SACRO CUORE DON CALABRIA

IRCCS

ECI CANCER CENTRE

SCUOLA DI METODOLOGIA DELLA RICERCA CLINICA

2025 - 11ª EDIZIONE

MODULI SPECIALISTICI - S4

VENERDÌ 16 - SABATO 17 MAGGIO 2025

NEGRAR DI VALPOLICELLA (VR)

Centro Formazione IRCCS "Sacro Cuore - Don Calabria"

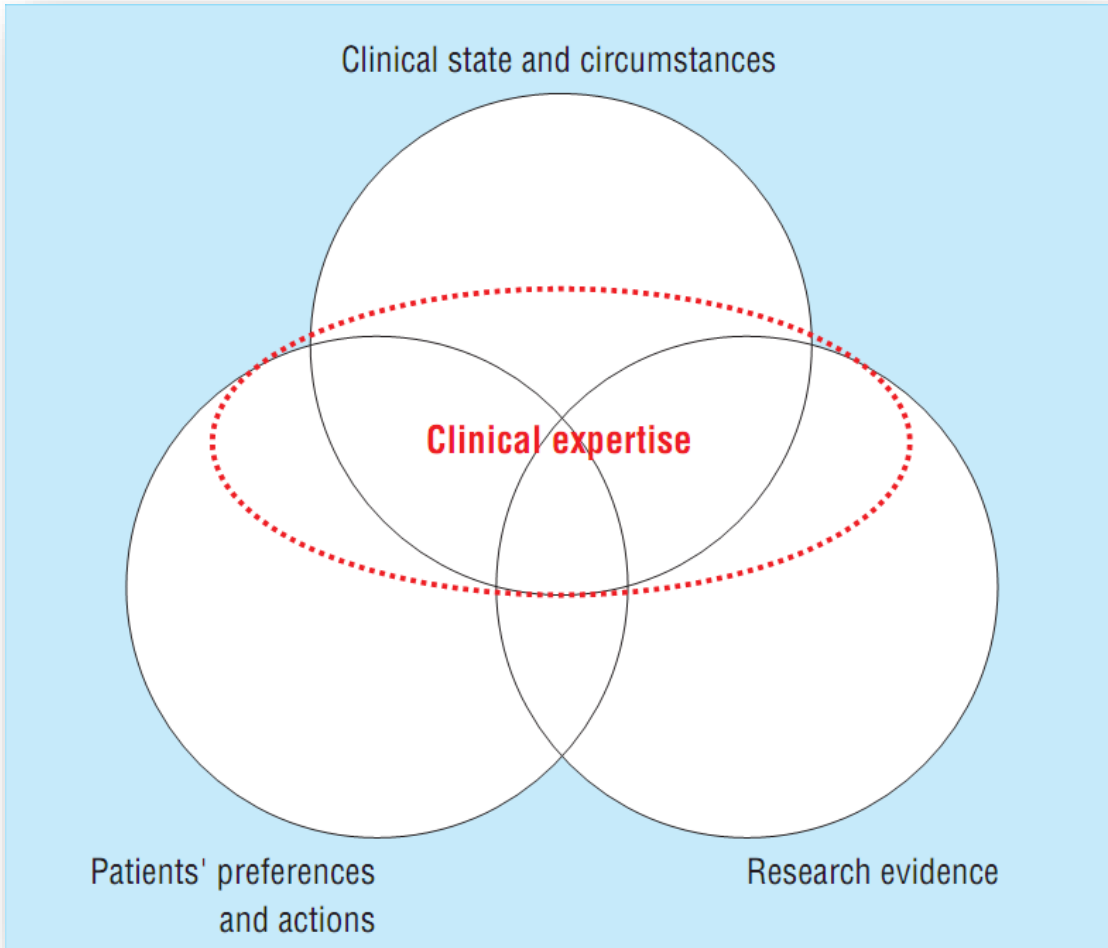
Una Premessa…

Physicians' and patients' choices in evidence based practice

Evidence does not make decisions, people do

R Brian Haynes P J Devereaux Gordon H Guyatt

BMJ 2002;324:1350

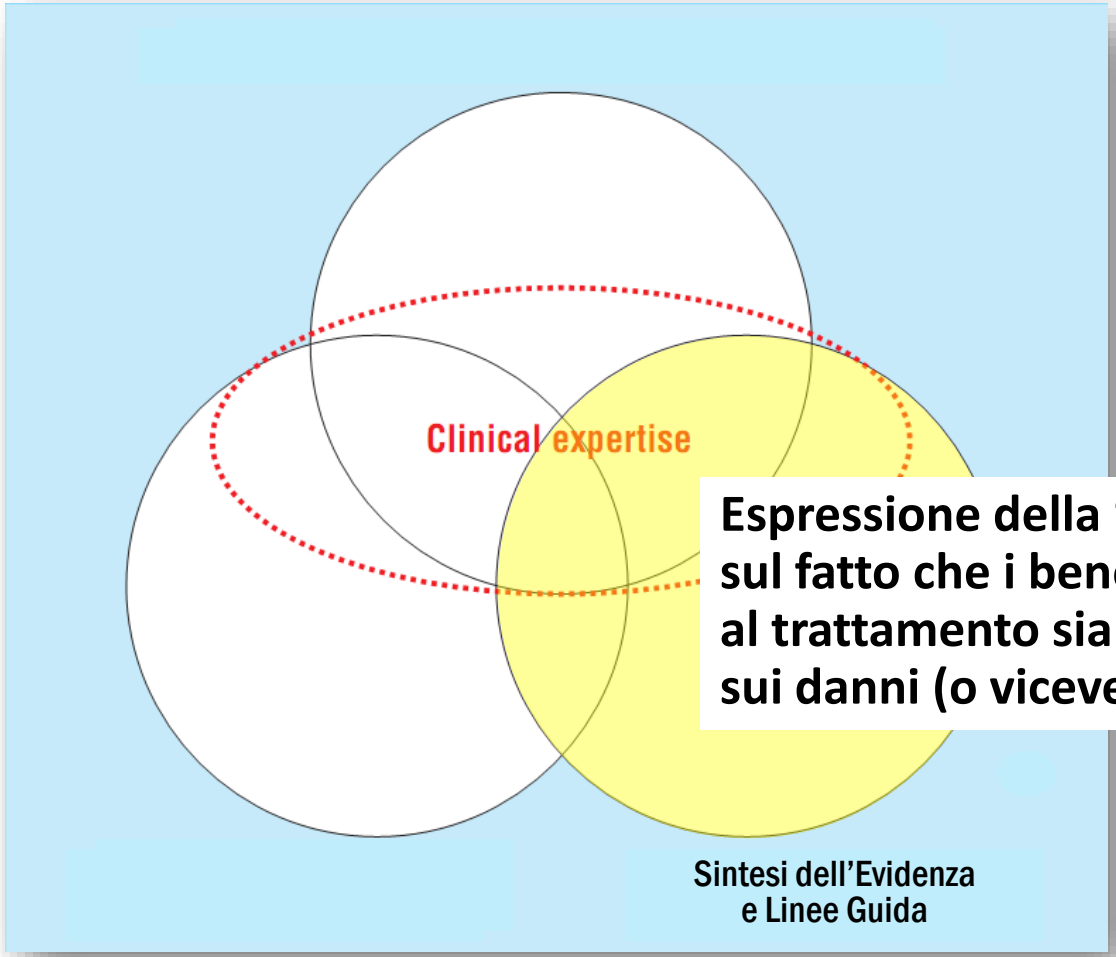


Physicians' and patients' choices in evidence based practice

Evidence does not make decisions, people do

R Brian Haynes P J Devereaux Gordon H Guyatt

BMJ 2002;324:1350



Espressione della "confidenza" sul fatto che i benefici ascrivibili al trattamento siano prevalenti sui danni (o viceversa)

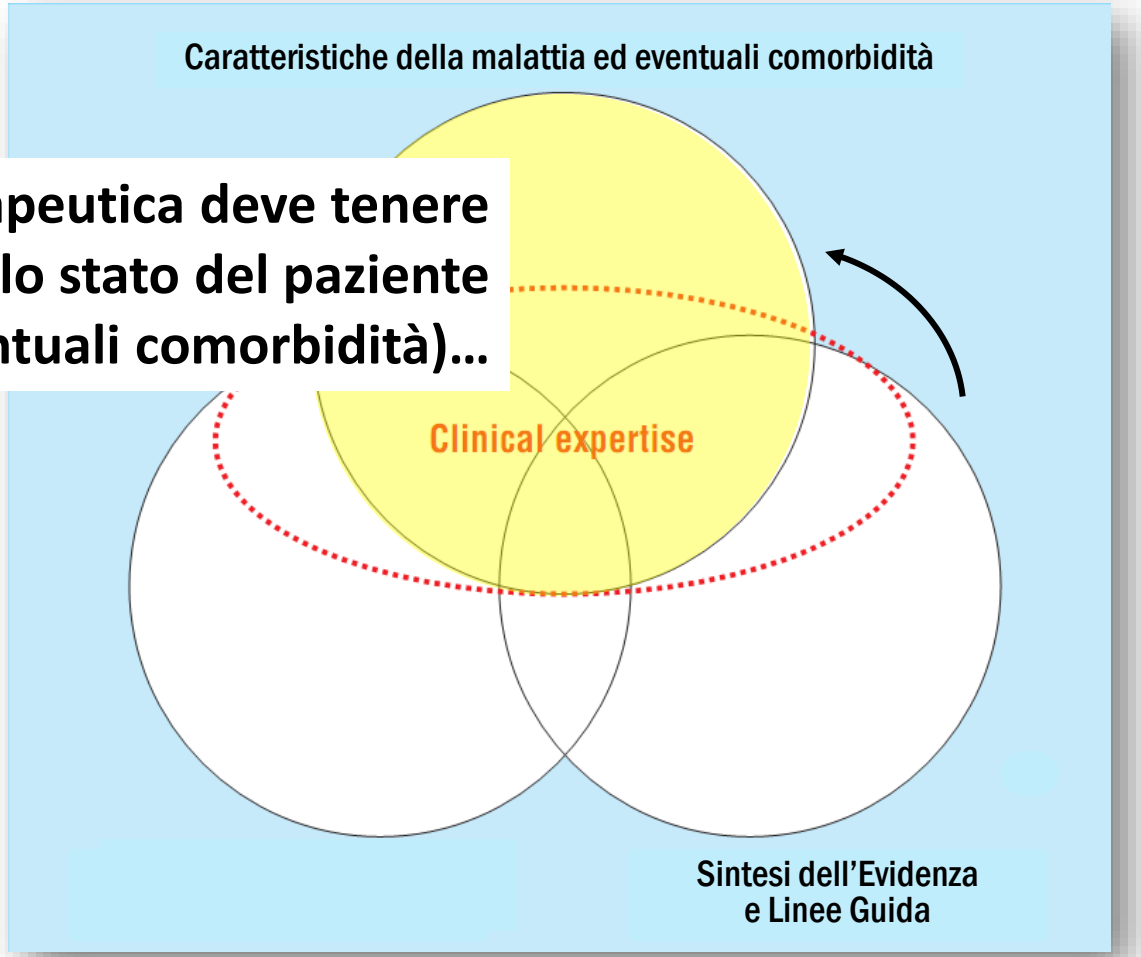
Physicians' and patients' choices in evidence based practice

Evidence does not make decisions, people do

R Brian Haynes P J Devereaux Gordon H Guyatt

BMJ 2002;324:1350

La proposta terapeutica deve tenere conto dello stato del paziente (malattia ed eventuali comorbidità)…

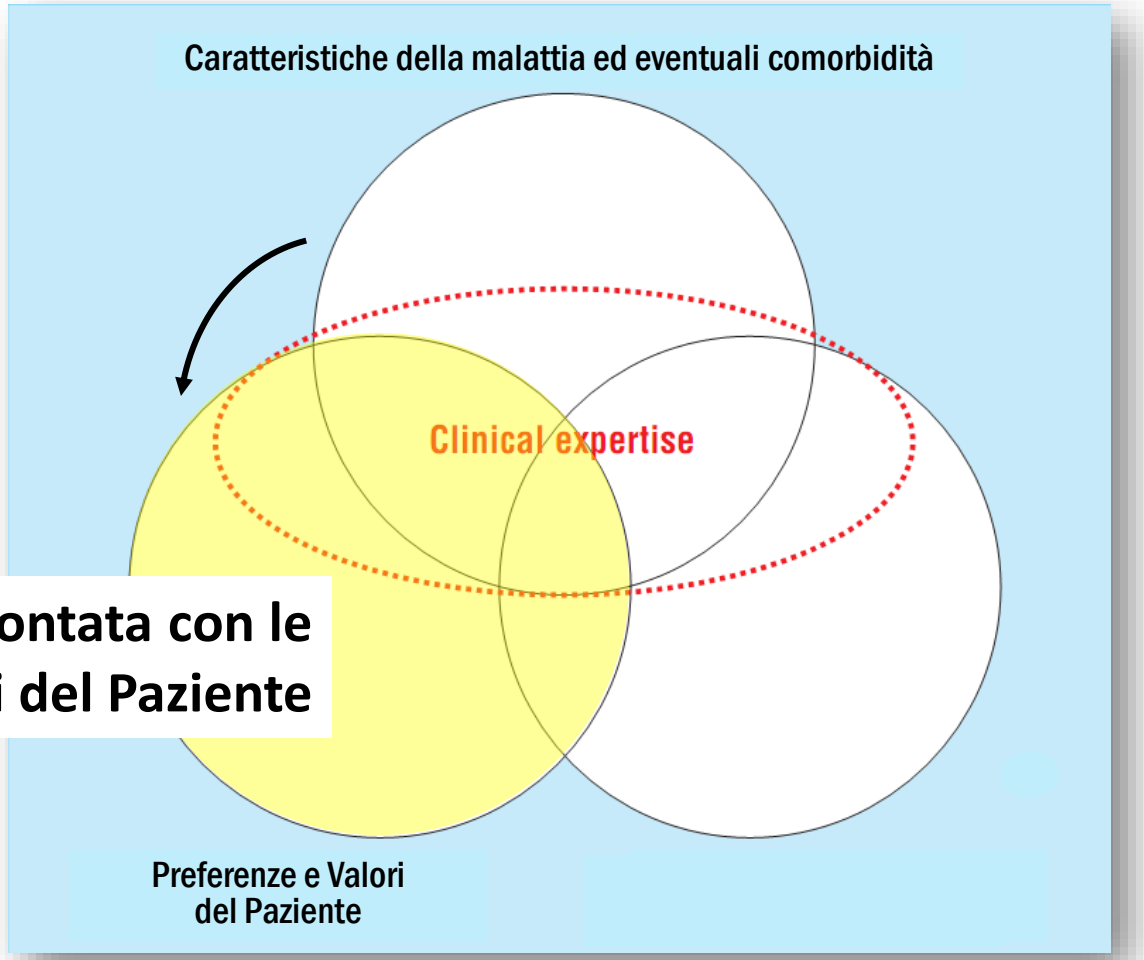


Physicians' and patients' choices in evidence based practice

Evidence does not make decisions, people do

R Brian Haynes P J Devereaux Gordon H Guyatt

BMJ 2002;324:1350



… e deve essere confrontata con le Preferenza e i Valori del Paziente

LEGGE 8 marzo 2017, n. 24

Disposizioni in materia di sicurezza delle cure e della persona assistita, nonche' in materia di responsabilita' professionale degli esercenti le professioni sanitarie. (17G00041) (GU Serie Generale n.64 del 17-03-2017)

Art. 5

Buone pratiche clinico-assistenziali e raccomandazioni previste dalle linee guida

1. Gli esercenti le professioni sanitarie, nell'esecuzione delle prestazioni sanitarie con finalita' preventive, diagnostiche, terapeutiche, palliative, riabilitative e di medicina legale, si attengono, salve le specificita' del caso concreto, alle raccomandazioni previste dalle linee guida pubblicate ai sensi del comma 3 ed elaborate da enti e istituzioni pubblici e privati nonche' dalle societa' scientifiche e dalle associazioni tecnico-scientifiche delle professioni sanitarie iscritte in apposito elenco istituito e regolamentato con decreto del Ministro della salute, da emanare entro novanta giorni dalla data di entrata in vigore della presente legge, e da aggiornare con cadenza biennale. In mancanza delle suddette raccomandazioni, gli esercenti le professioni sanitarie si attengono alle buone pratiche clinico-assistenziali.

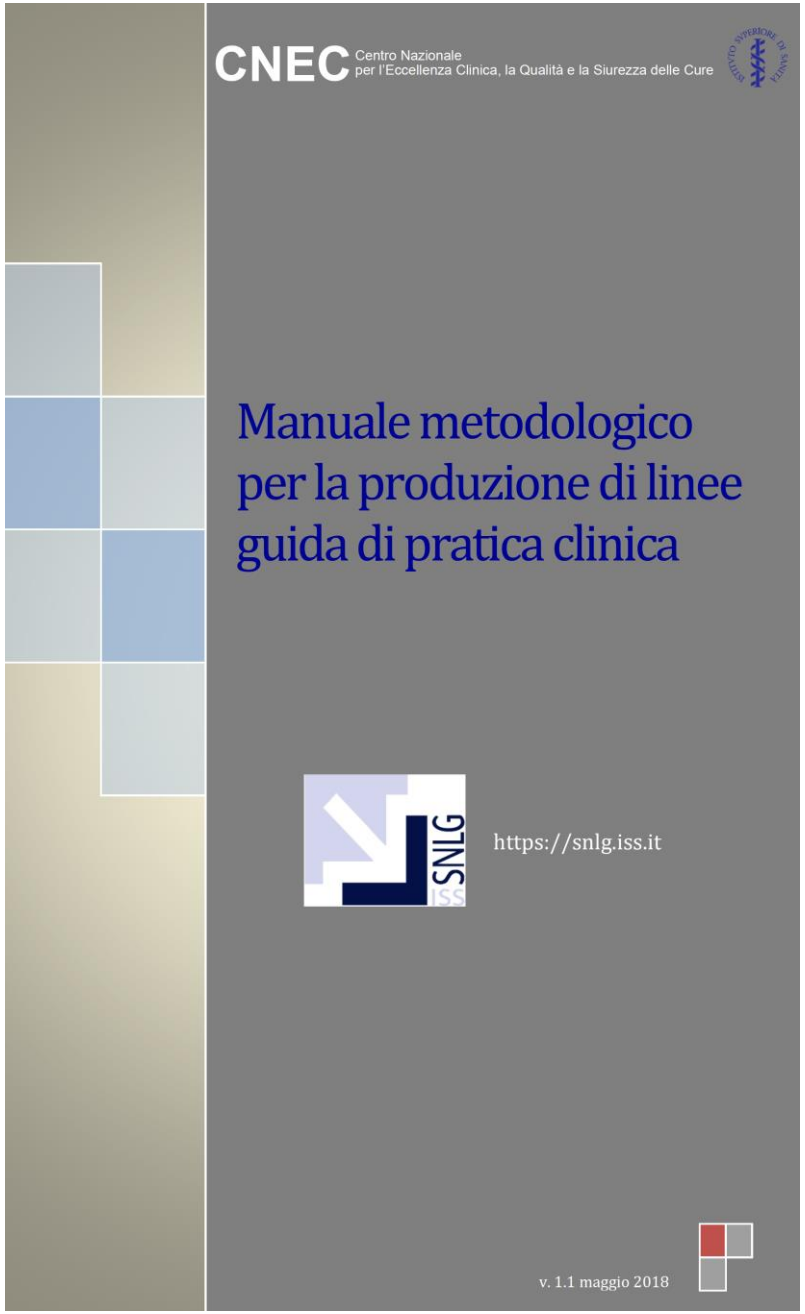
Obiettivo da perseguire:

determinare la "confidenza" sul fatto che i benefici ascrivibili al trattamento siano prevalenti sui danni (o viceversa)

Metodiche di produzione di Linee Guida a confronto

(G.L. Pappagallo)

Cominciamo con il GRADE…



Il metodo GRADE propone una valutazione della qualità delle prove più ampia e articolata di quella proposta da tutti gli altri sistemi di grading e rappresenta *lo standard metodologico di riferimento per la produzione di linee guida adottato sempre più diffusamente a livello internazionale.*

World Health Organization

GRADE

Working Group

Grades of **R**ecommendation **A**ssessment, **D**evelopment and **E**valuation

McMaster University
Inspiring Innovation and Discovery

- Aim: to develop a common, transparent and sensible system for grading the quality of evidence and the strength of recommendations (over 100 systems)
- International group of guideline developers, methodologists & clinicians from around the world (>200 contributors) – since 2000
- International group: WHO, ACCP, AHRQ, Australian NMRC, BMJ Clinical Evidence, CC, CDC, McMaster, NICE, Oxford CEBM, SIGN, UpToDate, USPSTF

CMAJ 2003, BMJ 2004, BMC 2004, BMC 2005, AJRCCM 2006, Chest 2006, BMJ 2008

Systems for grading the quality of evidence and the strength of recommendations I: Critical appraisal of existing approaches The GRADE Working Group

David Atkins[1], Martin Eccles[2], Signe Flottorp[3], Gordon H Guyatt[4], David Henry[5], Suzanne Hill[5], Alessandro Liberati[6], Dianne O'Connell[7], Andrew D Oxman[3], Bob Phillips[8], Holger Schünemann[4,9], Tessa Tan-Torres Edejer[10], Gunn E Vist*[3], John W Williams Jr[11] and The GRADE Working Group[3]

