rule of thumb

When studies yield widely differing estimates of effect... or heterogeneity....

- Look for reasons for heterogeneity
  (e.g. differences in populations, interventions, outcomes)

Your confidence in the results is lower when there is unexplained heterogeneity

➡ lower quality of the evidence

Determinants of quality

5 factors that can lower quality

1. limitations of detailed design and execution *(risk of bias criteria)*
2. Inconsistency *(or heterogeneity)*
3. Indirectness *(PICO and applicability)*
4. Imprecision *(number of events and confidence intervals)*
5. Publication bias

The members of the Grade Working Group

BMJ | 26 APRIL 2008 | VOLUME 336 **924**

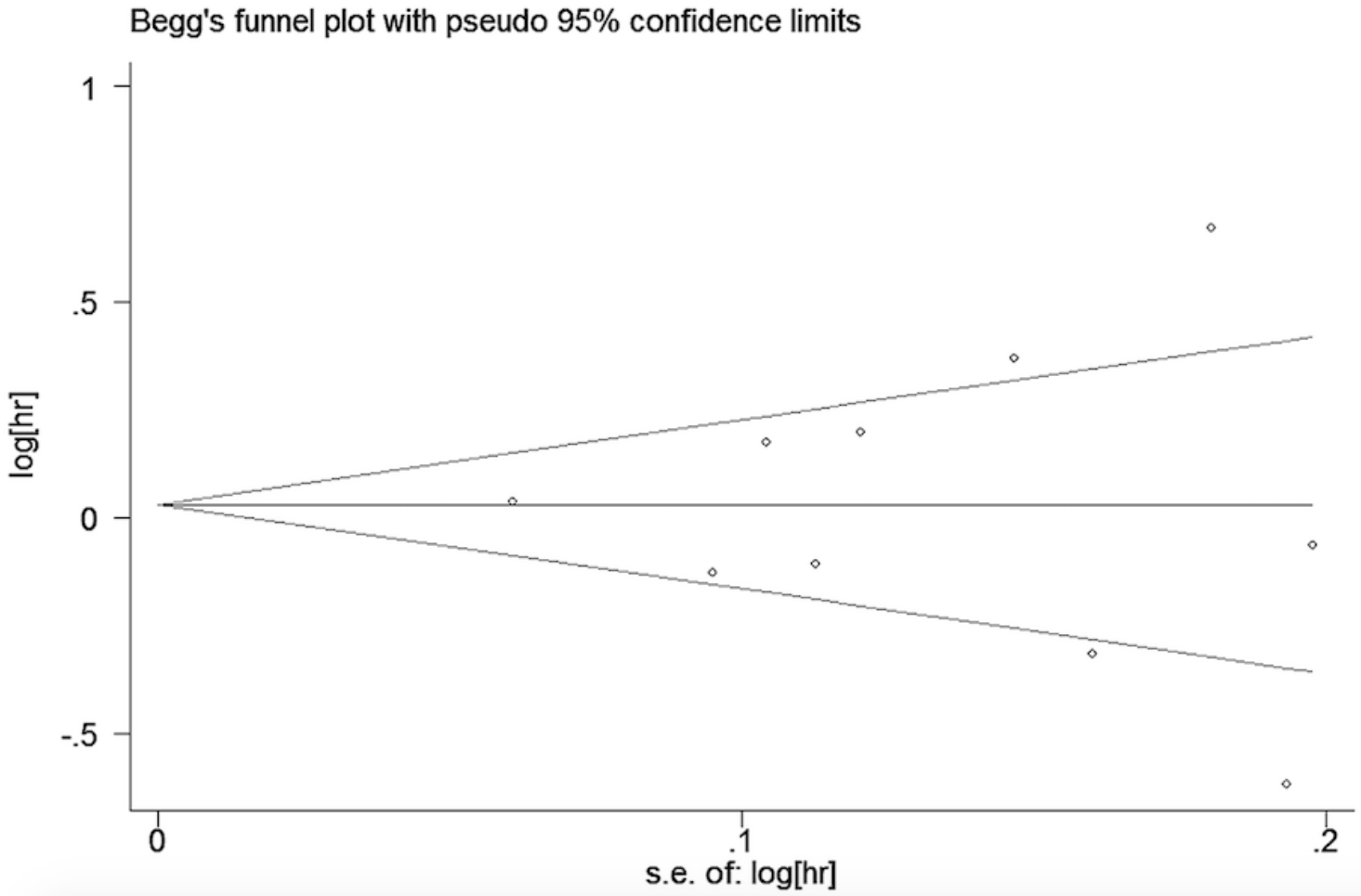
Publication bias

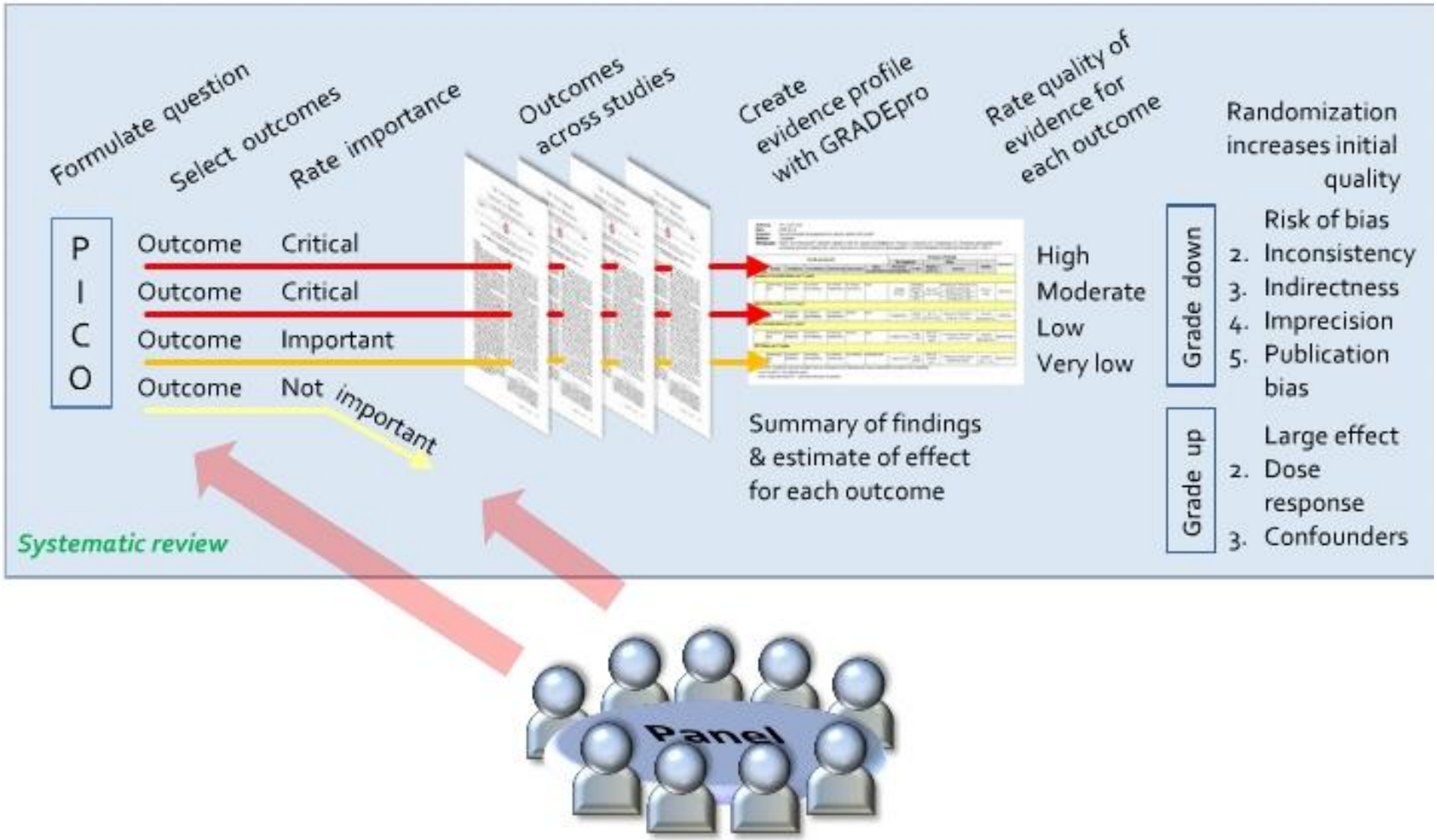
- Trials with statistically significant results ("positive trials") are
  - ✓ more likely to get published
  - ✓ more likely to get published early (estimates are in years)
  - ✓ more likely to get multiple publications
- Meta-analyses based on only published results are biased

Meta-Analysis of EGFR Tyrosine Kinase Inhibitors Compared with Chemotherapy as Second-Line Treatment in Pretreated Advanced Non-Small Cell Lung Cancer