BMC Health Services Research 2004, **4**:38

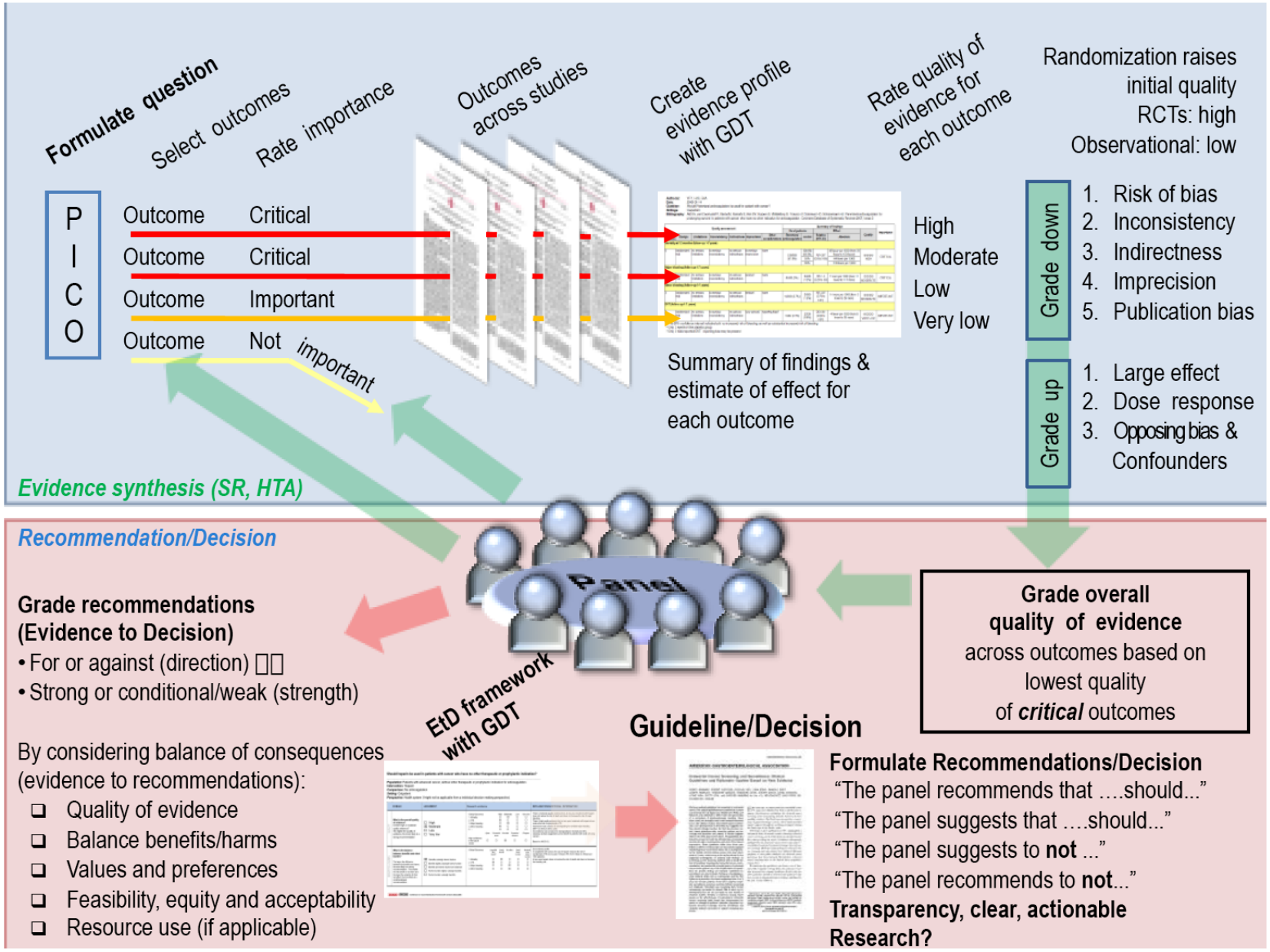
Table 1: Summary of assessments of the sensibility of six approaches to rating levels of evidence and strength of recommendation

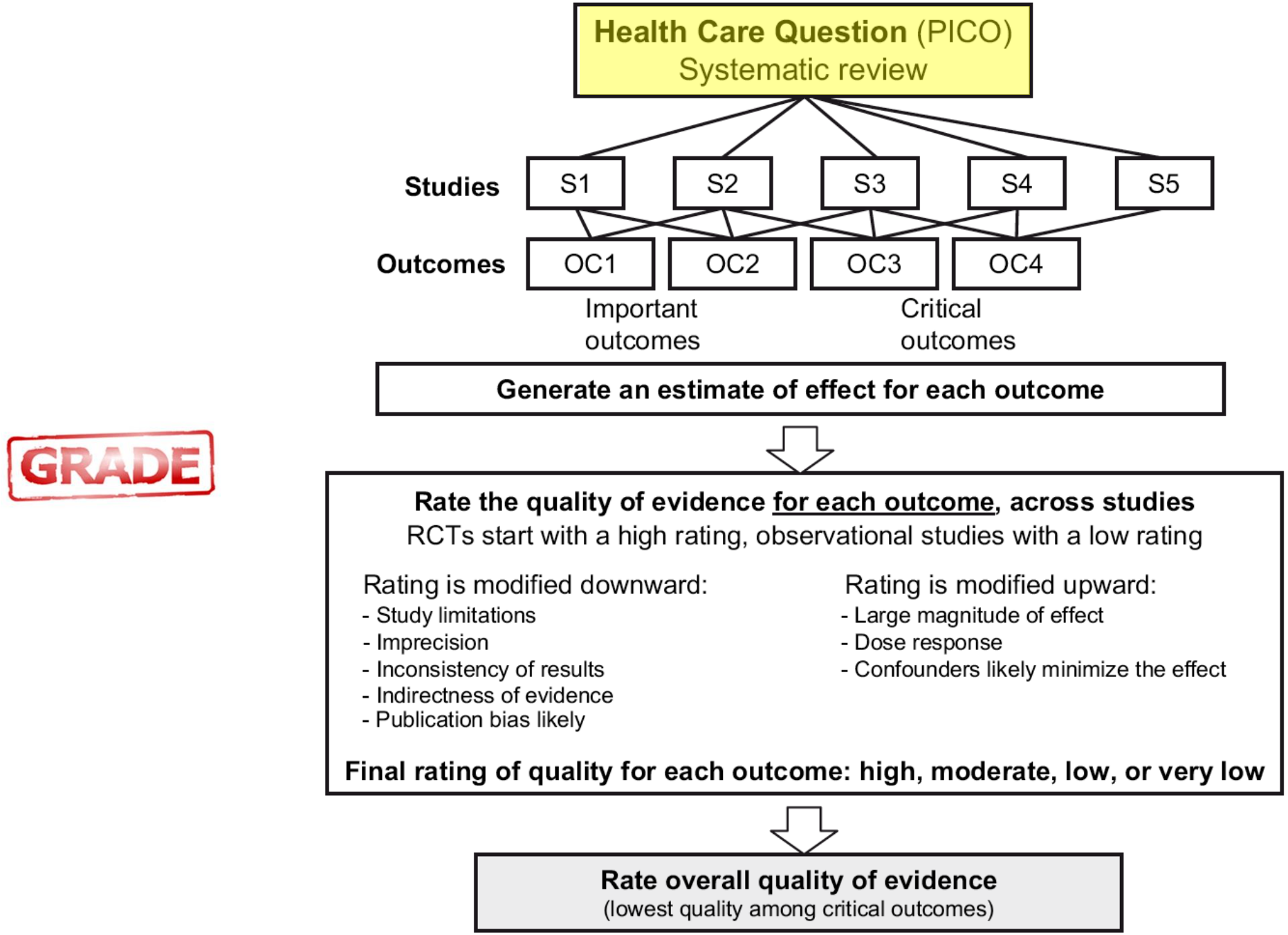
Criteria[1]	ACCP			ANHMRC[2]			USTFCPS			OCEBM			SIGN			USPSTF[3]		
	No		Yes	No		Yes	No		Yes	No		Yes	No		Yes	No		Yes
1. Applicable to different questions:																		
Effectiveness			12		2	8		1	11			12	1		11		2	9
Harm		1	11		5	5	1	7	4		1	11	1	3	8	2	2	7
Diagnosis	7	3	2	4	4	2	9	3				12	5	2	5	2	2	7
Prognosis	6	3	3	2	5	3	9	2	1			11	4	3	5	3	3	5
2. Can be used by:																		
Professionals		1	11	1	5	3		7	4	1	6	5		5	7		3	8
Policy makers	1	5	6	1	5	3	1	2	9	3	7	2	2	6	4	1	4	6
Patients	4	5	3	5	5		6	3	3	9	3		7	5		4	6	1
3. Clear and simple	1	5	6	2	6	1	2	8	2	2	4	5	1	5	6	1	4	7
4. Information not available		8	4	1	5	3	1	6	5		4	8	1	7	4	1	9	2
5. Subjective decisions	2	10		5	2	2	5	5	2		7	5	5	7		2	9	
6. Inappropriate dimensions	1	3	8		1	6	2	4	6		1	10	1	2	8	1	4	6
7. Missing dimensions	1	6	5	2	2	4	5	4	3	9	3	1	5	4	3	2	5	4
Aggregation of dimensions:																		
8. Clear and simple	1	5	6	4	1	2	2	2	7	3	4	4		6	6	2	7	2
9. Appropriate		6	5	3	1	1	3	4	4	2	5	4	1	4	6	1	5	5
10. Sufficient categories	1	4	6	4	2	1		5	7	2	2	7	1	2	9		1	10
11. Likely to discriminate		7	5	2	5	1	1	9	2	2	4	6		5	7		4	7
12. Assessments reproducible	1	8	3	4	4		2	7	2		7	4	1	8	2		10	

Based on discussions of the strengths and limitations of current approaches to grading levels of evidence and the strength of recommendations, we agreed to develop an approach that addresses the major limitations that we identified.

The GRADE approach

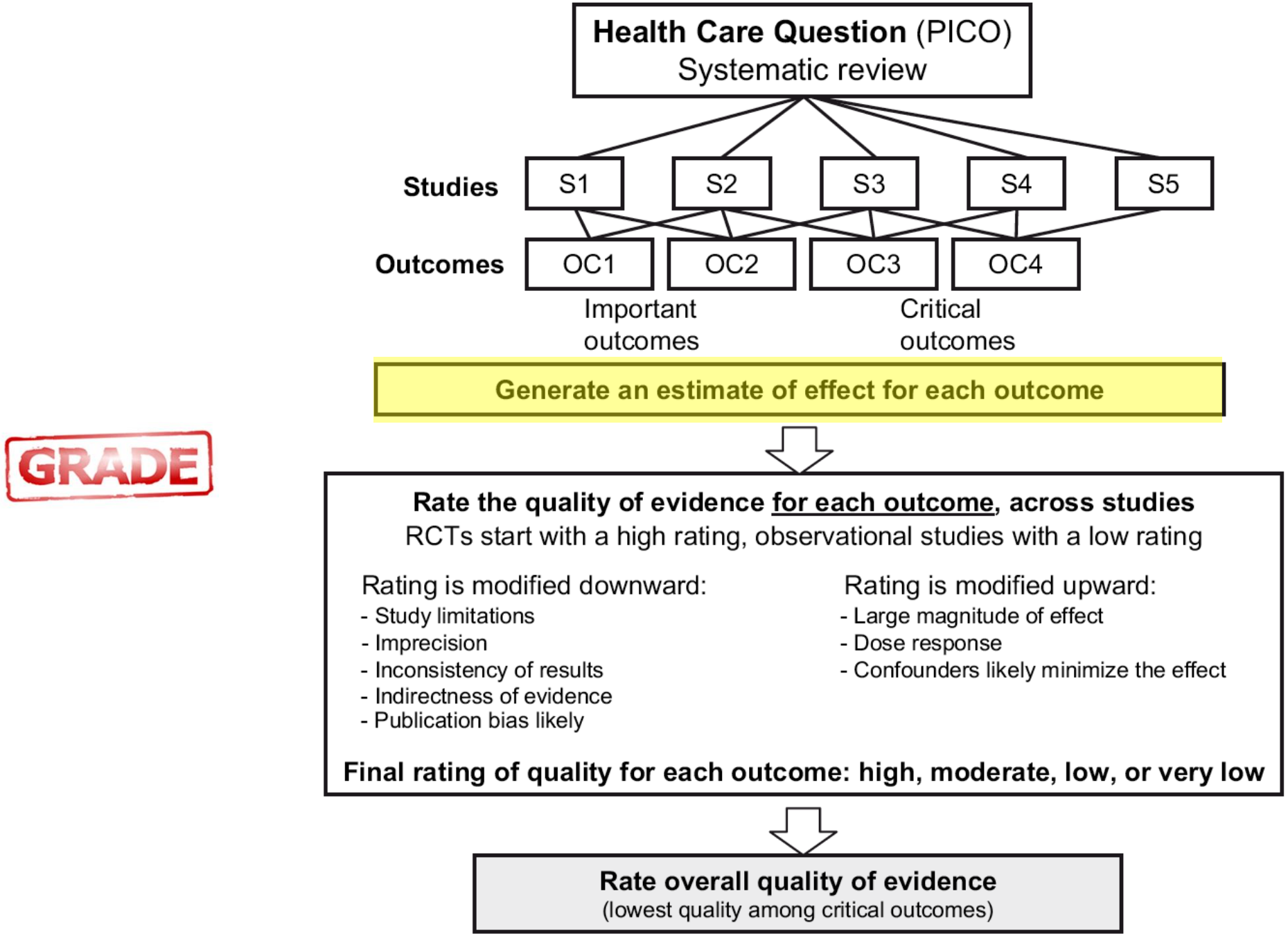
- Considers
  - the evidence for each outcome in the review separately
  - magnitude of the effect
  - all factors to determine how confident we are in the results – quality of evidence
- Ensures
  - systematic process
  - transparency

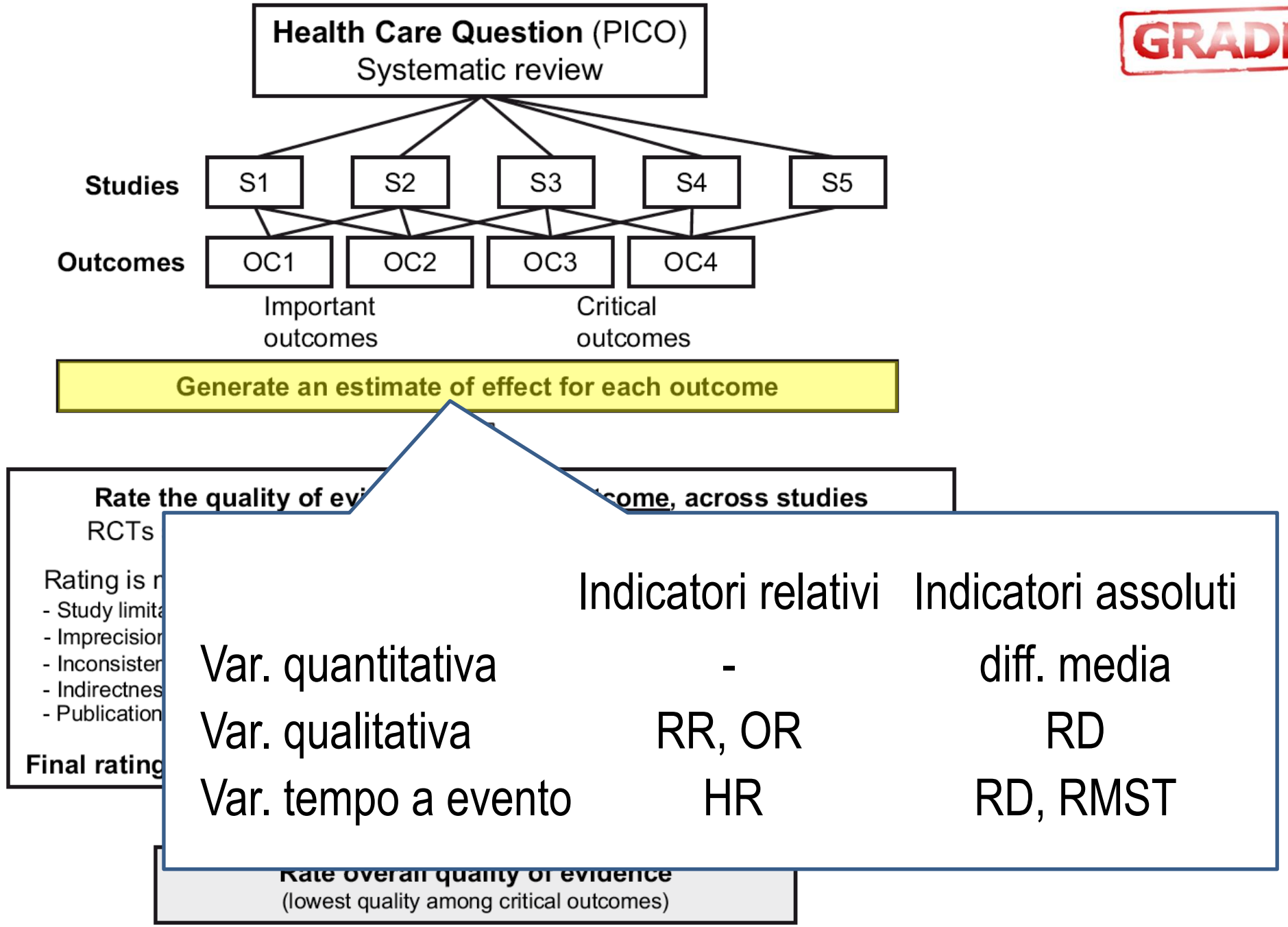

Formulate question
Select outcomes
Rate importance
Outcomes across studies
Create evidence profile with GDT
Rate quality of evidence for each outcome
Randomization raises initial quality
RCTs: high
Observational: low
P
I
C
O
Outcome Critical
Outcome Critical
Outcome Important
Outcome Not important
Summary of findings & estimate of effect for each outcome
High
Moderate
Low
Very low
Grade down
1. Risk of bias
2. Inconsistency
3. Indirectness
4. Imprecision
5. Publication bias
Grade up
1. Large effect
2. Dose response
3. Opposing bias & Confounders
Evidence synthesis (SR, HTA)
Recommendation/Decision
Panel
Grade overall quality of evidence across outcomes based on lowest quality of critical outcomes
Grade recommendations (Evidence to Decision)
• For or against (direction)
• Strong or conditional/weak (strength)
By considering balance of consequences (evidence to recommendations):
Quality of evidence
Balance benefits/harms
Values and preferences
Feasibility, equity and acceptability
Resource use (if applicable)
EtD framework with GDT
Guideline/Decision
Formulate Recommendations/Decision
"The panel recommends that ….should..."
"The panel suggests that ….should..."
"The panel suggests to not ..."
"The panel recommends to not..."
Transparency, clear, actionable
Research?


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
GRADE
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)

The GRADE approach

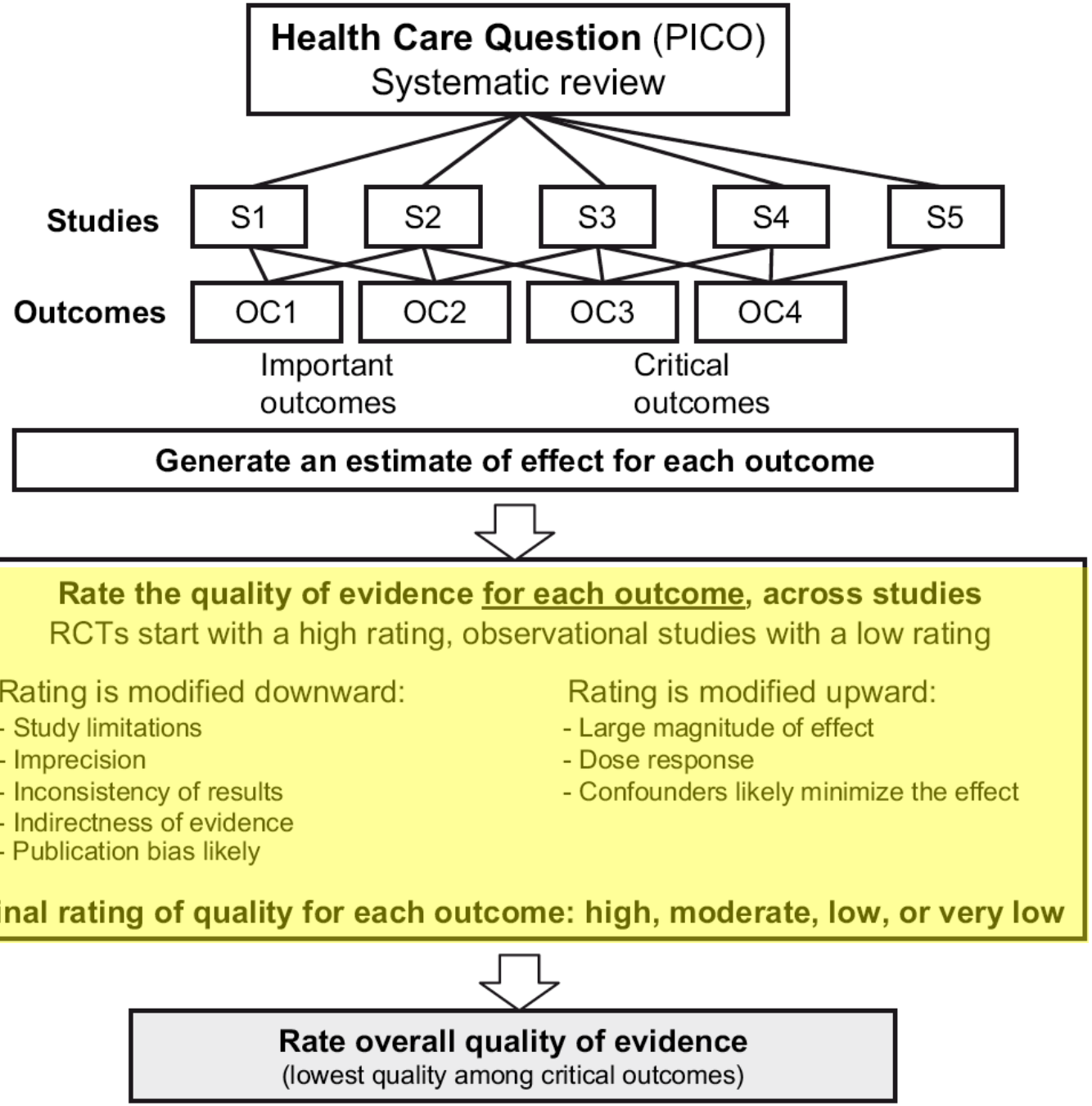
- Considers
  - the evidence for each outcome in the review separately
  - magnitude of the effect
  - all factors to determine how confident we are in the results – quality of evidence
- Ensures
  - systematic process
  - transparency


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
GRADE
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 evi
come, across studies
RCTs
Rating is r
- Study limita
- Imprecisior
- Inconsister
- Indirectnes
- Publication
Final rating
Rate overall quality of evidence
(lowest quality among critical outcomes)
Indicatori relativi
Indicatori assoluti
Var. quantitativa
-
diff. media
Var. qualitativa
RR, OR
RD
Var. tempo a evento
HR
RD, RMST

The GRADE approach

- Considers
  - the evidence for each outcome in the review separately
  - magnitude of the effect
  - all factors to determine how confident we are in the results – quality of evidence
- Ensures
  - systematic process
  - transparency


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
GRADE
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)

E ora un cenno alle altre metodiche…

National Comprehensive Cancer Network

Your Best Resource in the Fight Against Cancer®

NCCN Categories of Evidence and Consensus

Category 1: The recommendation is based on high-level evidence (e.g. randomized controlled trials) and there is uniform NCCN consensus.

Category 2A: The recommendation is based on lower-level evidence and there is uniform NCCN consensus.

Category 2B: The recommendation is based on lower-level evidence and there is nonuniform NCCN consensus (but no major disagreement).

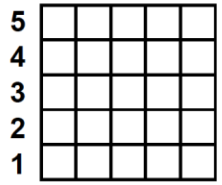
Category 3: The recommendation is based on any level of evidence but reflects major disagreement.

All recommendations are category 2A unless otherwise noted.

- Non quesito clinico specifico
- Non revisione sistematica (sufficiente il rimborso delle assicurazioni)
- Non chiara definizione della certezza delle prove
- Importanza eccessiva del consenso
- Non chiaro riferimento al rapporto tra benefici e danni

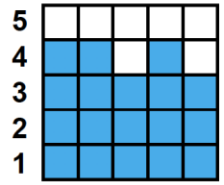
NCCN Evidence Blocks™

NCCN EVIDENCE BLOCKS CATEGORIES AND DEFINITIONS



E = Efficacy of Regimen/Agent
S = Safety of Regimen/Agent
Q = Quality of Evidence
C = Consistency of Evidence
A = Affordability of Regimen/Agent

Example Evidence Block



E = 4
S = 4
Q = 3
C = 4
A = 3

Efficacy of Regimen/Agent

5	Survival
4	
3	
2	
1	

Quality of Evidence

5	Study Design
4	
3	
2	
1	

Safety of Regimen/Agent

5	Interference with ADLs
4	
3	
2	
1	

Note: For significant chronic or long-term toxicities, score decreased by 1

Consistency of Evidence

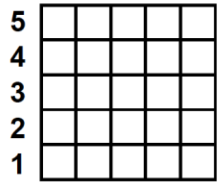
5	Consistency
4	
3	
2	
1	

Affordability of Regimen/Agent (includes drug cost, supportive care, infusions, toxicity monitoring, management of toxicity)

5	Costs
4	
3	
2	
1	

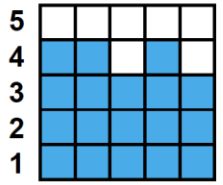
NCCN Evidence Blocks™

NCCN EVIDENCE BLOCKS CATEGORIES AND DEFINITIONS



E = Efficacy of Regimen/Agent
S = Safety of Regimen/Agent
Q = Quality of Evidence
C = Consistency of Evidence
A = Affordability of Regimen/Agent

Example Evidence Block



E = 4
S = 4
Q = 3
C = 4
A = 3

Efficacy of Regimen/Agent

5	**Highly effective:** Cure likely and often provides long-term survival advantage
4	**Very effective:** Cure unlikely but sometimes provides long-term survival advantage
3	**Moderately effective:** Modest impact on survival, but often provides control of disease
2	**Minimally effective:** No, or unknown impact on survival, but sometimes provides control of disease
1	**Palliative:** Provides symptomatic benefit only

Safety of Regimen/Agent

5	**Usually no meaningful toxicity:** Uncommon or minimal toxicities; no interference with activities of daily living (ADLs)
4	**Occasionally toxic:** Rare significant toxicities or low-grade toxicities only; little interference with ADLs
3	**Mildly toxic:** Mild toxicity that interferes with ADLs
2	**Moderately toxic:** Significant toxicities often occur but life threatening/fatal toxicity is uncommon; interference with ADLs is frequent
1	**Highly toxic:** Significant toxicities or life threatening/fatal toxicity occurs often; interference with ADLs is usual and severe

Note: For significant chronic or long-term toxicities, score decreased by 1

Quality of Evidence

5	**High quality:** Multiple well-designed randomized trials and/or meta-analyses
4	**Good quality:** One or more well-designed randomized trials
3	**Average quality:** Low quality randomized trial(s) or well-designed non-randomized trial(s)
2	**Low quality:** Case reports or extensive clinical experience
1	**Poor quality:** Little or no evidence

Consistency of Evidence

5	**Highly consistent:** Multiple trials with similar outcomes
4	**Mainly consistent:** Multiple trials with some variability in outcome
3	**May be consistent:** Few trials or only trials with few patients, whether randomized or not, with some variability in outcome
2	**Inconsistent:** Meaningful differences in direction of outcome between quality trials
1	**Anecdotal evidence only:** Evidence in humans based upon anecdotal experience

Affordability of Regimen/Agent (includes drug cost, supportive care, infusions, toxicity monitoring, management of toxicity)

5	**Very inexpensive**
4	**Inexpensive**
3	**Moderately expensive**
2	**Expensive**
1	**Very expensive**

Oxford Centre for Evidence-based Medicine Levels of Evidence (March 2009)

Level	Therapy/Prevention, Aetiology/Harm
1a	SR (with homogeneity*) of RCTs
1b	Individual RCT (with narrow Confidence Interval‡)
1c	All or none§
2a	SR (with homogeneity*) of cohort studies
2b	Individual cohort study (including low quality RCT; e.g., <80% follow-up)
2c	"Outcomes" Research; Ecological studies
3a	SR (with homogeneity*) of case-control studies
3b	Individual Case-Control Study
4	Case-series (and poor quality cohort and case-control studies§§)
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"

Grades of Recommendation

A	consistent level 1 studies
B	consistent level 2 or 3 studies or extrapolations from level 1 studies
C	level 4 studies or extrapolations from level 2 or 3 studies
D	level 5 evidence or troublingly inconsistent or inconclusive studies of any level

- Prevista revisione sistematica
- Certezza delle prove chiaramente definita (rischio di bias, imprecisione, eterogeneità)
- Non riferimento al rapporto tra benefici e danni
- Raccomandazioni espresse in base alla sola certezza delle prove (possibili equivoci)

GOOD SCIENCE
BETTER MEDICINE
BEST PRACTICE

SOPs/Instructions for Authors and templates for standard ESMO Clinical Practice Guidelines (CPGs), eUpdates and ESMO-MCBS Scores

ESMO Guidelines Committee, October 2015, latest revision October 2016

Levels of evidence

I	Evidence from at least one large randomised, controlled trial of good methodological quality (low potential for bias) or meta-analyses of well-conducted randomised trials without heterogeneity
II	Small randomised trials or large randomised trials with a suspicion of bias (lower methodological quality) or meta-analyses of such trials or of trials with demonstrated heterogeneity
III	Prospective cohort studies
IV	Retrospective cohort studies or case–control studies
V	Studies without control group, case reports, experts opinions

- Non quesito clinico specifico
- Non revisione sistematica
- Non chiara definizione della certezza delle prove
- Non riferimento al rapporto tra benefici e danni

A standardised, generic, validated approach to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies: the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS)

N. I. Cherny[1*], R. Sullivan[2], U. Dafni[3], J. M. Kerst[4], A. Sobrero[5], C. Zielinski[6], E. G. E. de Vries[7] & M. J. Piccart[8,9]

Annals of Oncology 26: 1547–1573, 2015

Table 1. Potential benefits of a new treatment

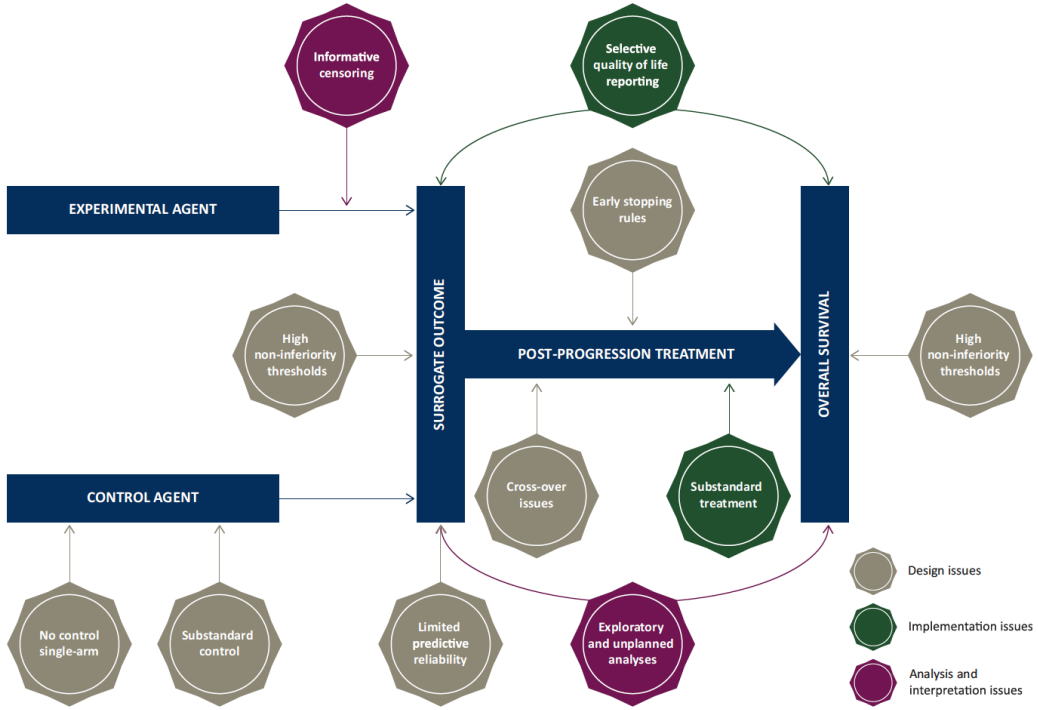
Living longer
Improved OS
Improved surrogate of OS
DFS (when OS data are immature in adjuvant setting)
Improved PFS
Living better
Improved quality of life
Improved surrogate of quality of life
Improved PFS
Reduced toxicity

- Beneficio riferito al braccio di controllo dello studio e NON allo *standard of care*
- Prevalenza della statistica (*P-value*) sulla clinica (*absolute benefit*)

Biases in study design, implementation, and data analysis that distort the appraisal of clinical benefit and ESMO-Magnitude of Clinical Benefit Scale (ESMO-MCBS) scoring

B. Gyawali[1,2,3*], E. G. E. de Vries[4], U. Dafni[5,6], T. Amaral[7], J. Barriuso[8], J. Bogaerts[9], A. Calles[10], G. Curigliano[11,12], C. Gomez-Roca[13], B. Kiesewetter[14], S. Oosting[4], A. Passaro[15], G. Pentheroudakis[16], M. Piccart[17], F. Roitberg[18,19], J. Tabernero[20], N. Tarazona[21], D. Trapani[12], R. Wester[22], G. Zarkavelis[23], C. Zielinski[24], P. Zygoura[6] & N. I. Cherny[25]

https://doi.org/10.1016/j.esmoop.2021.100117



Conclusion: Interpretation of the ESMO-MCBS scores requires critical appraisal of trials to understand caveats in trial design, implementation, and data analysis that may have biased results and conclusions. These will be addressed in future iterations of the ESMO-MCBS.

European Association of Urology

Level	Type of evidence
1a	Evidence obtained from meta-analysis of randomised trials
1b	Evidence obtained from at least one randomised trial
2a	Evidence obtained from one well-designed controlled study without randomisation
2b	Evidence obtained from at least one other type of well-designed quasi-experimental study
3	Evidence obtained from well-designed non-experimental studies, such as comparative studies, correlation studies and case reports
4	Evidence obtained from expert committee reports or opinions or clinical experience of respected authorities

From 2018 onwards, the EAU Guidelines have been using a **modified GRADE approach** for the grading of recommendations.

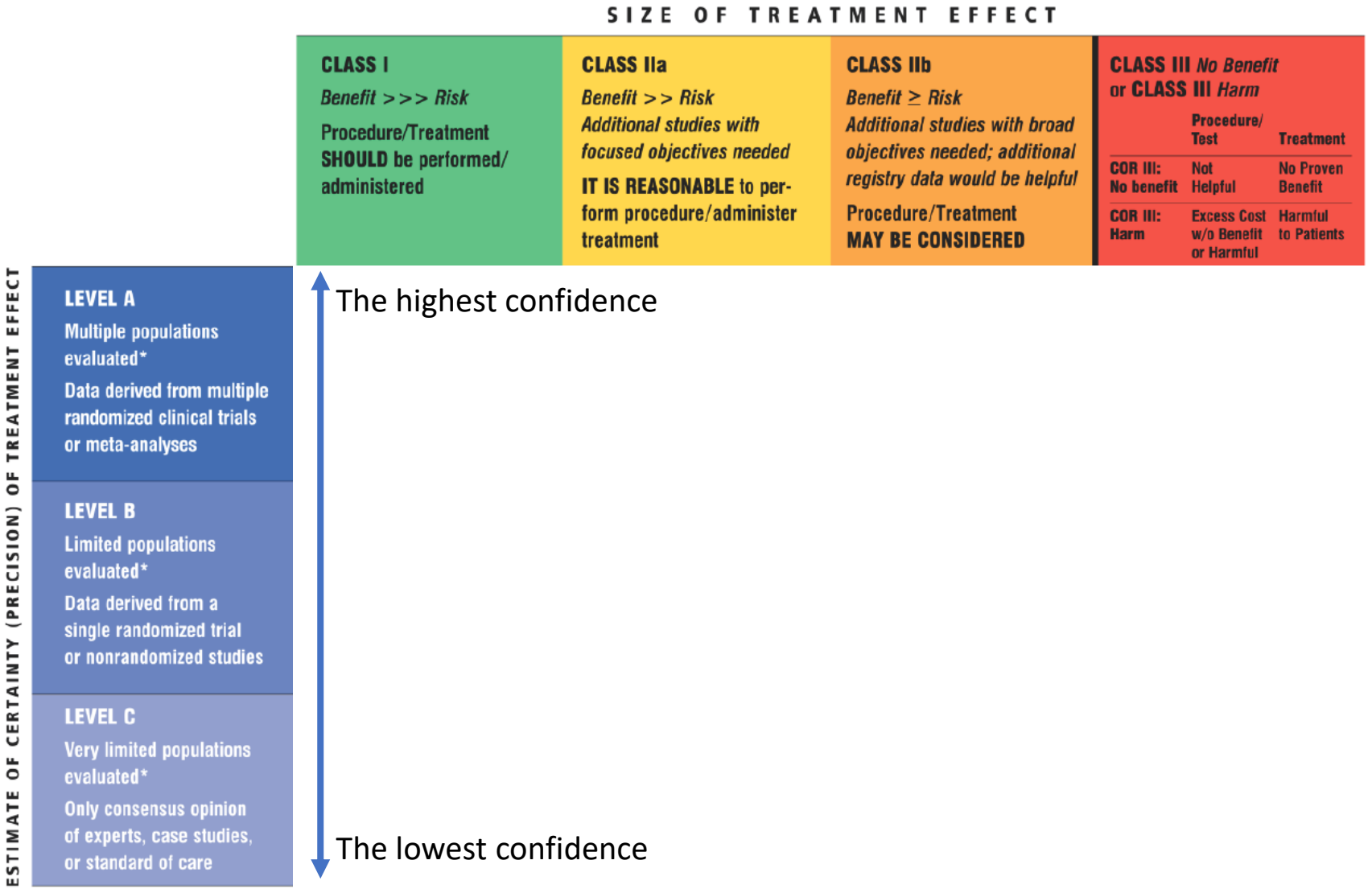
The strength of each recommendation is represented by the words 'strong' or 'weak'.

European Association of Urology. Guidelines Office development handbook. Arnhem, The Netherlands; EAU; 2022. https://uroweb.org/eau-guidelines/methodology-policies.

- Quesito clinico P.I.C.O?
- Revisione sistematica?
- Certezza delle prove chiaramente definita (*v. Oxford CEBM*)
- Riferimento al rapporto tra benefici e danni?
- *Summary of Evidence* (SOE) *tables* non disponibili

G.L. Pappagallo, 2023

2013 ACC/AHA Cardiovascular Risk Guideline



2013 ACC/AHA Cardiovascular Risk Guideline

SIZE OF TREATMENT EFFECT

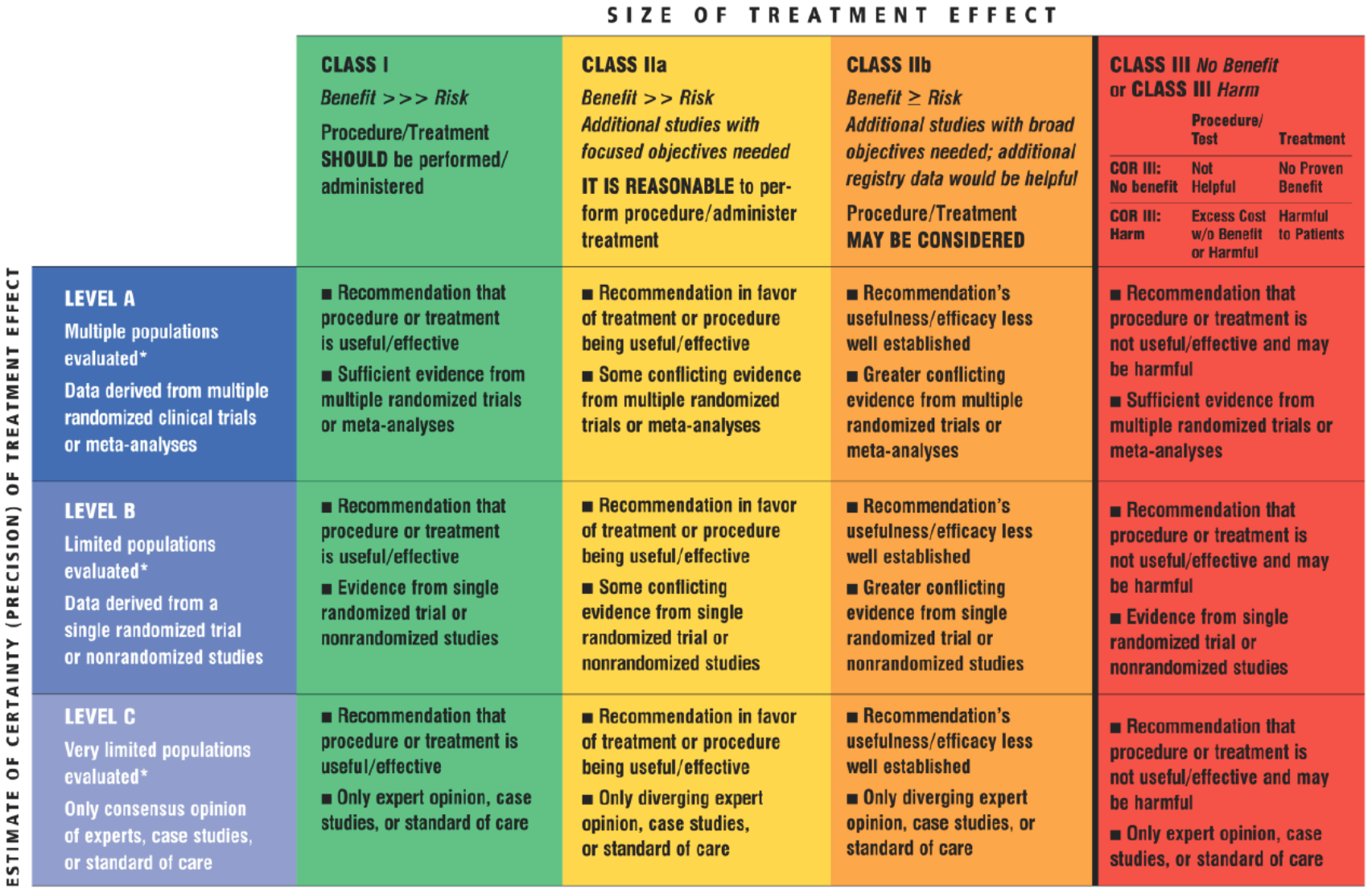
CLASS I	CLASS IIa	CLASS IIb	CLASS III *No Benefit* or CLASS III *Harm*
Benefit >>> Risk	*Benefit >> Risk*	*Benefit ≥ Risk*	
Procedure/Treatment **SHOULD** be performed/ administered	*Additional studies with focused objectives needed*	*Additional studies with broad objectives needed; additional registry data would be helpful*	
	IT IS REASONABLE to perform procedure/administer treatment	Procedure/Treatment **MAY BE CONSIDERED**	

	Procedure/ Test	Treatment
COR III: No benefit	Not Helpful	No Proven Benefit
COR III: Harm	Excess Cost w/o Benefit or Harmful	Harmful to Patients

Benefit outweights harm

Harm outweights benefit

2013 ACC/AHA Cardiovascular Risk Guideline



ASCO® Guidelines

Methodology Manual

The quality and usability of ASCO's guidelines is enhanced by transparency about the quality and strength of evidence that informs guideline recommendations. ASCO adopted the GRADE Methodology as a recognized standard in guideline development methodology.