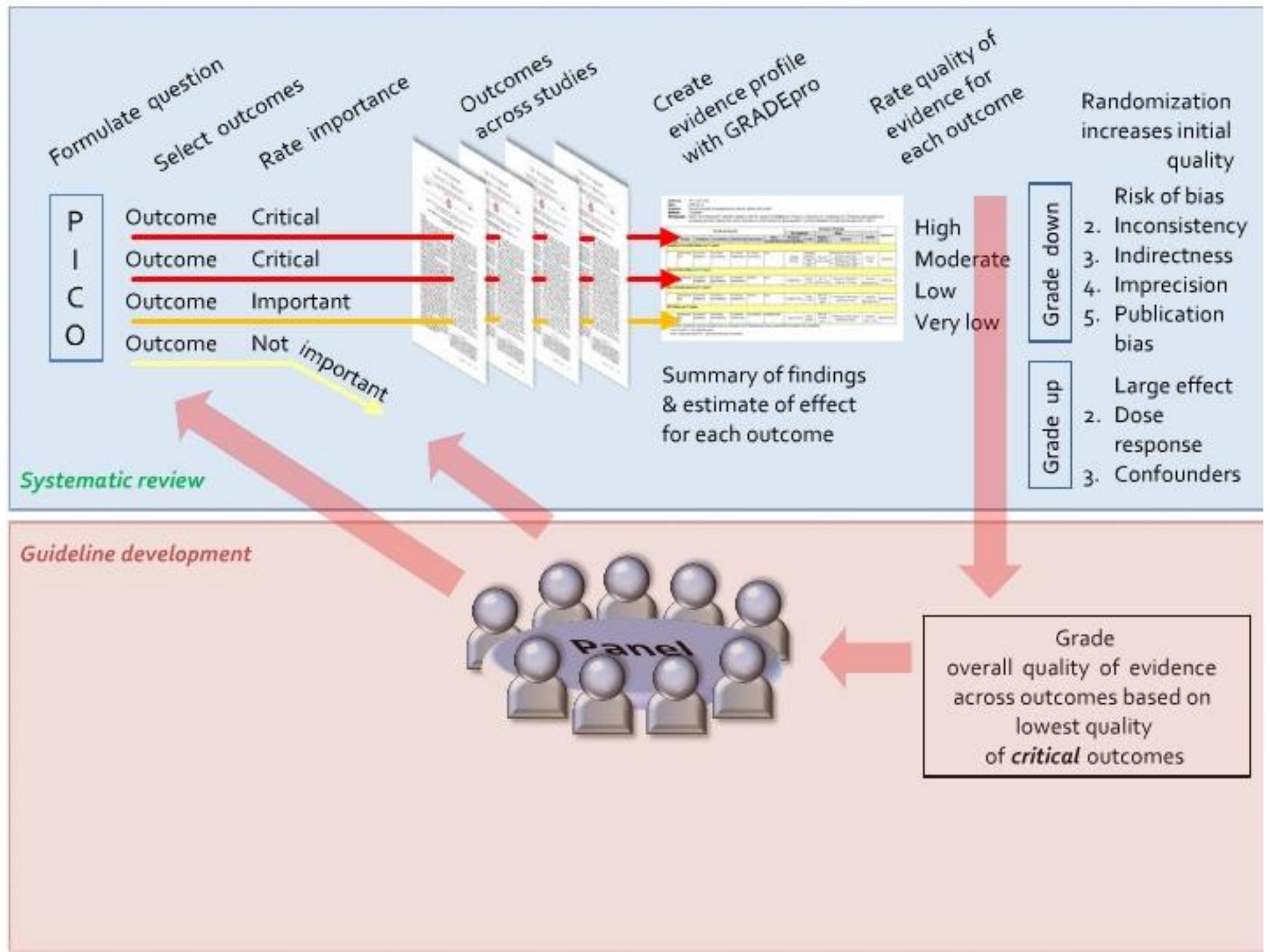
Ning Li[1☯], Lu Yang[2☯], Wei Ou[1], Liang Zhang[3], Song-liang Zhang[1], Si-yu Wang[1*]

PLoS ONE 9(7): e102777. doi:10.1371/journal.pone.0102777




Formulate question
Select outcomes
Rate importance
Outcomes across studies
Create evidence profile with GRADEpro
Rate quality of evidence for each outcome
P
I
C
O
Outcome Critical
Outcome Critical
Outcome Important
Outcome Not important
High
Moderate
Low
Very low
Summary of findings & estimate of effect for each outcome
Randomization increases initial quality
Grade down
1. Risk of bias
2. Inconsistency
3. Indirectness
4. Imprecision
5. Publication bias
Grade up
1. Large effect
2. Dose response
3. Confounders
Systematic review