History:

Approved by the Society Board of Directors on May 30, 2019

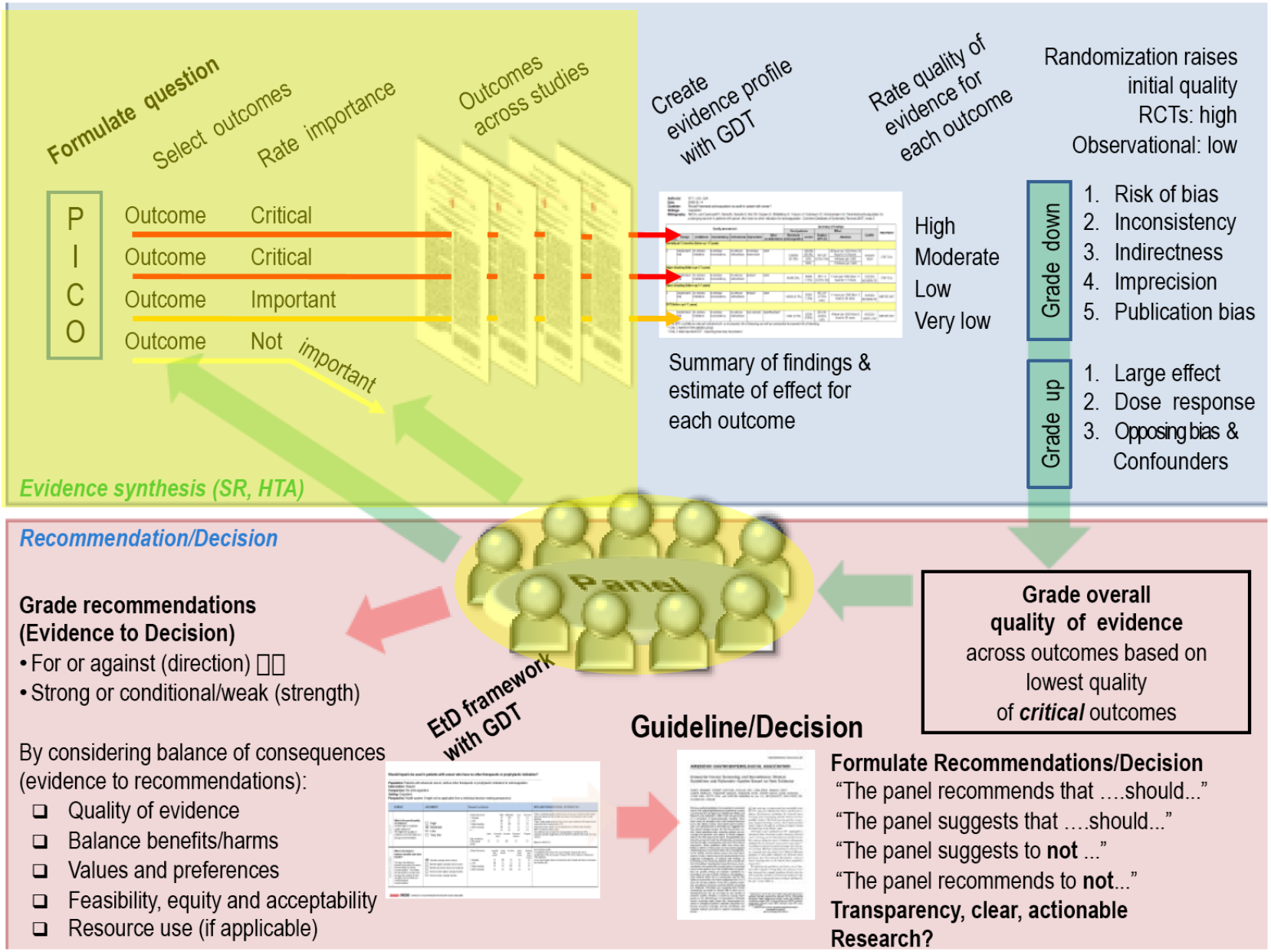
Amended and approved by the Society Executive Committee on December 19, 2019, July 23, 2020, September 9, 2021, November 19, 2021, September 16, 2022, and December 18, 2024

GRADE

Dal quesito clinico alla raccolta sistematica delle prove

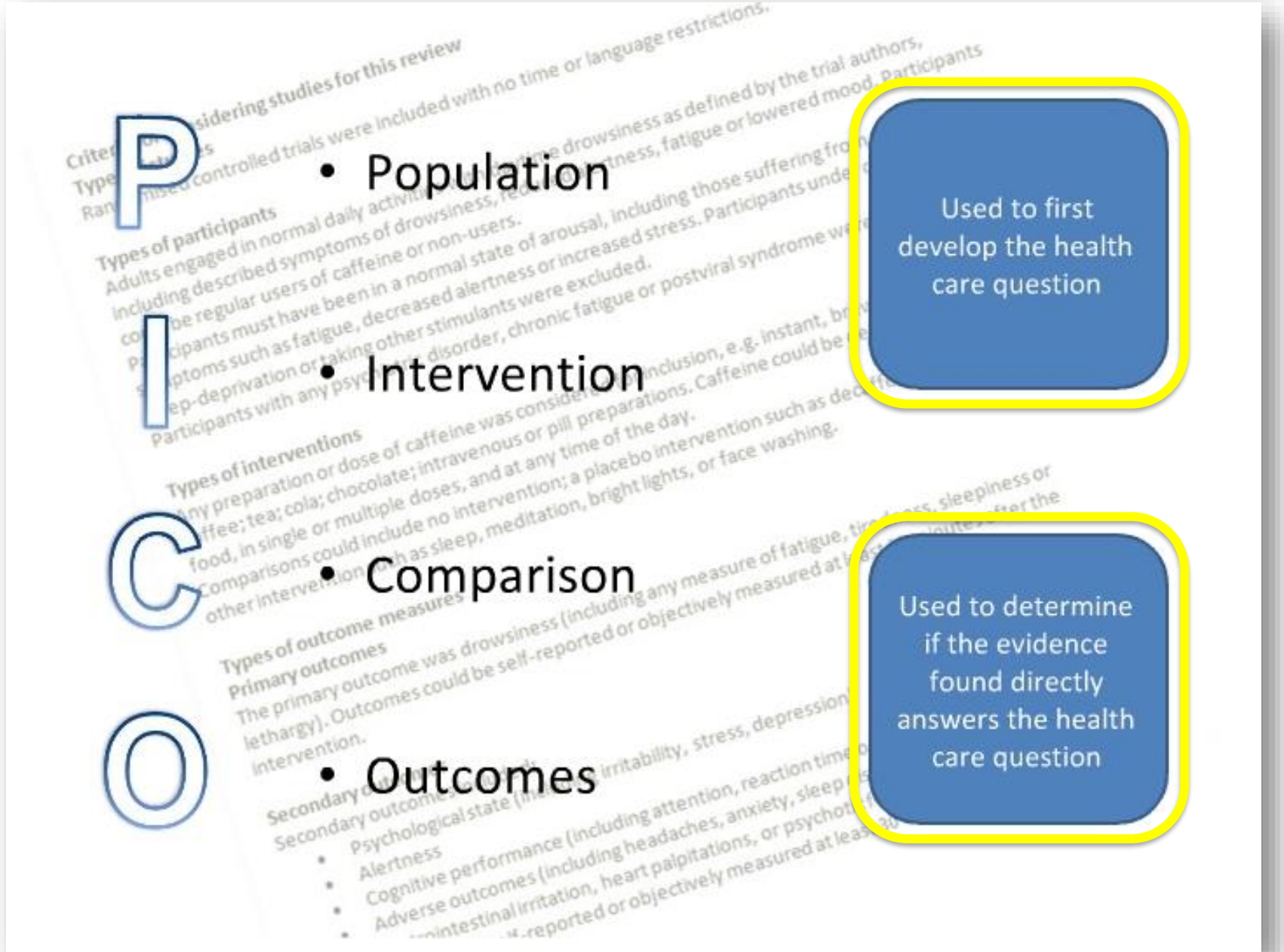
(Giovanni Pappagallo)

The GRADE process in developing guidelines



Important Questions

Should be
from practice
NOT
evidence driven



- **P** Population
- **I** Intervention
- **C** Comparison
- **O** Outcomes

Used to first develop the health care question

Used to determine if the evidence found directly answers the health care question

GRADE

Strutturazione del Quesito Clinico sec. modello P.I.C.O.

P	Nei **P**azienti con…	Specifiche caratteristiche di malattia (stadio, classe di rischio, ecc.)
I	l'**I**ntervento…	Intervento terapeutico oggetto del quesito clinico
C	(è suscettibile di impiego) in **C**onfronto con…	Trattamento altrimenti considerabile in alternativa all'intervento in esame
O	riguardo agli **O**utcome di beneficio/danno…	Parametri clinico-laboratoristici ritenuti essenziali per la proposta terapeutica

Strutturazione del Quesito Clinico sec. modello P.I.C.O.

P	Nei **P**azienti con…	Specifiche caratteristiche di malattia (stadio, classe di rischio, ecc.)
I	l'**I**ntervento…	Intervento terapeutico oggetto del quesito clinico
C	(è suscettibile di impiego) in **C**onfronto con…	Trattamento altrimenti considerabile in alternativa all'intervento in esame
O	riguardo agli **O**utcome di beneficio/danno…	Parametri clinico-laboratoristici ritenuti essenziali per la proposta terapeutica

Strutturazione del Quesito Clinico sec. modello P.I.C.O.

P	Nei **P**azienti con…	Specifiche caratteristiche di malattia (stadio, classe di rischio, ecc.)
I	l'**I**ntervento…	Intervento terapeutico oggetto del quesito clinico
C	(è suscettibile di impiego) in **C**onfronto con…	Trattamento altrimenti considerabile in alternativa all'intervento in esame
O	riguardo agli **O**utcome di beneficio/danno…	Parametri clinico-laboratoristici ritenuti essenziali per la proposta terapeutica

Strutturazione del Quesito Clinico sec. modello P.I.C.O.

P	Nei **P**azienti con…	Specifiche caratteristiche di malattia (stadio, classe di rischio, ecc.)
I	l'**I**ntervento…	Intervento terapeutico oggetto del quesito clinico
C	(è suscettibile di impiego) in **C**onfronto con…	Trattamento altrimenti considerabile in alternativa all'intervento in esame
O	riguardo agli **O**utcome di beneficio/danno…	Parametri clinico-laboratoristici ritenuti essenziali per la proposta terapeutica

Outcomes

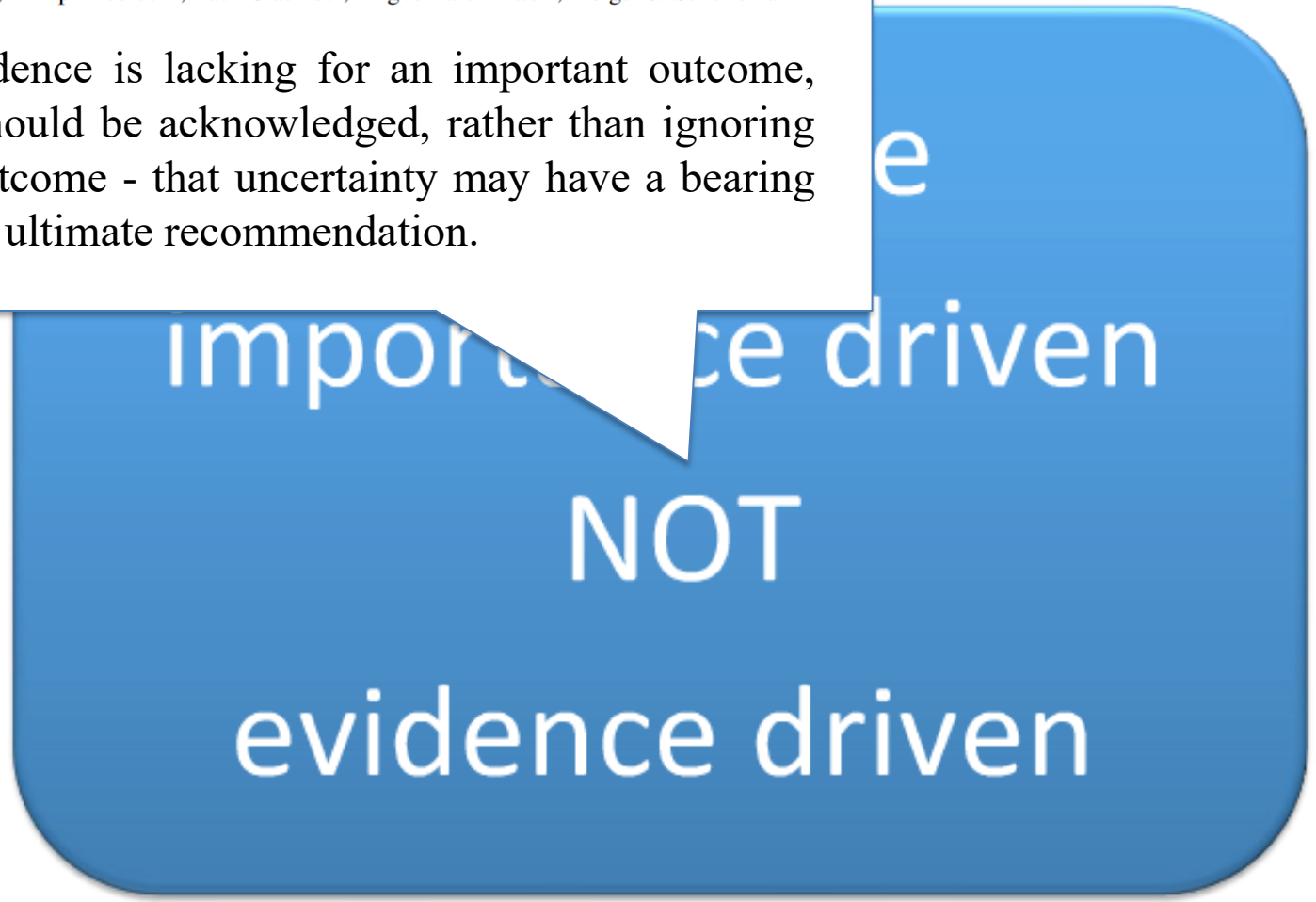
Should be
importance driven
NOT
evidence driven

Journal of Clinical Epidemiology 64 (2011) 395–400

GRADE guidelines: 2. Framing the question and deciding on important outcomes

Gordon H. Guyatt[a,*], Andrew D. Oxman[b], Regina Kunz[c], David Atkins[d], Jan Brozek[a], Gunn Vist[b], Philip Alderson[e], Paul Glasziou[f], Yngve Falck-Ytter[g], Holger J. Schünemann[a]

If evidence is lacking for an important outcome, this should be acknowledged, rather than ignoring the outcome - that uncertainty may have a bearing on the ultimate recommendation.



Choosing outcomes

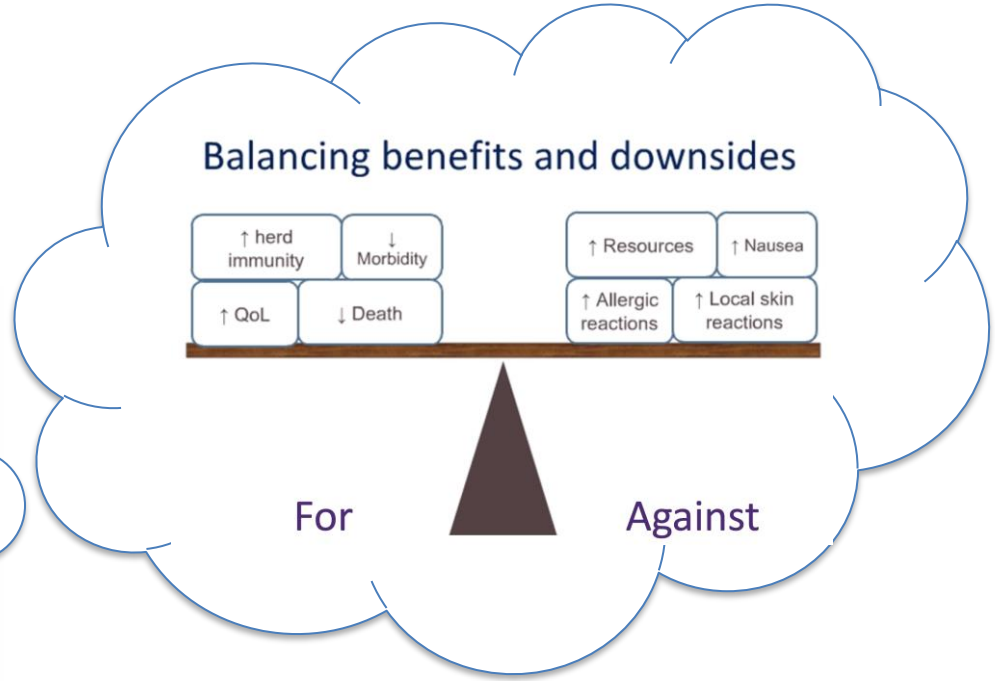
Desirable outcomes

- lower mortality
- reduced hospital stay
- reduced duration of disease
- reduced resource expenditure

Undesirable outcomes

- adverse reactions
- the development of resistance
- costs of treatment

Recommendations must consider desirable and undesirable outcomes



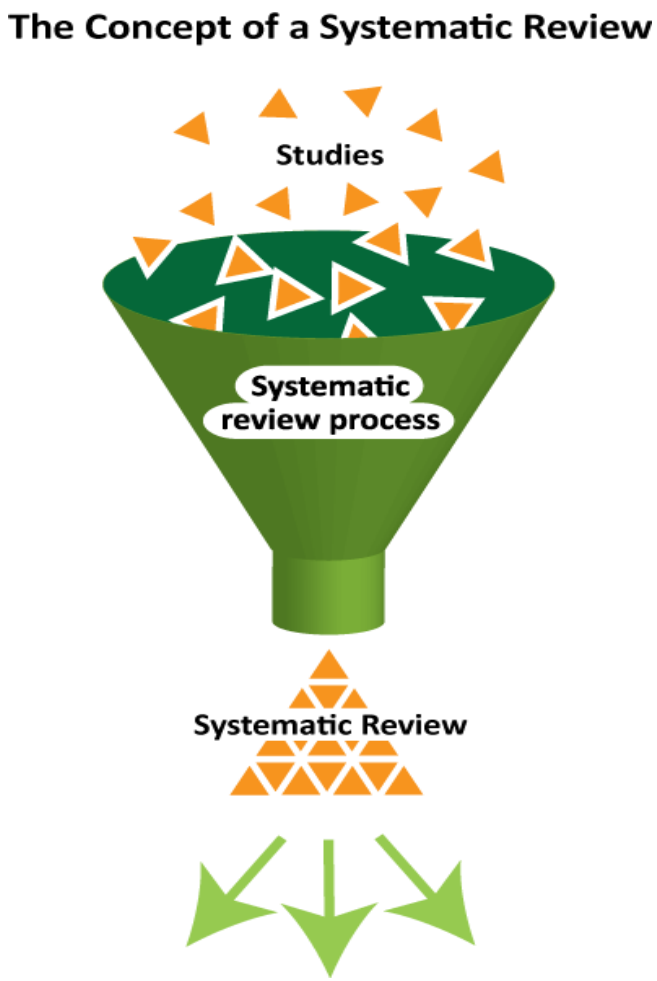
Valutazione dell'importanza degli *outcome* mediante votazione individuale dei panelisti, utilizzando una scala a 9 punti e assegnando l'outcome a una delle tre categorie sulla base del punteggio mediano ottenuto.

Classificazione degli *outcome* proposta dal metodo GRADE

Rating **(mediana del voto)**	**Importanza**	**Incluso in**
7 8 9	*outcome* importanti ed essenziali	tabelle sulla qualità delle prove: SÌ raccomandazione: SÌ
4 5 6	*outcome* importanti ma non essenziali	tabelle sulla qualità delle prove: SÌ raccomandazione: NO
1 2 3	*outcome* non importanti	tabelle sulla qualità delle prove: NO raccomandazione: NO

Revisione Sistematica

Metodo esplicito e trasparente per identificare, valutare e riassumere i risultati di singoli studi sugli effetti di un intervento sanitario.



Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation

BMJ 2015;349:g7647

Larissa Shamseer[1], David Moher[1], Mike Clarke[2], Davina Ghersi[3], Alessandro Liberati (deceased)[4], Mark Petticrew[5], Paul Shekelle[6], Lesley A Stewart[7], the PRISMA-P Group

[1]Ottawa Hospital Research Institute and University of Ottawa, Canada; [2]Queen's University Belfast, Ireland; [3]National Health and Medical Research Council, Australia; [4]University of Modena, Italy; [5]London School of Hygiene and Tropical Medicine, UK; [6]Southern California Evidence-based Practice Center, USA; [7]Centre for Reviews and Dissemination, University of York, UK

An international group of experts has created a guideline to improve the transparency, accuracy, completeness, and frequency of documented systematic review and meta-analysis protocols—PRISMA-P (for protocols) 2015.

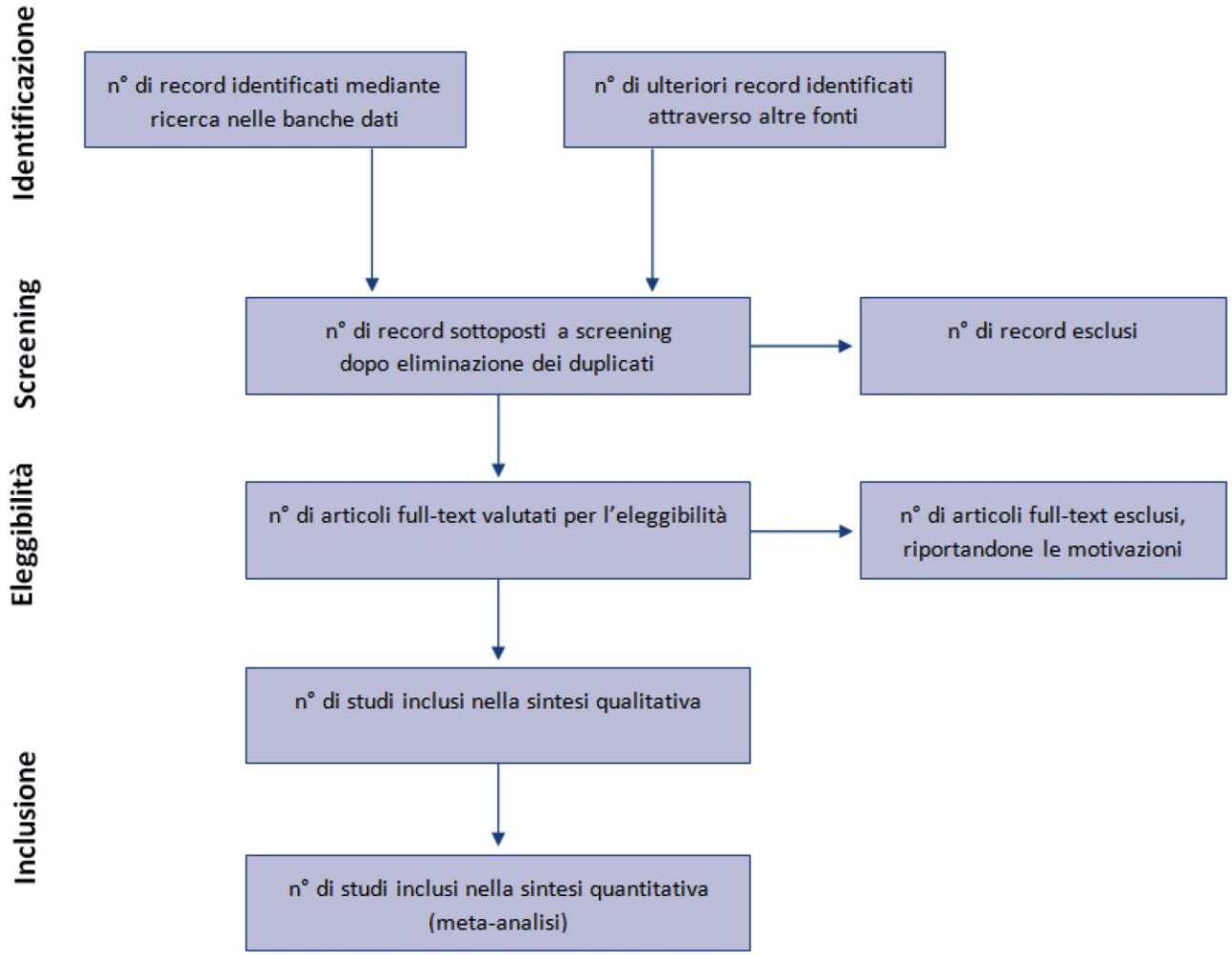
The PRISMA-P checklist contains 17 items considered to be essential and minimum components of a systematic review or meta-analysis protocol.

This PRISMA-P 2015 Explanation and Elaboration paper provides readers with a full understanding of and evidence about the necessity of each item as well as a model example from an existing published protocol.

Linee guida per il reporting di revisioni sistematiche e meta-analisi: il PRISMA Statement

David Moher[1,2*], Alessandro Liberati[3,4], Jennifer Tetzlaff[1], Douglas G. Altman[5], The PRISMA Group[6]

Evidence 2015;7(6): e1000114



GRADE Adolopment

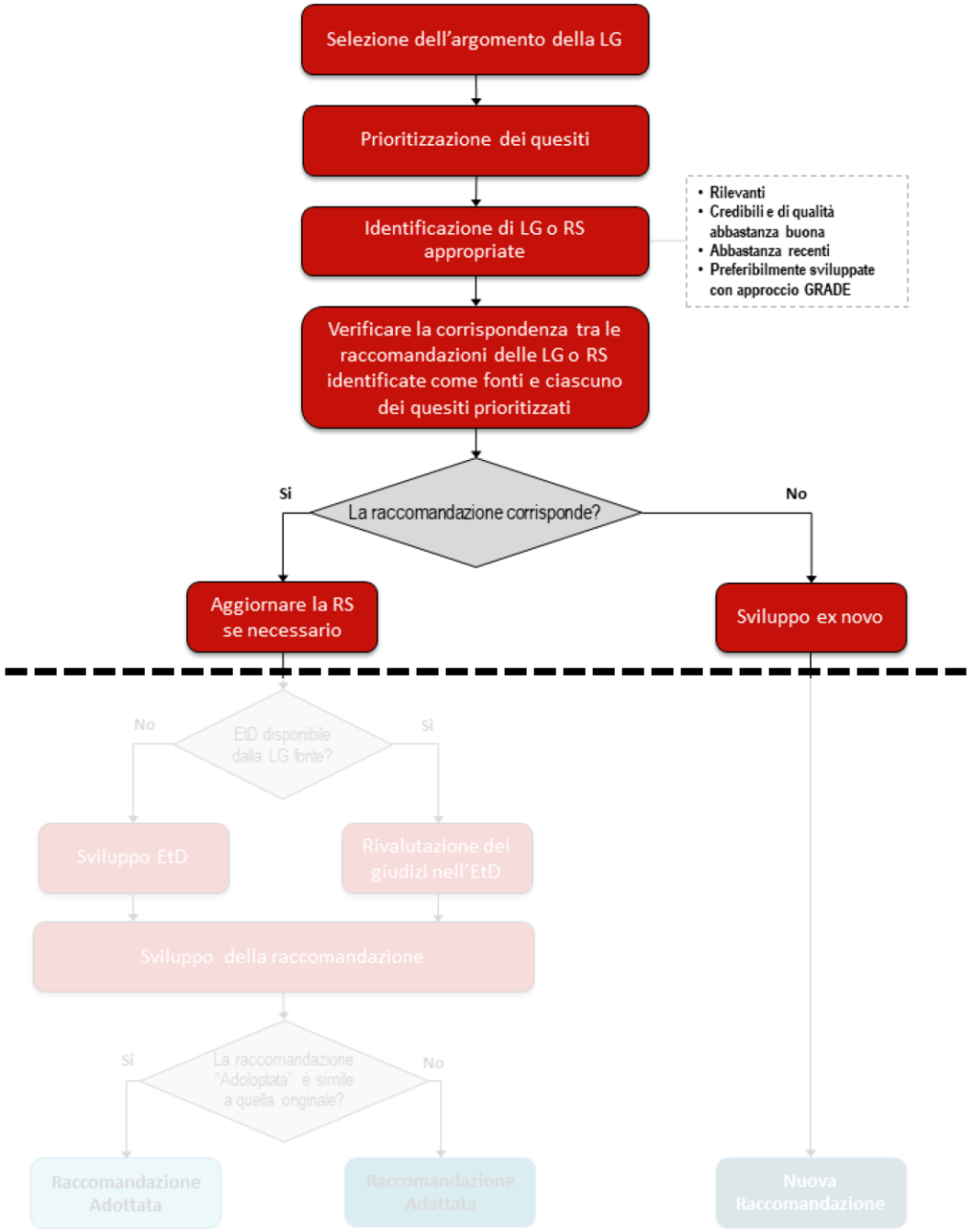
Evoluzione del metodo GRADE che consente di valutare se è opportuno produrre una LG *ex novo* ovvero è sufficiente un adattamento al contesto o l'adozione di raccomandazioni esistenti per rispondere a determinati quesiti PICO.

Decisione sulla base di requisiti di **credibilità**, **stato di aggiornamento**, **accettabilità** e **applicabilità** al contesto culturale e organizzativo.

Journal of Clinical Epidemiology 81 (2017) 101−110

GRADE Evidence to Decision (EtD) frameworks for adoption, adaptation, and de novo development of trustworthy recommendations: GRADE-ADOLOPMENT

Holger J. Schünemann[a,b,*], Wojtek Wiercioch[a], Jan Brozek[a,b], Itziar Etxeandia-Ikobaltzeta[a], Reem A. Mustafa[a,c,d], Veena Manja[e,f], Romina Brignardello-Petersen[g,h], Ignacio Neumann[a,i], Maicon Falavigna[j,k], Waleed Alhazzani[a,b], Nancy Santesso[a], Yuan Zhang[a], Jörg J. Meerpohl[l,m], Rebecca L. Morgan[a], Bram Rochwerg[a], Andrea Darzi[d], Maria Ximenas Rojas[n], Alonso Carrasco-Labra[a,i], Yaser Adi[o], Zulfa AlRayees[p], John Riva[a,q], Claudia Bollig[l], Ainsley Moore[a,q], Juan José Yepes-Nuñez[a], Carlos Cuello[a,r], Reem Waziry[s,t], Elie A. Akl[a,s]



GRADE Adolopment

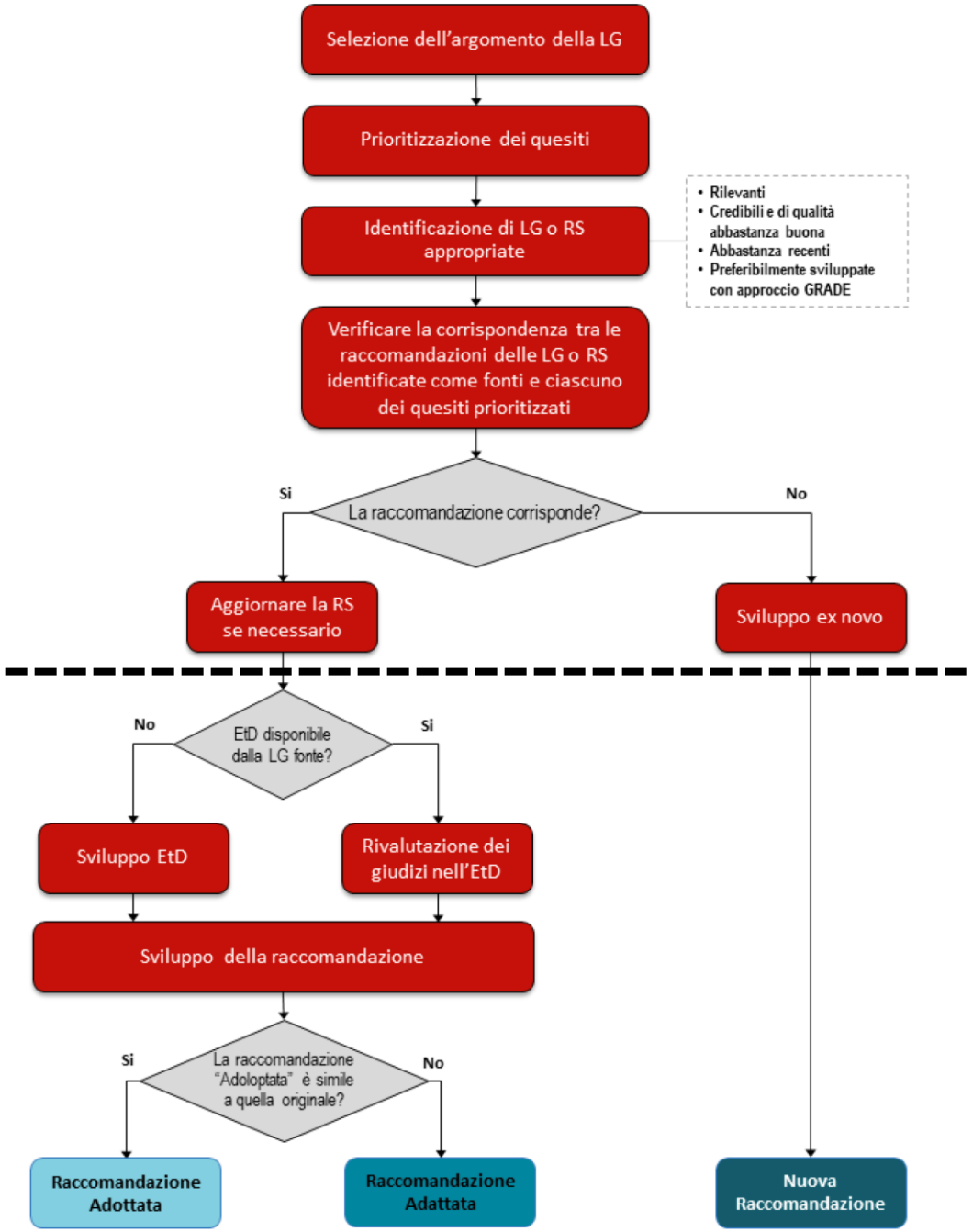
Evoluzione del metodo GRADE che consente di valutare se è opportuno produrre una LG *ex novo* ovvero è sufficiente un adattamento al contesto o l'adozione di raccomandazioni esistenti per rispondere a determinati quesiti PICO.

Decisione sulla base di requisiti di **credibilità**, **stato di aggiornamento**, **accettabilità** e **applicabilità** al contesto culturale e organizzativo.

Journal of Clinical Epidemiology 81 (2017) 101–110

GRADE Evidence to Decision (EtD) frameworks for adoption, adaptation, and de novo development of trustworthy recommendations: GRADE-ADOLOPMENT

Holger J. Schünemann[a,b,*], Wojtek Wiercioch[a], Jan Brozek[a,b], Itziar Etxeandia-Ikobaltzeta[a], Reem A. Mustafa[a,c,d], Veena Manja[e,f], Romina Brignardello-Petersen[g,h], Ignacio Neumann[a,i], Maicon Falavigna[j,k], Waleed Alhazzani[a,b], Nancy Santesso[a], Yuan Zhang[a], Jörg J. Meerpohl[l,m], Rebecca L. Morgan[a], Bram Rochwerg[a], Andrea Darzi[d], Maria Ximenas Rojas[n], Alonso Carrasco-Labra[a,i], Yaser Adi[o], Zulfa AlRayees[p], John Riva[a,q], Claudia Bollig[l], Ainsley Moore[a,q], Juan José Yepes-Nuñez[a], Carlos Cuello[a,r], Reem Waziry[s,t], Elie A. Akl[a,s]



SACRO CUORE DON CALABRIA

IRCCS

ECI CANCER CENTRE

SCUOLA DI METODOLOGIA DELLA RICERCA CLINICA

2025 - 11ª EDIZIONE

MODULI SPECIALISTICI - S4

VENERDÌ 16 - SABATO 17 MAGGIO 2025

NEGRAR DI VALPOLICELLA (VR)

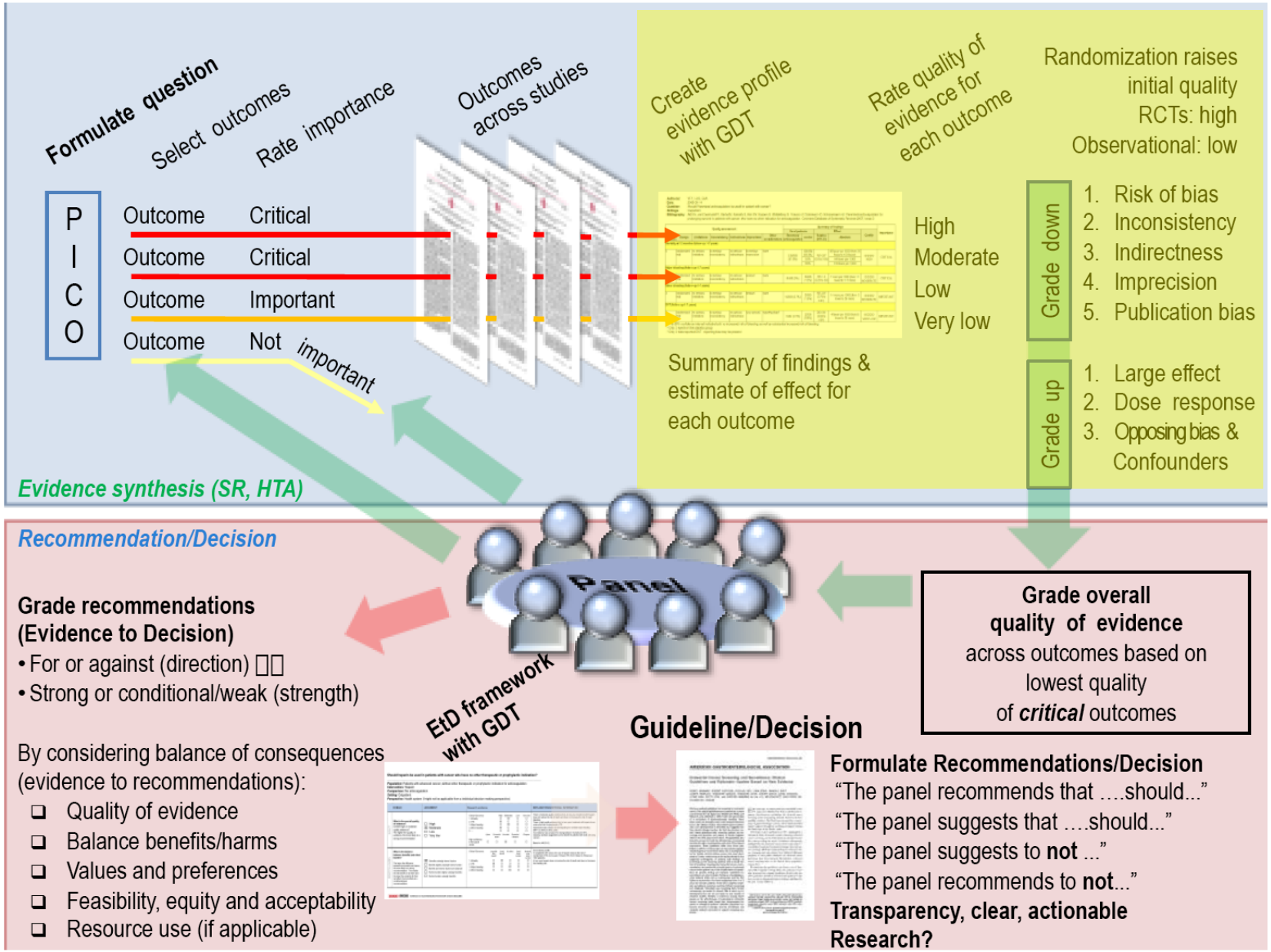
Centro Formazione IRCCS "Sacro Cuore - Don Calabria"

GRADE

La tabella sinottica delle evidenze

(Ivan Moschetti)

The GRADE process in developing guidelines



La Table of evidence (ToEs)

Evidence profile: use of antibiotics (penicillin) versus no use of antibiotics in children with sickle cell disease. Source: Hirst et al. 4

Quality assessment							№ of patients		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Penicillin prophylaxis	Standard care	Relative (95% CI)	Absolute (95% CI)		
Incidence of pneumococcal infection, for initiation of treatment - Initiation of penicillin												
2	Randomized trials	Not serious [1]	serious [2]	not serious [3]	serious [4]	none [5]	9/248 (3.6%)	19/209 (9.1%)	**OR 0.37** (0.16 to 0.86)	55 fewer per 1000 (from 12 fewer to 75 fewer)	⊕⊕◯◯ LOW	CRITICAL
Deaths, for initiation of treatment - Initiation of penicillin												
1	randomized trials	not serious [6]	not serious	not serious	serious [4]	none [5]	0/105 (0.0%)	4/110 (3.6%)	**OR 0.11** (0.01 to 2.11)	32 fewer per 1000 (from 36 fewer to 37 more)	⊕⊕⊕◯ MODERATE	CRITICAL
Adverse drug effects - Nausea and vomiting												
1	randomized trials	not serious [6]	not serious	not serious	serious [4]	none [5]	2/201 (1.0%)	1/199 (0.5%)	**OR 1.99** (0.18 to 22.12)	5 more per 1000 (from 4 fewer to 95 more)	⊕⊕⊕◯ MODERATE	CRITICAL

1.blinding and concealment were not clear for one of the two studies
2.heterogeneity exists; p-value for testing heterogeneity is 0.07 and I2=69%
3.the question addressed is the same for the evidence regarding the population, intervention, comparator and outcome
4.total sample size is small, and the total number of events is <300 (a threshold rule-of-thumb value)
5.insufficient number of studies to assess publication bias
6.unclear allocation concealment

Domanda: Agopuntura rispetto a trattamento non farmacologico per lombalgia cronica aspecifica

Certainty assessment							Nº di pazienti		Effetto		Certo	Importanza
Nº degli studi	Disegno dello studio	Rischio di distorsione	Mancanza di riproducibilità dei risultati	Mancanza di generalizzabilità	Imprecisione	Ulteriori considerazioni	agopuntura	trattamento non farmacologico	Relativo (95% CI)	Assoluto (95% CI)		
dolore al termine del trattamento (follow up: intervallo 4 settimane a 5 settimane; valutato con: VAS)												
3[1,2,3]	studi randomizzati	serio[a]	non importante	non importante	serio[b]	nessuno	76	65	-	**MD 0.1 maggiore** (15.05 inferiore a 15.25 maggiore)	⨁⨁◯◯ Bassa	CRITICAL
Qualità della vita al termine del trattamento Agopuntura media intensità (follow up: medio 4 settimane; valutato con: SF 36 totale)												
1[1]	studi randomizzati	serio[a]	non importante	non importante	molto serio[c]	nessuno	39	29	-	**MD 4.2 inferiore** (10.11 inferiore a 1.71 maggiore)	⨁◯◯◯ Molto bassa	CRITICAL
partecipanti con uso analgesici al termine del trattamento. agopuntura media intensità (follow up: medio 10 settimane)												
1[4]	studi randomizzati	serio[a]	non importante	non importante	molto serio[d]	nessuno	48/94 (51.1%)	37/78 (47.4%)	**RR 1.08** (0.79 a 1.46)	**4 più per 100** (da 10 meno a 22 più)	⨁◯◯◯ Molto bassa	CRITICAL

CI: Confidence interval; **MD:** Mean difference; **RR:** Risk ratio; **SMD:** Standardised mean difference

Spiegazioni

a. abbassato di un livello per alto rischio di performance e detection bias

b. abbassato di un livello per imprecisione: meno di 400 partecipanti

c. abbassato di due livelli per imprecisione: meno di 100 partecipanti

d. abbassato di due livelli per imprecisione: meno di 100 eventi e intervalli di confidenza molto ampi

Sintesi dei risultati:

Agopuntura rispetto a trattamento non farmacologico per lombalgia cronica aspecifica

Paziente o popolazione: lombalgia cronica aspecifica

Intervento: agopuntura

Confronto: trattamento non farmacologico

Esiti	Effetto assoluto anticipato* (95% CI)		Effetto relativo (95% CI)	№ dei partecipanti (studi)	Certezza delle prove (GRADE)	Commenti
	Rischio con trattamento non farmacologico	Rischio con agopuntura				
dolore al termine del trattamento valutato con: VAS follow up: intervallo 4 settimane a 5 settimane	La media dolore al termine del trattamento era **37.64**	MD **0.1 maggiore** (15.05 inferiore a 15.25 maggiore)	-	141 (3 RCT)[1,2,3]	⨁⨁◯◯ Bassa[a,b]	
Qualità della vita al termine del trattamento Agopuntura media intensità valutato con: SF 36 totale follow up: medio 4 settimane	La media qualità della vita al termine del trattamento Agopuntura media intensità era **32.6**	MD **4.2 inferiore** (10.11 inferiore a 1.71 maggiore)	-	68 (1 RCT)[1]	⨁◯◯◯ Molto bassa[a,c]	
partecipanti con uso analgesici al termine del trattamento. agopuntura media intensità follow up: medio 10 settimane	47 per 100	**51 per 100** (37 a 69)	**RR 1.08** (0.79 a 1.46)	172 (1 RCT)[4]	⨁◯◯◯ Molto bassa[a,d]	

* **Il rischio nel gruppo di intervento** (e il suo intervallo di confidenza (IC) al 95%) si basa sul rischio assunto nel gruppo di controllo e **sull'effetto relativo** dell'intervento (e il suo IC al 95 %).