Panel

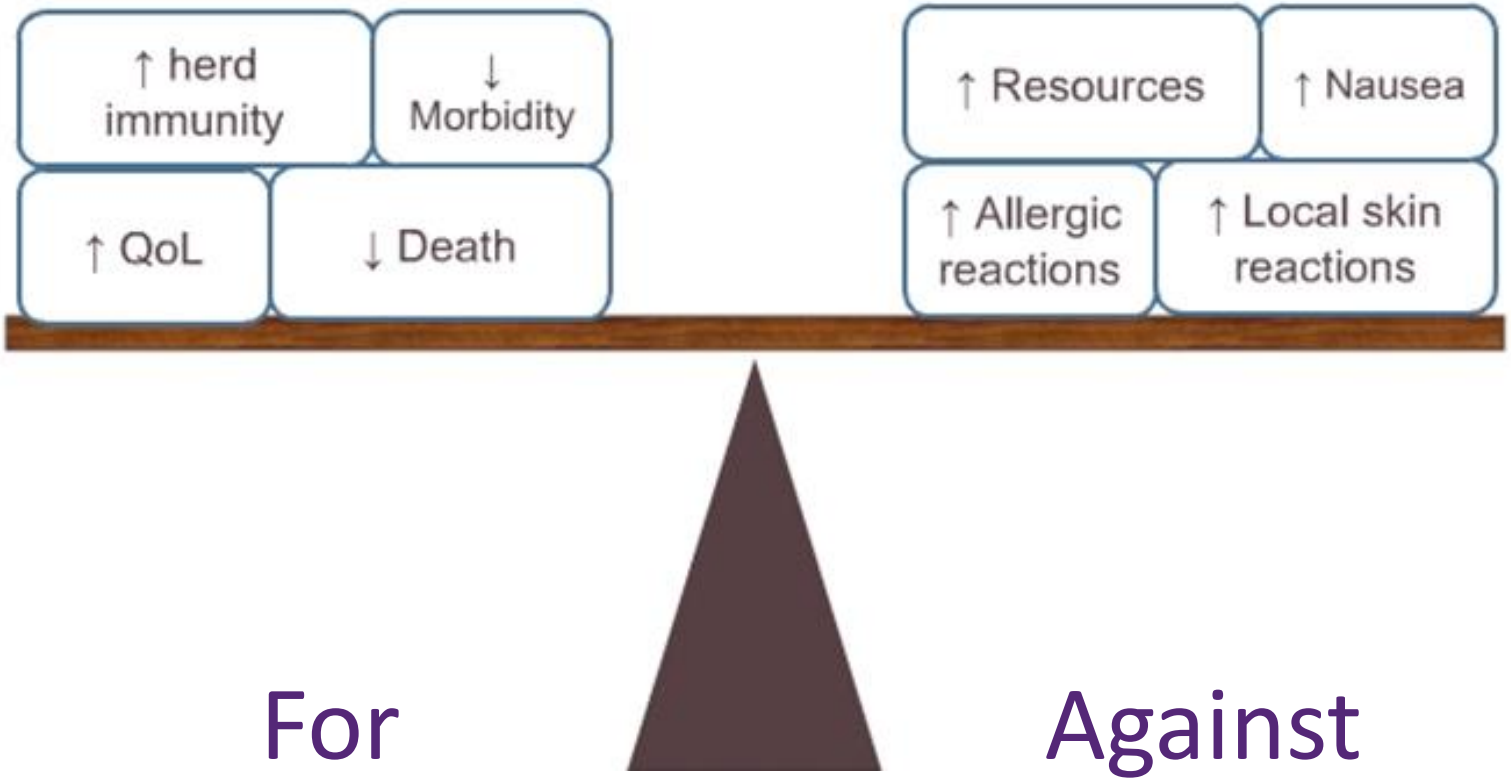

Formulate question
Select outcomes
Rate importance
Outcomes across studies
Create evidence profile with GRADEpro
Rate quality of evidence for each outcome
P
I
C
O
Outcome Critical
Outcome Critical
Outcome Important
Outcome Not important
Summary of findings & estimate of effect for each outcome
High
Moderate
Low
Very low
Randomization increases initial quality
Grade down
1. Risk of bias
2. Inconsistency
3. Indirectness
4. Imprecision
5. Publication bias
Grade up
1. Large effect
2. Dose response
3. Confounders
Systematic review
Guideline development
Panel
Grade overall quality of evidence across outcomes based on lowest quality of critical outcomes