SACRO CUORE DON CALABRIA

ECI CANCER CENTRE

I R C C S

SCUOLA DI METODOLOGIA DELLA RICERCA CLINICA

2025 - 11ª EDIZIONE

MODULI SPECIALISTICI - S4

VENERDÌ 16 - SABATO 17 MAGGIO 2025

NEGRAR DI VALPOLICELLA (VR)

Centro Formazione IRCCS "Sacro Cuore - Don Calabria"

GRADE

La valutazione della qualità delle prove (1): *risk of bias*

(Michela Cinquini)

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

Tabella 5- Criteri* per l'aumento (*upgrading*) o la diminuzione (*downgrading*) del giudizio di qualità (alta, moderata, bassa, molto bassa) delle prove

Tipo di prove	Studio controllato e randomizzato = alta Studio osservazionale = bassa Qualsiasi altro tipo di informazione = molto basso
A. Diminuzione della categoria di attribuzione (es. da "alta" a "moderata")	1. Limiti gravi (-1 livello) o molto gravi (-2 livelli) nella qualità di conduzione dello studio 2. Incoerenza nei risultati tra studi diversi sullo stesso quesito (-1 o -2 livelli) 3. Alcune (-1 livello) o importanti (-2 livelli) incertezze circa la diretta trasferibilità dei risultati (directness) 4. Imprecisione o dati insufficienti (*sparse data*) (-1 o -2 livelli) 5. Possibilità di pubblicazione selettiva dei dati (*publication* e *reporting bias*) (-1 o-2 livelli)
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. Aver considerato tutti i possibili fattori di confondimento che avrebbero potuto alterare le stime di effetto (+1 livello)

* non sono da considerarsi un algoritmo

Tabella 5- Criteri* per l'aumento (*upgrading*) o la diminuzione (*downgrading*) del giudizio di qualità (alta, moderata, bassa, molto bassa) delle prove

Tipo di prove	Studio controllato e randomizzato = alta Studio osservazionale = bassa Qualsiasi altro tipo di informazione = molto basso
A. Diminuzione della categoria di attribuzione (es. da "alta" a "moderata")	1. Limiti gravi (-1 livello) o molto gravi (-2 livelli) nella qualità di conduzione dello studio 2. Incoerenza nei risultati tra studi diversi sullo stesso quesito (-1 o -2 livelli) 3. Alcune (-1 livello) o importanti (-2 livelli) incertezze circa la diretta trasferibilità dei risultati (directness) 4. Imprecisione o dati insufficienti (*sparse data*) (-1 o -2 livelli) 5. Possibilità di pubblicazione selettiva dei dati (*publication* e *reporting bias*) (-1 o-2 livelli)
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. Aver considerato tutti i possibili fattori di confondimento che avrebbero potuto alterare le stime di effetto (+1 livello)

* non sono da considerarsi un algoritmo

Tabella 5- Criteri* per l'aumento (*upgrading*) o la diminuzione (*downgrading*) del giudizio di qualità (alta, moderata, bassa, molto bassa) delle prove

Tipo di prove	Studio controllato e randomizzato = alta Studio osservazionale = bassa Qualsiasi altro tipo di informazione = molto basso
A. Diminuzione della categoria di attribuzione (es. da "alta" a "moderata")	1. Limiti gravi (-1 livello) o molto gravi (-2 livelli) nella qualità di conduzione dello studio 2. Incoerenza nei risultati tra studi diversi sullo stesso quesito (-1 o -2 livelli) 3. Alcune (-1 livello) o importanti (-2 livelli) incertezze circa la diretta trasferibilità dei risultati (directness) 4. Imprecisione o dati insufficienti (*sparse data*) (-1 o -2 livelli) 5. Possibilità di pubblicazione selettiva dei dati (*publication* e *reporting bias*) (-1 o-2 livelli)
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. Aver considerato tutti i possibili fattori di confondimento che avrebbero potuto alterare le stime di effetto (+1 livello)

* non sono da considerarsi un algoritmo

- **RCTs ⊕⊕⊕⊕**
- **observational studies ⊕⊕◯◯**
- **5 factors that can lower quality**
  1. limitations in study design, execution and reporting *(risk of bias criteria)*
  2. Inconsistency *(or heterogeneity)*
  3. Indirectness *(PICO and applicability)*
  4. Imprecision
  5. Publication bias
- **3 factors can increase quality**
  1. large magnitude of effect
  2. opposing plausible residual bias or confounding
  3. dose-response gradient

ELSEVIER

Journal of Clinical Epidemiology 64 (2011) 407–415

Journal of Clinical Epidemiology

GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias)

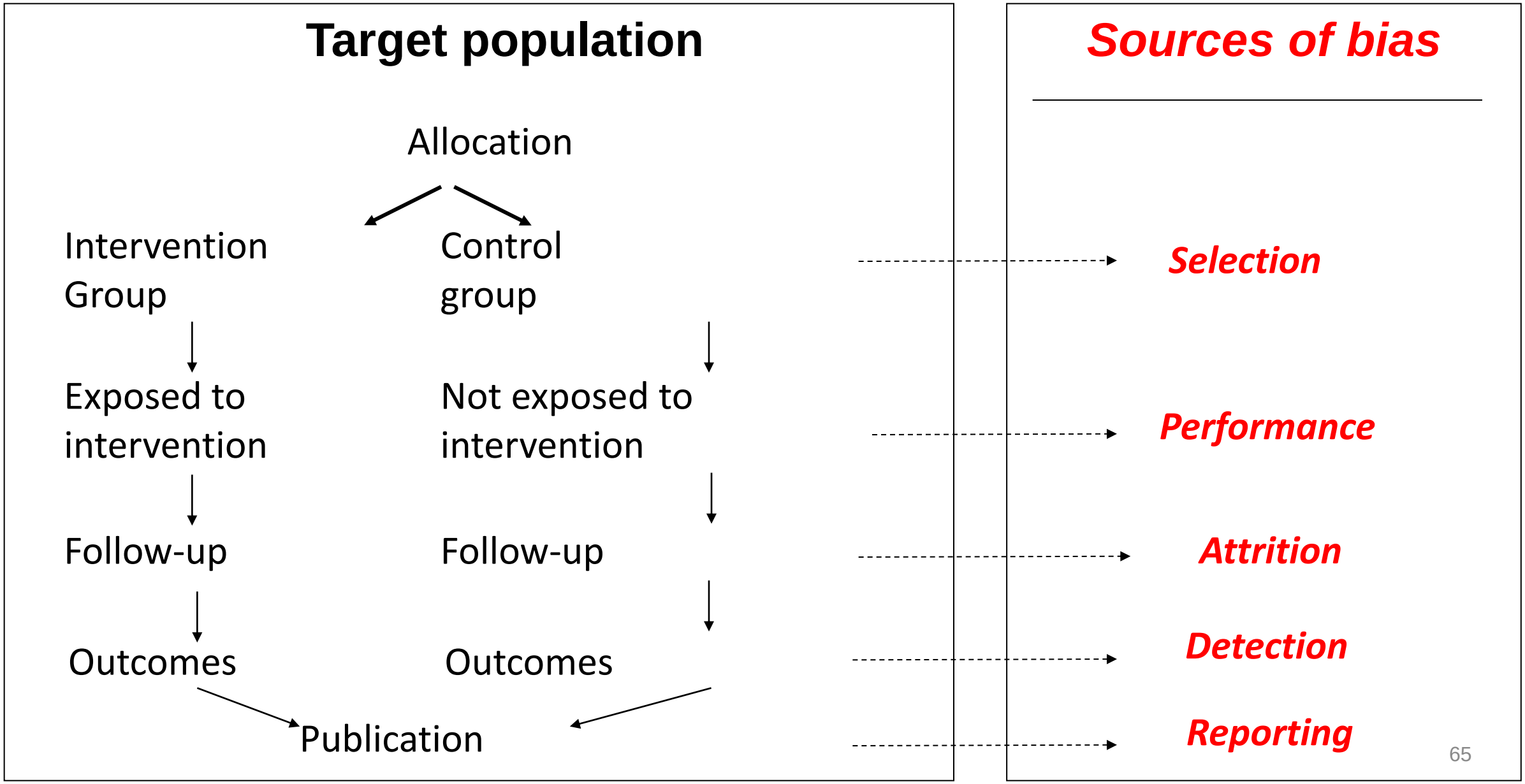
Gordon H. Guyatt[a,*], Andrew D. Oxman[b], Gunn Vist[b], Regina Kunz[c], Jan Brozek[a], Pablo Alonso-Coello[d], Victor Montori[e], Elie A. Akl[f], Ben Djulbegovic[g,h,i], Yngve Falck-Ytter[j], Susan L. Norris[k], John W. Williams Jr.[l], David Atkins[m], Joerg Meerpohl[n,o], Holger J. Schünemann[a]

Bias

Systematic distortion of the estimated intervention effect away from the truth, caused by **inadequacies** in the **design**, **conduct**, or **analysis** of a trial , or in the **publication of its results**. In other words, in a biased trial, the results observed reflect other factors in addition to (or, in extreme cases, instead of) the effect of the tested therapeutic procedure alone.

Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001;134:663–94

Trial as a flow



Selection bias: due componenti

Generazione lista di randomizzazione

- **Basso rischio di bias**. Uso di metodi realmente casuali come ad esempio: tavole di numeri random, sistemi computerizzati, lancio di una moneta o di un dado, sorteggio.
- **Alto rischio di bias.** Uso di metodi **NON** realmente casuali come ad esempio: giorno di nascita o di ammissione in ospedale, giudizio del medico, preferenze del paziente, risultati di test di laboratorio, disponibilità del trattamento, alternanza
- **Rischio incerto**. Quando non sono disponibili sufficienti informazioni per definire il rischio (bassa qualità del reporting)

Nascondimento della sequenza di randomizzazione*

- **Basso rischio di bias**. Sperimentatori che arruolano i pazienti non possono prevedere in quale gruppo verrà inserito il paziente perché si usa uno dei seguenti metodi: randomizzazione centralizzata (telefonica, via web, o gestita da personale esterno alla sperimentazione - farmacista, statistico); buste chiuse e opache.
- **Alto rischio di bias.** Sperimentatori che arruolano i pazienti possono prevedere in quale gruppo verrà inserito il paziente perché si usa uno dei seguenti metodi: liste di randomizzazione, buste aperte o non opache, alternanza, data di nascita, numero di cartella, ect.
- **Rischio incerto**. Non sono disponibili sufficienti informazioni per definire il rischio (bassa qualità del reporting)

*allocation concealment

Esempio 1

To prevent the introduction of bias, randomisation will be via an online system, accessed via http://www.ctu.co.uk, hosted by the Clinical Trials Unit (CTU), King's College London (KCL). Eligible participants will be randomised to either the nilvadipine or placebo treatment group. The

***Generazione della sequenza**: basso rischio*

***Allocation concealment:** basso rischio*

Esempio 2

Eligible patients with even-number hospital records were assigned to treatment (5000 U twice daily), those with odd-number records served as controls

***Generazione della sequenza**: alto rischio*

***Allocation concealment:** alto rischio*

Esempio 3

Subjects were assigned at random in a 2:1 ratio to naltrexone or control (Cornish 1977)

***Sequence generation**: rischio incerto*

***Allocation concealment:** rischio incerto*

Esempio 4

Patients were randomly assigned in a 1:1:1 ratio to one of three treatment groups: ledipasvir–sofosbuvir for 8 weeks, ledipasvir–sofosbuvir plus ribavirin for 8 weeks, or ledipasvir–sofosbuvir for 12 weeks. Randomization was stratified according to HCV genotype (1a or 1b).

***Generazione della sequenza**: rischio incerto*

***Allocation concealment:** rischio incerto*

Blinding

- Sperimentatori e partecipanti non conoscono gruppo di allocazione (*performance bias*)

- Valutatori degli esiti non conoscono gruppo di allocazione (*detection bias*)

Singolo cieco
i pazienti inclusi nello studio non conoscono il gruppo al quale sono stati assegnati

Doppio cieco
i pazienti e gli sperimentatori non conoscono il gruppo al quale (i pazienti) sono stati assegnati

Triplo cieco
i pazienti, gli sperimentatori e i valutatori degli esiti non conoscono il gruppo di allocazione

…

Non sempre il significato è questo … è sempre bene valutare chi è davvero in cieco!

Performance bias

Si verifica quando i partecipanti allo studio (sperimentatori o pazienti) modificano i loro comportamenti perché sanno a quale gruppo è assegnato un dato paziente

Esempi:

Lo sperimentatore controlla la presenza di effetti avversi più frequentemente nei pazienti assegnati al gruppo di trattamento.

Un paziente nel gruppo placebo assume altri farmaci, fa più (o meno) visite di controllo.

Detection bias

Si verifica quando la valutazione degli esiti dello studio viene influenzata dalla conoscenza del gruppo al quale è assegnato un dato paziente

Esempi:

Interpretazione di esiti radiologici, risoluzione dei sintomi, valutazione delle ricadute di malattia diversa nei pazienti assegnati al trattamento e al controllo

Performance bias

Cecità di pazienti e sperimentatori

- **Basso rischio di bias.** Pazienti e sperimentatori non conoscono l'assegnazione dei pazienti al gruppo di controllo o di trattamento oppure è poco probabile che la mancanza di cecità influenzi la performance di pazienti e sperimentatori
- **Alto rischio di bias.** Pazienti e sperimentatori conoscono l'assegnazione dei pazienti o, durante lo studio, diventa chiaro a quale gruppo di trattamento sono allocati (rottura del cieco). Studi definiti come "open label"
- **Rischio incerto**. Quando non sono disponibili sufficienti informazioni per definire il rischio (bassa qualità del reporting)

Detection bias

Cecità del valutatore degli esiti dello studio (outcome)

- **Basso rischio di bias.** L'esito dello studio è valutato senza conoscere l'assegnazione dei pazienti al gruppo di controllo o di intervento; oppure è poco probabile che la mancanza di cecità influenzi la valutazione
- **Alto rischio di bias.** L'esito dello studio è valutato conoscendo l'assegnazione dei pazienti al gruppo di controllo o di intervento ed è probabile che la mancanza di cecità influenzi la valutazione
- **Rischio incerto**. Quando non sono disponibili sufficienti informazioni per definire il rischio (bassa qualità del reporting)

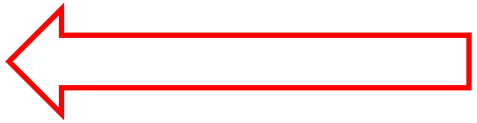
Performance and detection bias

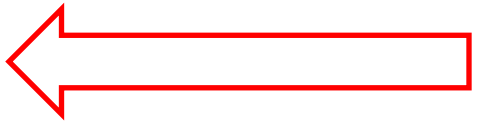
- Impatto diverso su outcome **soggettivi** e **oggettivi** (quindi la valutazione va fatta separatamente)
- Se studio su **farmaco in doppio cieco** e dice che tutti gli operatori erano all'oscuro dell'assegnazione è probabile che sia in cieco anche l'outcome assessor, anche se non espressamente detto
- Se studio su **interventi che non possono essere in doppio cieco** (psicosociali, educativi, chirurgici, riabilitativi) importante che sia in cieco l'outcome assessor e deve essere specificato
  - Performance: high risk per outcomes soggettivi sempre
  - Detection: low risk se c'è blinding of outcome assessor anche per outcomes soggettivi

Esempio 1

Eligible participants will be randomised to either the nilvadipine or placebo treatment group. The nilvadipine capsules and placebo capsules will be packaged and labelled identically. Randomisation will be at the level of the individual patient, using block randomisation with randomly varying block sizes and stratified by country site. Once the patient has been randomised, the online system will automatically recognise which treatment packs are located in each study pharmacy at the recruiting study site and will randomly select a pack in the appropriate trial arm to be dispensed to the patient. All study staff at all sites will be blinded to treatment allocation and will remain blind until the end of the trial.

The primary outcome measure is the change from baseline to week 78 in cognitive function, as assessed by the ADAS-Cog 12.





Performance bias: *basso rischio*

Detection bias: *basso rischio*

Esempio 2

PROCEDURES

After providing written informed consent, all trial participants were randomly assigned to receive one of two doses of dabigatran, or to receive warfarin, by means of a central, interactive, automated telephone system. Dabigatran was administered, in a blinded fashion, in capsules containing either 110 mg or 150 mg of the drug, to be taken twice daily. Warfarin was administered, in an unblinded fashion, in tablets of 1, 3, or 5 mg and was adjusted locally to an international normalized ratio (INR) of 2.0 to 3.0, with the INR measured at least monthly. The time that the INR was with-

.....

group and were centrally blinded. Each primary and secondary outcome event was adjudicated by two independent investigators who were unaware of the treatment assignments.

Performance bias: *alto rischio*

Detection bias: *basso rischio*

Esempio 3

Study medications were prepared by a research pharmacist, who had no direct contact with participants. Buprenorphine mono tablets (containing only buprenorphine) and placebo tablets that appeared identical were provided by the manufacturer. Naltrexone was purchased for the study: tablets were crushed, and the study pharmacist placed naltrexone or placebo inside capsules that appeared identical. To mask slight taste differences between active and placebo buprenorphine tablets, participants gargled with a mentholated antiseptic mouthwash before taking the sublingual tablets (Schottenfeld 2008)

Performance bias: basso rischio

Detection bias: basso rischio

Attrition bias

- Quando non tutti i soggetti randomizzati completano lo studio
- i soggetti non escono a caso dallo studio: è possibile che quelli che escono siano sistematicamente diversi da quelli che non escono: i gruppi non sono più randomizzati
- **Validità esterna** : es: escono tutti i più giovani, o i meno gravi, o i maschi: posso trarre conclusioni solo su quelli che rimangono
- **Validità interna (Bias):** se la probabilità di uscire dallo studio è legata all'intervento o all'outcome, cioè se quelli che escono hanno sistematicamente probabilità più alte o più basse di avere l'outcome di quelli che restano

Attrition bias

Low risk of bias

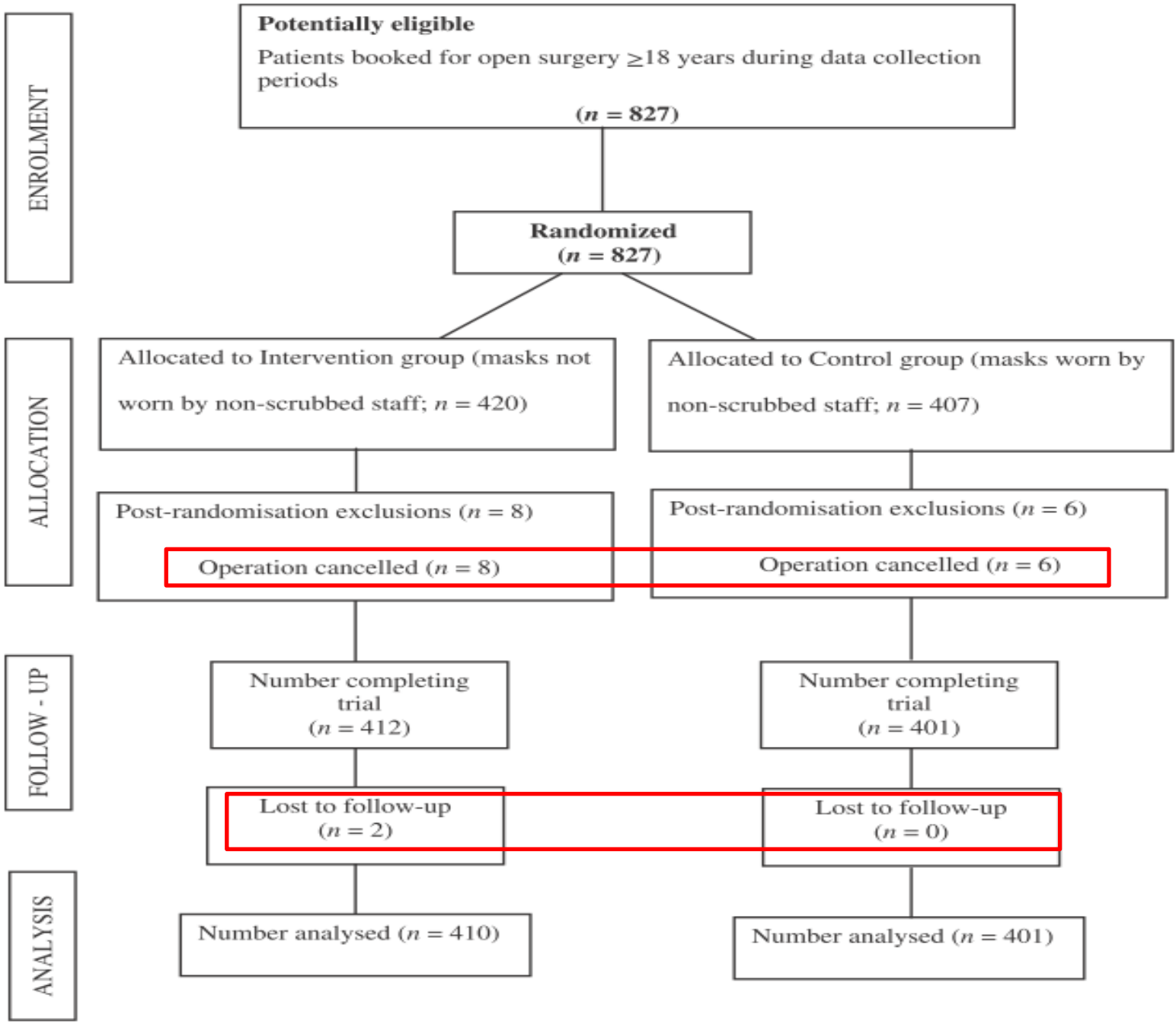
- No missing outcome data;
- the proportion of missing outcomes compared with observed event risk not enough to have a relevant impact on the intervention effect;
- Missing outcome data balanced in numbers across intervention groups, with similar reasons across groups;
- Missing data imputed using appropriate methods
- All patients analysed in the group they were allocated to by randomisation irrespective of non-compliance and co-interventions (intention to treat)

High risk of bias:

- the proportion of missing outcomes compared with observed event risk enough to induce relevant bias in intervention effect estimate
- Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups;

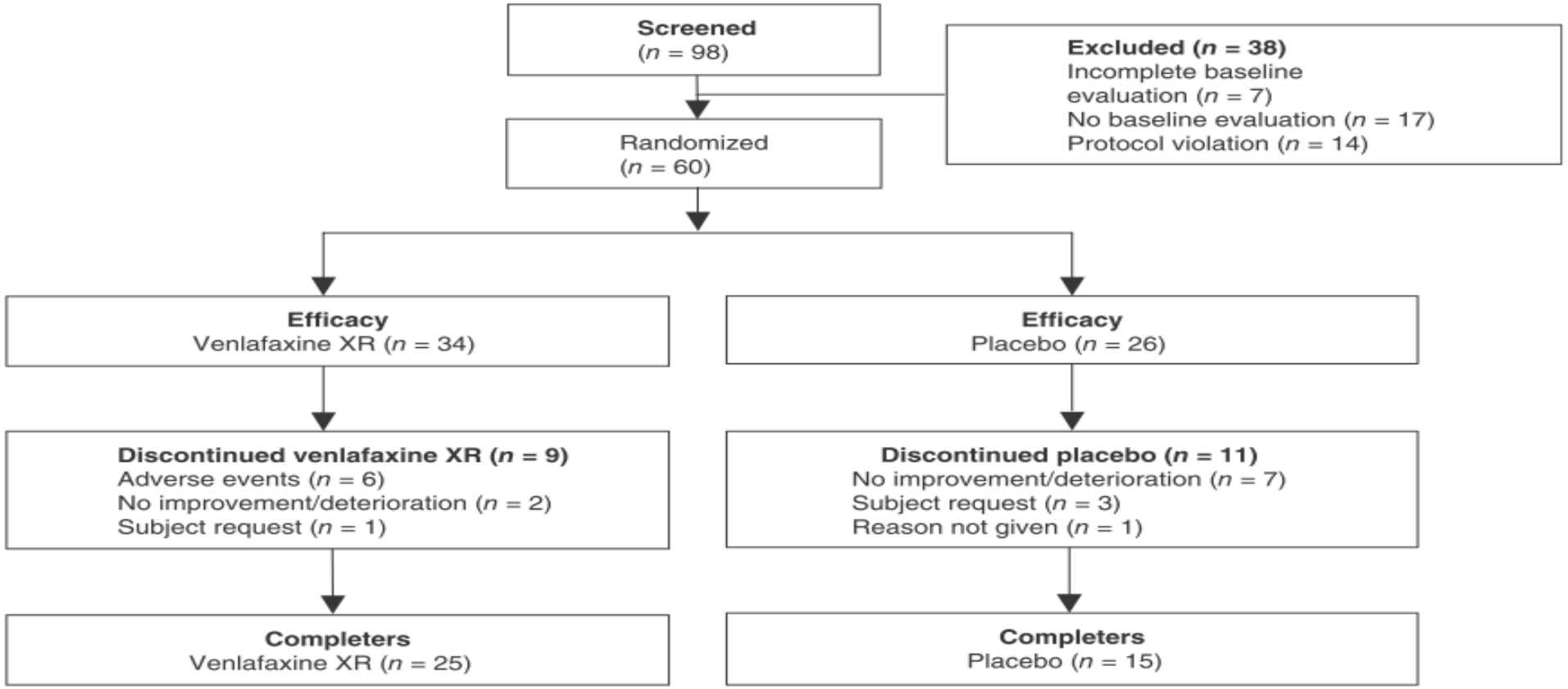
Esempio 1

A tota
enrolled and 811 (98.1%) patients complete
group and 410 No Mask group (Fig. 1).



Attrition bias:
basso rischio

Esempio 2



Totale:
33%
(20 su 60)

Intervento: 26%
(9 su 34)

Placebo: 42%
(11 su 26)

Attrition bias: *alto rischio*

Esempio 3

L’informazione nella sezione risultati

A total of 1395 (91.3%) randomized patients completed the doubleblind treatment phase. There were no patterns or trends suggesting any differences in discontinuation rates or the reasons for discontinuation between the treatment groups (Figure 1). (Bays 2004)

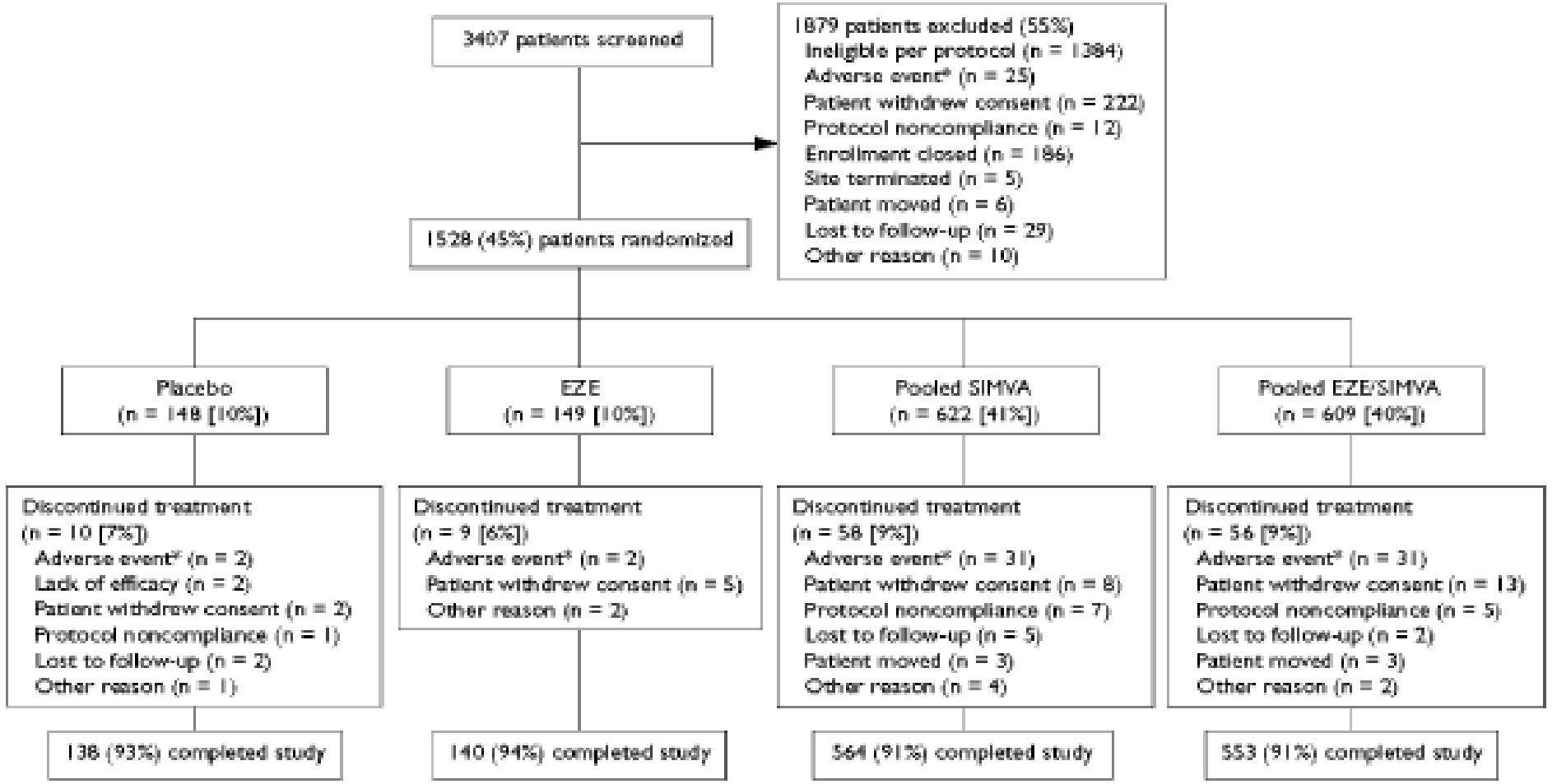


Figure 1. Disposition of patients in the study. Number of patients who were randomized, discontinued prematurely, and completed the study are shown for the placebo, ezetimibe monotherapy (EZE), pooled simvastatin monotherapy (pooled SIMVA), and pooled ezetimibe/simvastatin tablet (pooled EZE/SIMVA) groups. *Number of patients with clinical and laboratory adverse events.

Attrition bias: *basso rischio*

Reporting bias is selection bias

- Reporting bias is perhaps the greatest source of selection bias
- Originally defined as the publication or non-publication of studies depending on the direction and statistical significance of the results
- Is a complex phenomenon

Full — n and effect size, plus precision / *p*-value for continuous data

Partial — Effect size or precision (± n or *p*-value)

Qualitative — *p*-value

Unreported

Incompletely reported outcomes

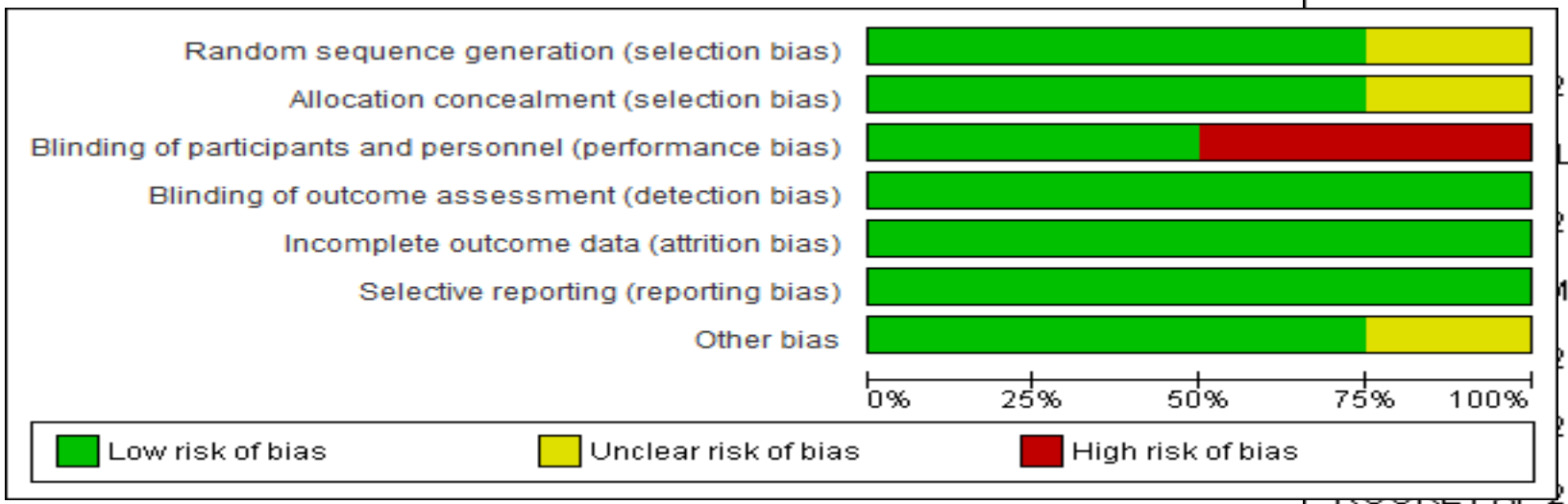
(Chan, 2004)

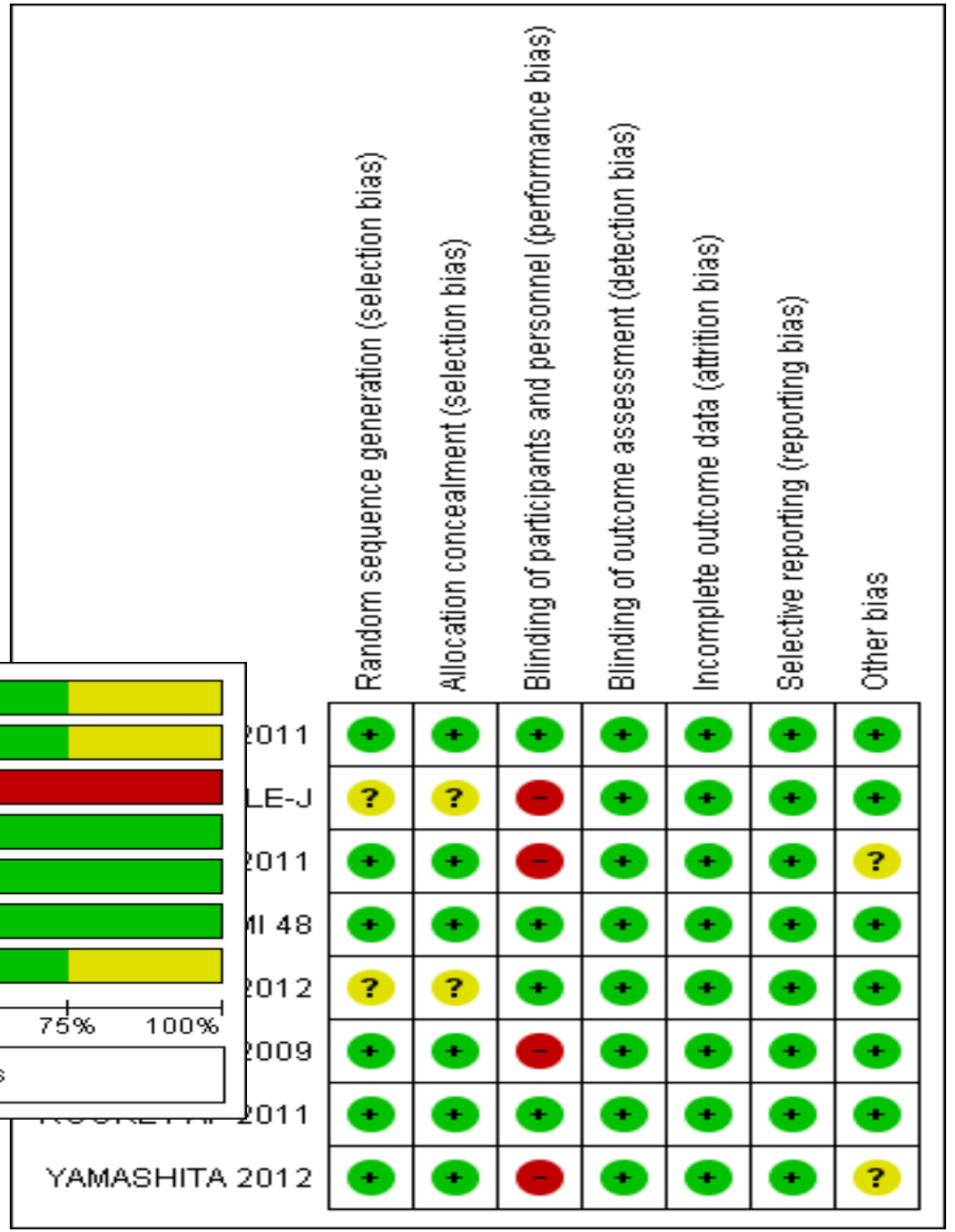
Risk of bias in one study

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	NR
Allocation concealment (selection bias)	Unclear risk	NR
Blinding of participants and personnel (performance bias)	High risk	open label
Blinding of outcome assessment (detection bias)	Low risk	An independent blinded endpoint committee adjudicated all reported bleeding and efficacy events
Incomplete outcome data (attrition bias)	Low risk	ITT. all patients followed up
Selective reporting (reporting bias)	Low risk	
Other bias	Low risk	

Risk of bias across studies/domains





Dal Risk of bias al GRADE

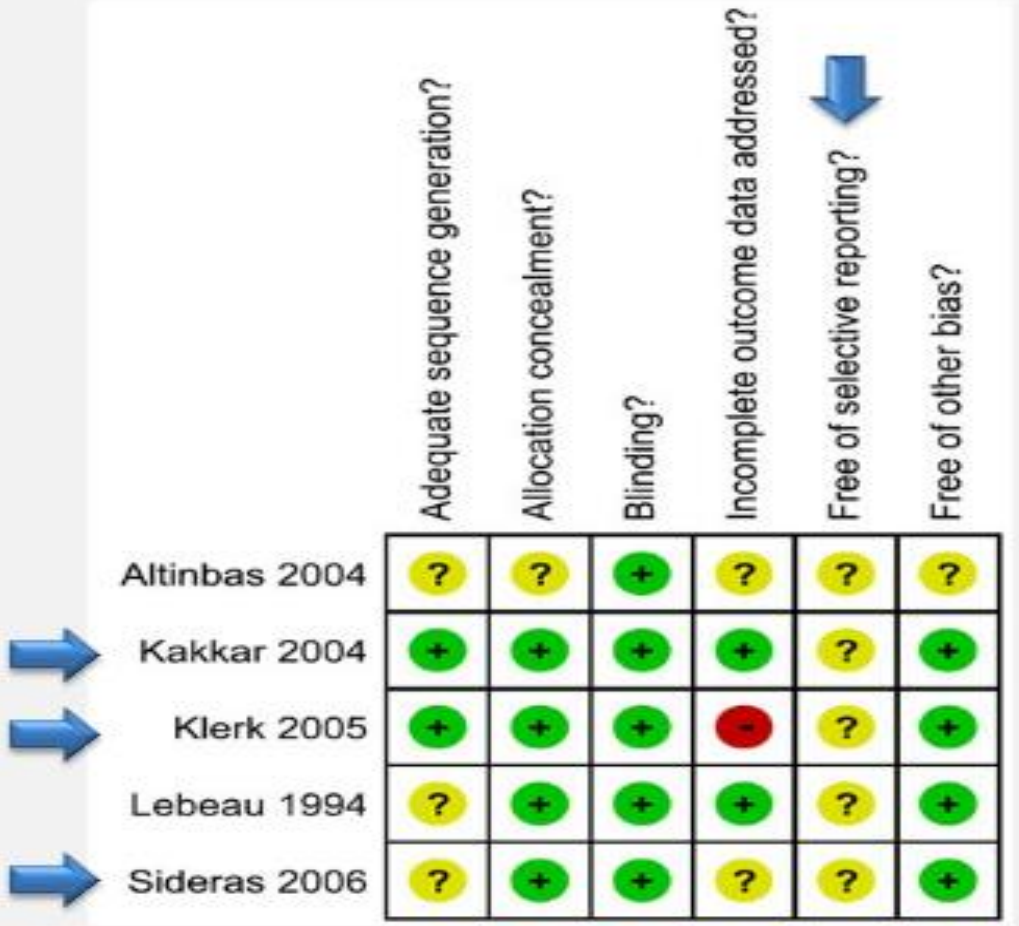
- La fiducia nelle prove complessiva, riferita ad un **singolo outcome**, può essere abbassata se si ritiene che esista un **importante rischio di bias** negli studi che contribuiscono a quelle evidenze

- Fiducia (certainty) minore nell'affidabilità dei risultati

- **APPROCCIO GRADE**
  - Se non si sospettano bias importanti, non si abbassa «not serious» (no change in quality)
  - Se si sospettano bias importanti, si abbassa di 1 livello «serious» (downgrade quality of evidence 1 level)
  - Se si sospettano bias molto importanti, si abbassa di 2 livelli «very serious» (downgrade quality of evidence 2 level)

Example: Major Bleeding with anticoagulation

Population: *people with cancer*
Intervention: *anti-coagulants, such as heparin*
Outcome: *major bleeding*



Alkl et al, CDSR 2008

Concern maggiore su selective reporting of outcome
Solo tre studi riportano l'outcome major bleeding

!!! Do not count number of green, yellow, red spots!!!

Would you downgrade for risk of bias?

☐ No, there are no serious limitations

☒ Yes, there are serious limitations

Most people would agree that selective reporting of outcome is concern

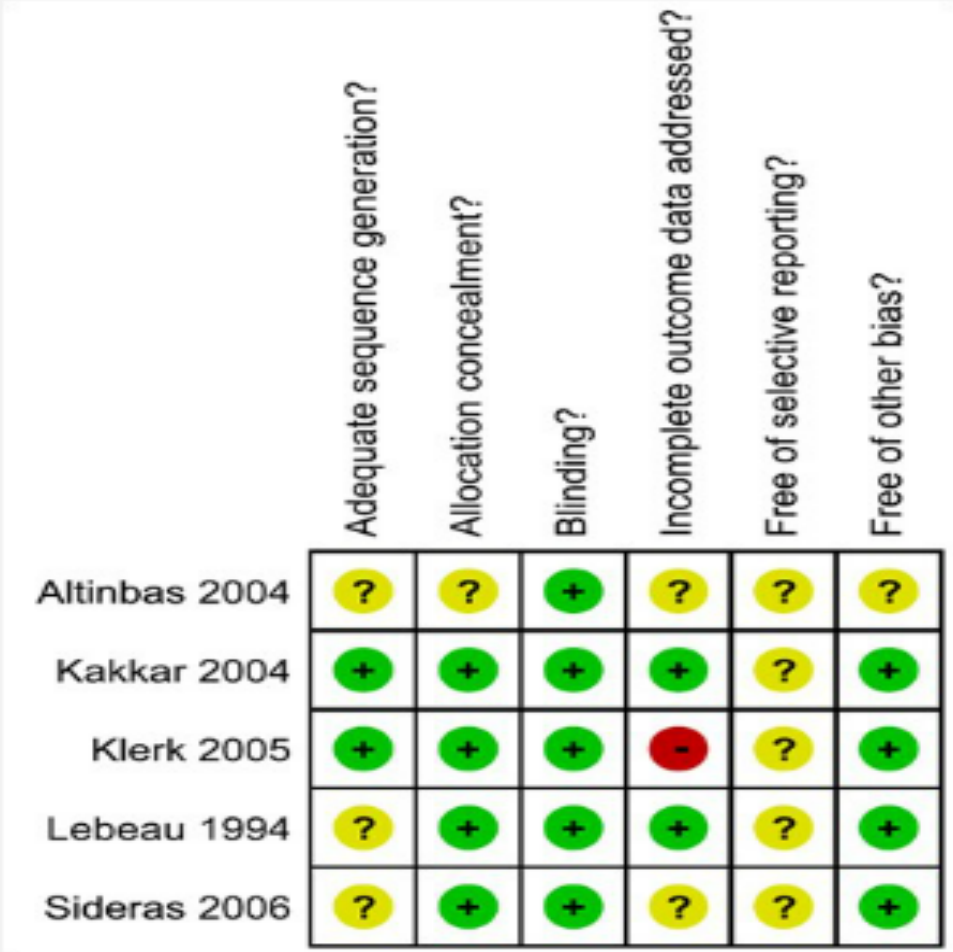
☐ Yes, there are very serious limitations

Example: Major Bleeding with anticoagulation

Population: *people with cancer*
Intervention: *anti-coagulants, such as heparin*
Outcome: *major bleeding*

Limitations likely to reduce confidence in effect
Downgrade 1 level

Footnotes
[1] Only 3 of 5 studies reported major bleeding, selective reporting likely



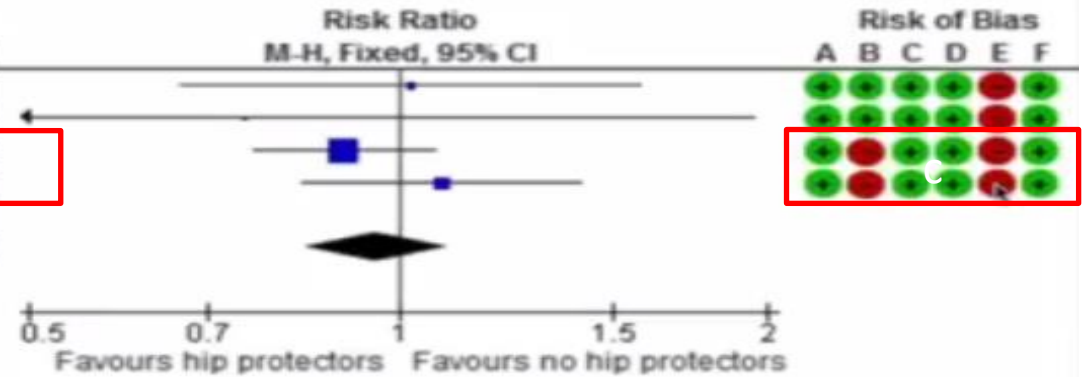
Alkl et al, CDSR 2008

Mortality (dichotomous outcome)

Mortality

Study or Subgroup	Hip protectors Events	Total	No hip protectors Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Cameron 2001	28	86	28	88	9.5%	1.02 [0.66, 1.58]
Jantti 1996	6	36	8	36	2.7%	0.75 [0.29, 1.94]
Meyer 2003	157	459	183	483	61.1%	0.90 [0.76, 1.07]
Van Schoor 2003	83	276	79	285	26.6%	1.08 [0.84, 1.41]
Total (95% CI)		**857**		**892**	**100.0%**	**0.96 [0.84, 1.10]**
Total events	274		298			

Heterogeneity: Chi² = 1.69, df = 3 (P = 0.64); I² = 0%
Test for overall effect: Z = 0.62 (P = 0.54)



A: random sequence generation
B: allocation concealment
C: Blinding of participants and personnel
D: Blinding of outcome assessor
E: Incomplete outcome data
F: selective outcome reporting

Would you downgrade for risk of bias?