Rapporto Beneficio / Danno

- **Bilancio tra gli effetti positivi (benefici) e negativi (effetti dannosi) dell'intervento**
- **Definito da:**
  - **importanza degli *outcomes***
  - **qualità dell'evidenza**
  - **rischio di base degli eventi che l'intervento dovrebbe essere in grado di ridurre**
  - **entità degli effetti (rilevanza clinico-epidemiologica)**

http://asr.regione.emilia-romagna.it/wcm/asr/collana_dossier/doss172.htm

Balancing benefits and downsides



Rapporto Beneficio / Danno

- Bilancio tra gli effetti positivi (benefici) e negativi (effetti dannosi) dell'intervento
- **Definito da:**
  - **importanza degli *outcomes***
  - **qualità dell'evidenza**
  - **rischio di base degli eventi che l'intervento dovrebbe essere in grado di ridurre**
  - **entità degli effetti (rilevanza clinico-epidemiologica)**

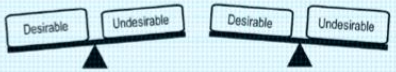
http://asr.regione.emilia-romagna.it/wcm/asr/collana_dossier/doss172.htm

Il Rapporto tra Benefici e Danni

Il rapporto tra benefici e danni	
in dettaglio	in sintesi
Evidenza che i benefici sono prevalenti sui danni	Favorevole
Incertezza sulla prevalenza dei benefici sui danni	Incerto
Incertezza sulla prevalenza dei danni sui benefici	
Evidenza che i danni sono prevalenti sui benefici	Sfavorevole

– **Net benefit**

– **No net benefit harm**

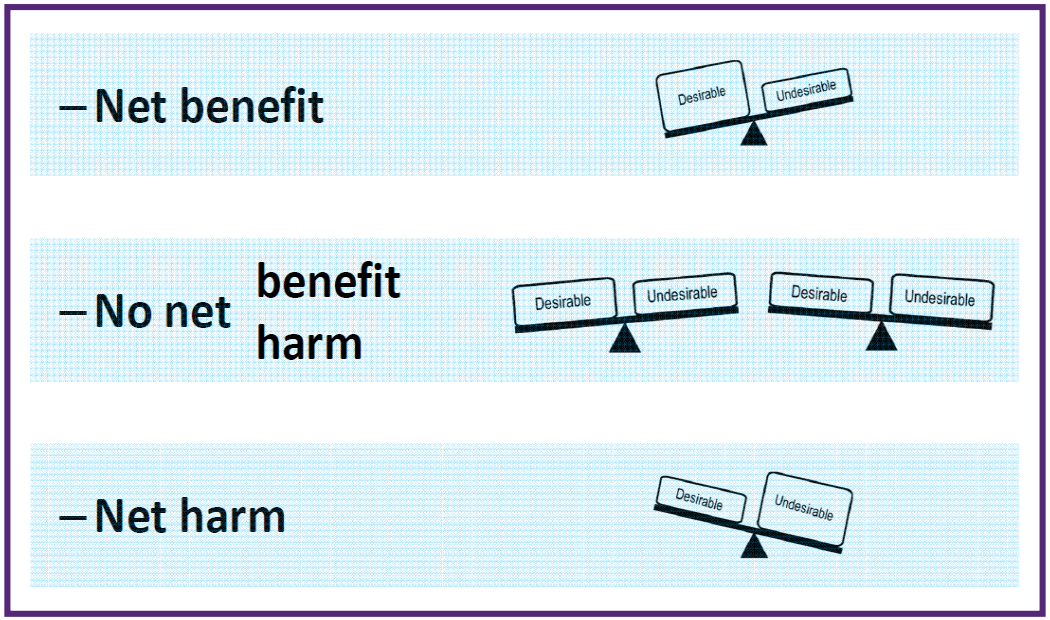


– **Net harm**

G.L. Pappagallo per il Gruppo Metodologico Linee Guida AIOM, 2015

Bilancio tra benefici e danni e direzione della raccomandazione

La direzione a favore o contro l'uso del trattamento si dovrebbe basare sul bilancio tra gli effetti positivi (benefici) e negativi (effetti dannosi) dell'intervento.



http://asr.regione.emilia-romagna.it/wcm/asr/collana_dossier/doss172.htm

LG Aiom 2017: La Sintesi

Il rapporto tra benefici e danni		La Forza	*L'intervento terapeutico…*
in dettaglio	in sintesi		
Evidenza che i benefici sono prevalenti sui danni	Favorevole	Positiva Forte	***… dovrebbe*** *essere preso in considerazione*
Incertezza sulla prevalenza dei benefici sui danni	Incerto	Positiva Debole	***… può*** *essere preso in considerazione*
Incertezza sulla prevalenza dei danni sui benefici		Negativa Debole	***… non dovrebbe*** *essere preso in considerazione*
Evidenza che i danni sono prevalenti sui benefici	Sfavorevole	Negativa Forte	***… non deve*** *essere preso in considerazione*