☐ No, there are no serious limitations

☒ Yes, there are serious limitations

Some concerns with allocation concealment and high losses to follow up

☐ Yes, there are very serious limitations

Dal Risk of bias al GRADE

- Non fare medie, non contare i domini a rischio o quelli adeguati
- Tutto si basa su giudizi, spesso soggettivi
- Calarsi nel contesto clinico e metodologico
- Trasparenza nelle decisioni (soprattutto nelle situazioni intermedie)
- Focus sugli studi a basso rischio di bias
- Focus sugli studi che portano un maggiore contributo informativo (peso nella meta-analisi)
- Siate conservativi
- Rischio di bias nel contesto delle altre dimensioni della qualità

GRADE

La valutazione della qualità delle prove (2): *imprecision*

(Michela Cinquini)

GRADE re-clarification of the construct of certainty of evidence

ELSEVIER

CrossMark

Journal of Clinical Epidemiology 87 (2017) 4–13

Journal of Clinical Epidemiology

GRADE UPDATE OF PAPERS

The GRADE Working Group clarifies the construct of certainty of evidence

Monica Hultcrantz[a,b,*], David Rind[c,d], Elie A. Akl[e,f], Shaun Treweek[g], Reem A. Mustafa[e,h], Alfonso Iorio[e,i], Brian S. Alper[j,k], Joerg J. Meerpohl[l,m], M Hassan Murad[n], Mohammed T. Ansari[o], Srinivasa Vittal Katikireddi[p], Pernilla Östlund[a,q], Sofia Tranæus[a,q,r], Robin Christensen[s], Gerald Gartlehner[t,u], Jan Brozek[e,i], Ariel Izcovich[v], Holger Schünemann[e,i], Gordon Guyatt[e,i]

ELSEVIER

Check for updates

Journal of Clinical Epidemiology 137 (2021) 163–175

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

GRADE guidelines 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings

Linan Zeng[a,b,*], Romina Brignardello-Petersen[b], Monica Hultcrantz[c], Reed A.C. Siemieniuk[b], Nancy Santesso[b], Gregory Traversy[d], Ariel Izcovich[e], Behnam Sadeghirad[b,f], Paul E. Alexander[b], Tahira Devji[b], Bram Rochwerg[b,g], Mohammad H. Murad[b], Rebecca Morgan[b], Robin Christensen[i,j], Holger J. Schünemann[b,g], Gordon H. Guyatt[b,g]

LSEVIER

Journal of Clinical Epidemiology 150 (2022) 225–242

GRADE GUIDANCE SERIES

GRADE guidance 35: update on rating imprecision for assessing contextualized certainty of evidence and making decisions

Holger J. Schünemann[a,b,c,q,*], Ignacio Neumann[b,d], Monica Hultcrantz[e], Romina Brignardello-Petersen[b], Linan Zeng[b,f], M Hassan Murad[g], Ariel Izcovich[h],

In either guidelines or systematic reviews, when we rate the certainty of evidence, we are assessing our confidence where the point effects lies relative to particular threshold(s) of interest.

Box 2 Steps for rating imprecision using the partially contextualized approach

Step 1. Define your outcome as dichotomous or continuous
Step 2. For the body of evidence, set thresholds for absolute effects of health outcomes that correspond to small, moderate or large effects, both desirable and undesirable (note, that outcomes, e.g., mortality, can be a desirable outcome if it is reduced or undesirable if it is increased): use existing evidence for these thresholds from other decision-makers, consensus by relevant stakeholders, empirical evidence about thresholds that integrate the absolute effect size and the relative importance of the outcomes or the best guess by content experts if nothing else is available. It is important to state how you set your thresholds.
Step 3. Choose the target of the rating of certainty of evidence in relation to those thresholds for the point estimate of the absolute effect. That is decide if you rate if the effect lies between two thresholds (i.e., between small desirable and undesirable effects or between small and moderate or moderate and large) or beyond the threshold for large effects or decide if you rate the certainty that the effect is beyond or below one of the thresholds.
Step 4. Calculate the absolute effect estimates for the body of evidence for the outcome of interest including its confidence intervals based on the baseline risk and relative effect or if relevant the meta-analytic estimate of the risk difference across studies (e.g., if there are very small number of events in the studies).
Step 5. Determine how many thresholds the confidence interval crosses regardless of if the effect estimates suggest a desirable or undesirable health effect (do not count the "no effect" as a threshold).
Step 6. Rate down by as many levels as thresholds are crossed.
Optional Step 7. If the effect is large (i.e., the point estimate falls beyond the threshold for large effects) and if it is based on an apparently small number of events or participants, consider using the review information size (RIS) by calculating the required sample size for small, moderate or large effects for the outcome of interest to determine if further rating down is required because the RIS is not met (see calculator in this article). Otherwise, use the rating in step 6. If the effect appears to be trivial or none (i.e., the point estimate falls within the threshold of trivial or no effect to small desirable and undesirable effects), check if the RIS for trivial to no effects is met, for example, for the assessment of equivalence of interventions, to determine if further rating down is required. Otherwise, use the rating in step 5.

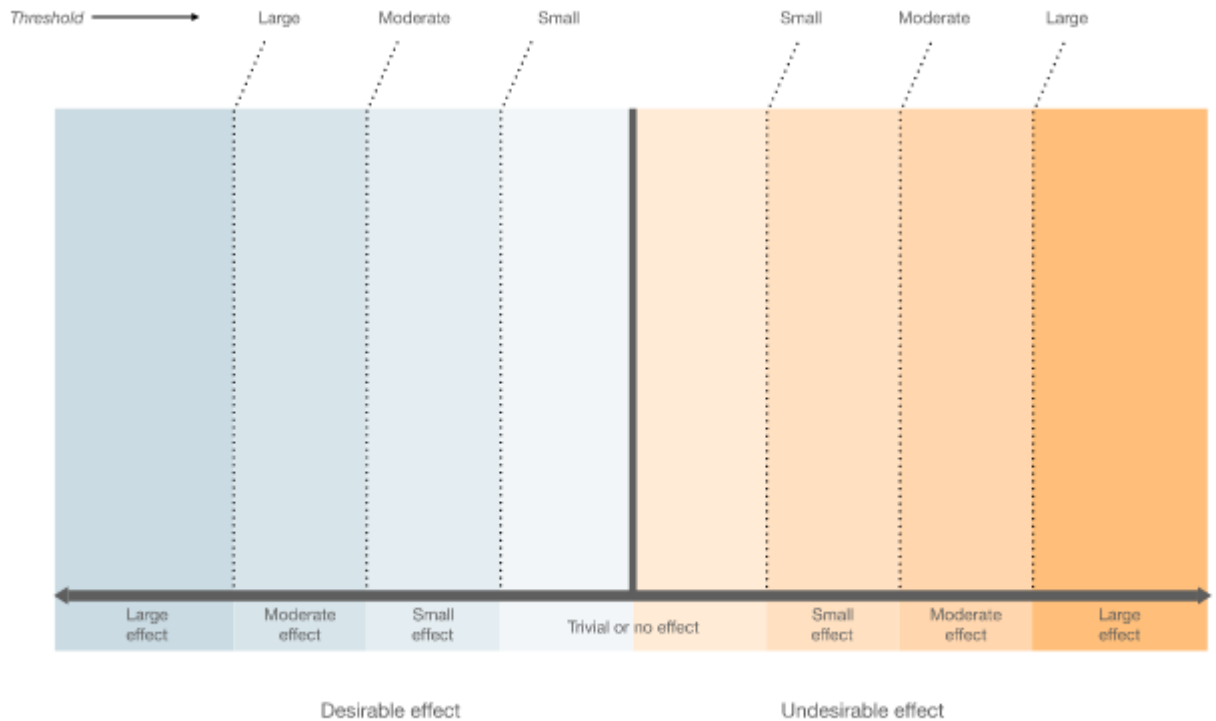
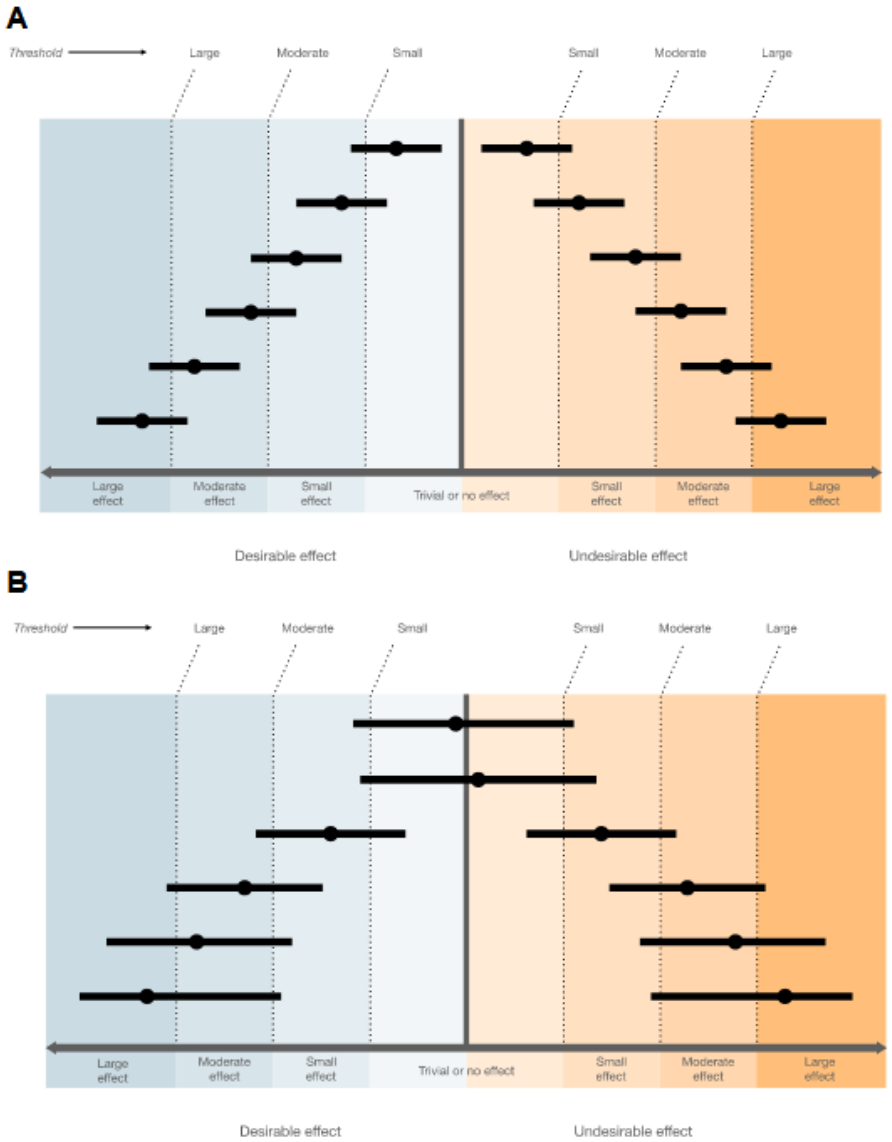


Fig. 1. Describes the thresholds and ranges for trivial, small, moderate and large effects.



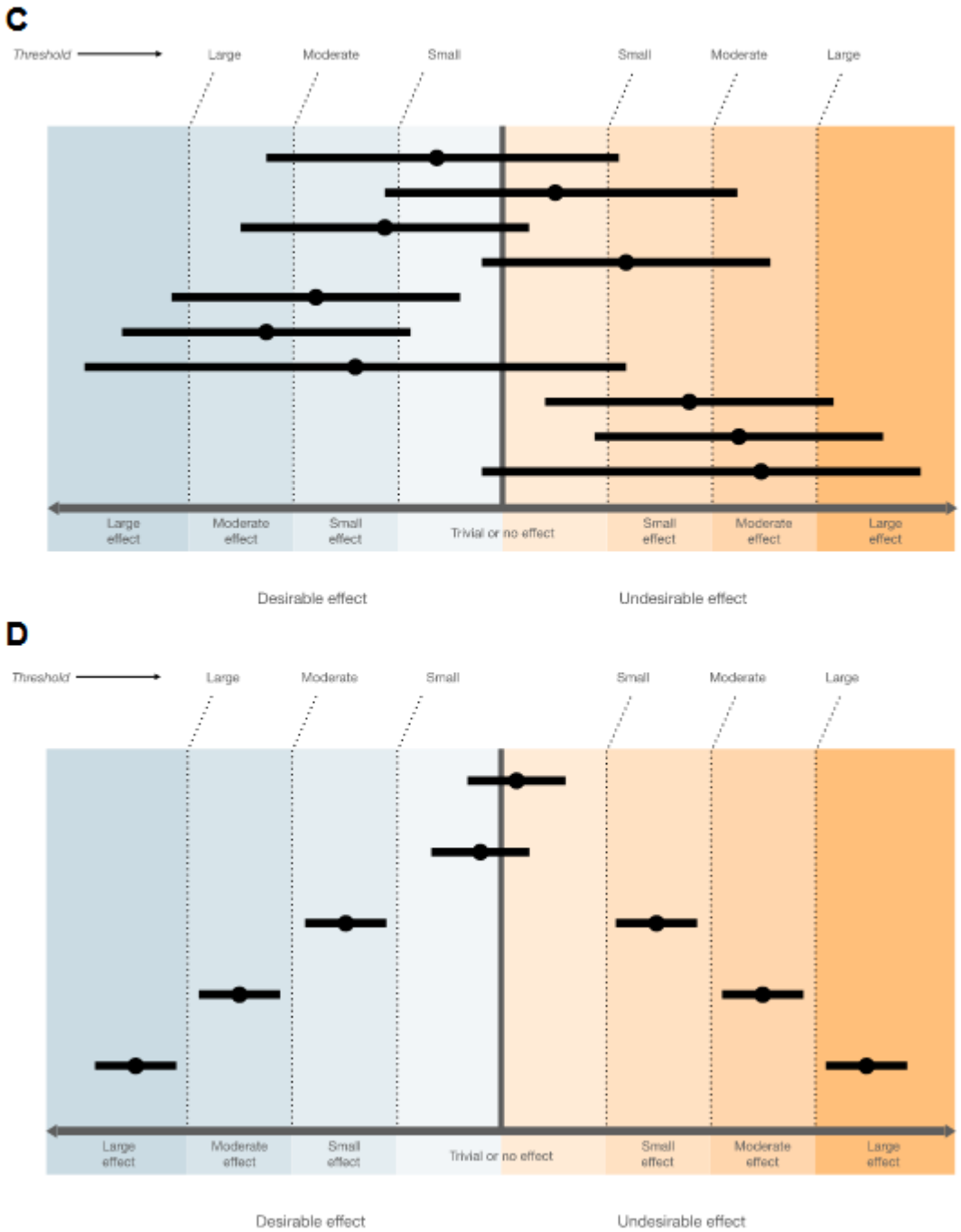
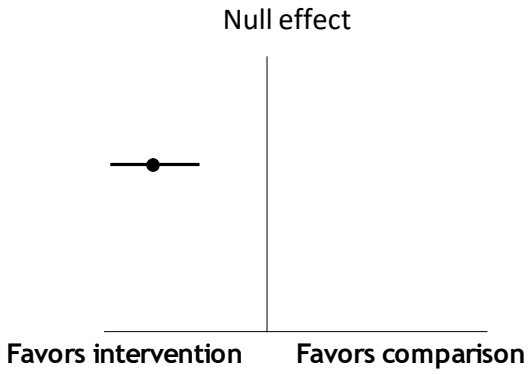
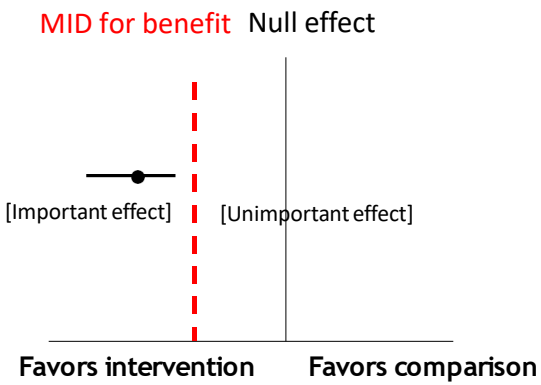


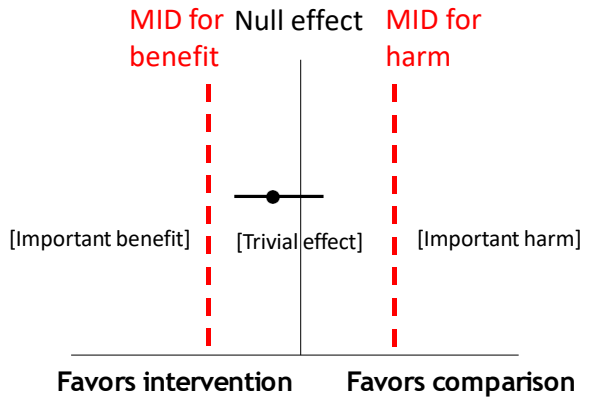
Fig. 2. Panel (A, B and C) describe a range of hypothetical scenarios to rating down for imprecision by one, two, and three levels respectivel panel (D) describes the conceptual approach to not rating down. Please note that for demonstration purposes the confidence intervals are s rical while for most scenarios the confidence intervals will be asymmetrical. Conceptually, one should rate down by one level when the CI absolute effect crosses one of the thresholds across the range of categories of magnitude of effect (in our example, there are six thresholds fo categories of magnitude of effect). One should rate down by two or three levels when the CI for the absolute effect crosses respectively two of the thresholds across the range of categories of magnitude of effect, respectively.

Threshold of interest, target of certainty of evidence rating in minimally contextualized approach

Threshold of interest	(a) Null effect	(b) Minimally important difference (MID)	(c) Range of trivial or no effect
Target of certainty rating	An effect is truly present or not.	An important effect is truly present or not.	A trivial or no effect is truly present or not.








DON'T TRY THIS AT HOME !

GRADE

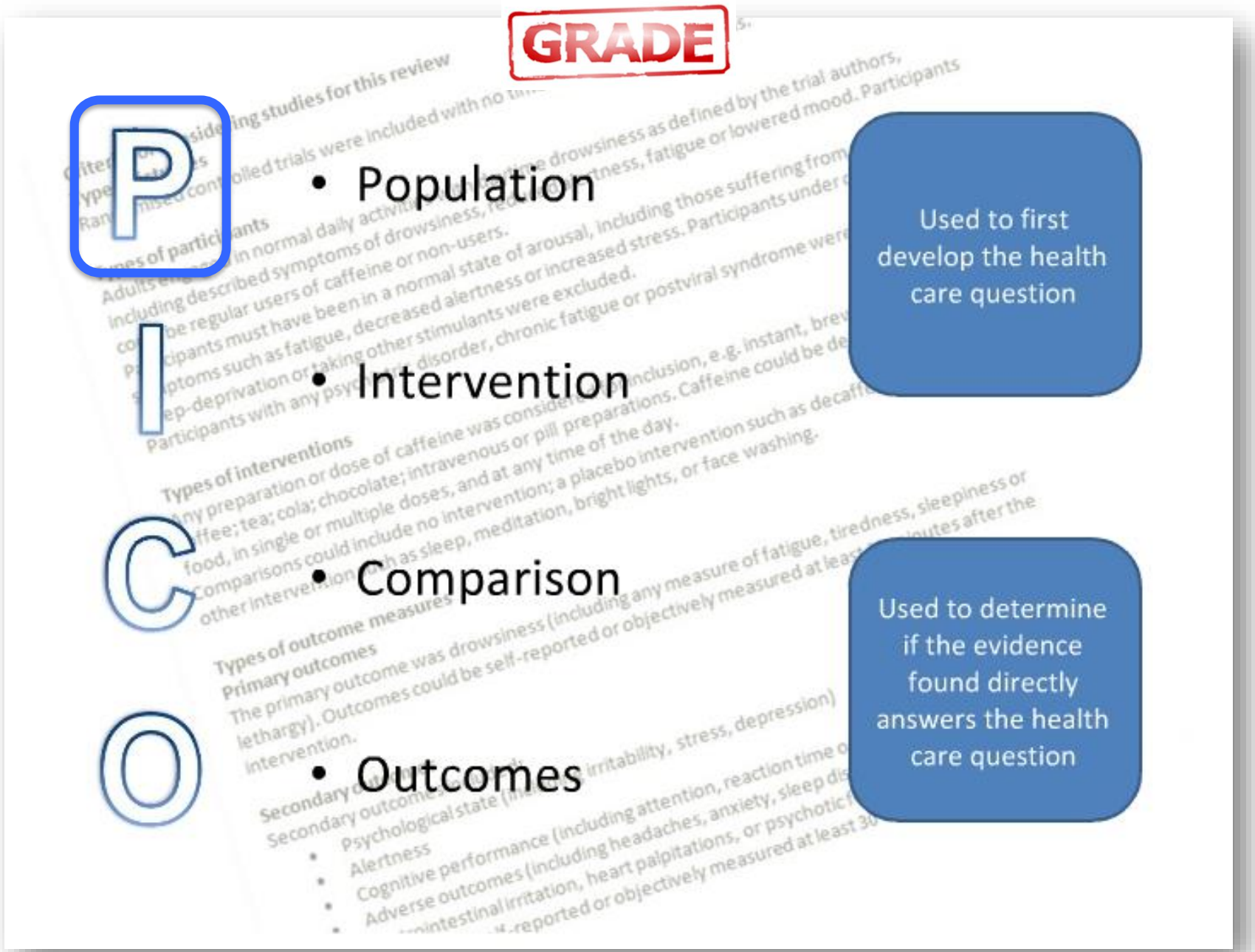
La valutazione della qualità delle prove (3): *indirectness*

(Giovanni Pappagallo)

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

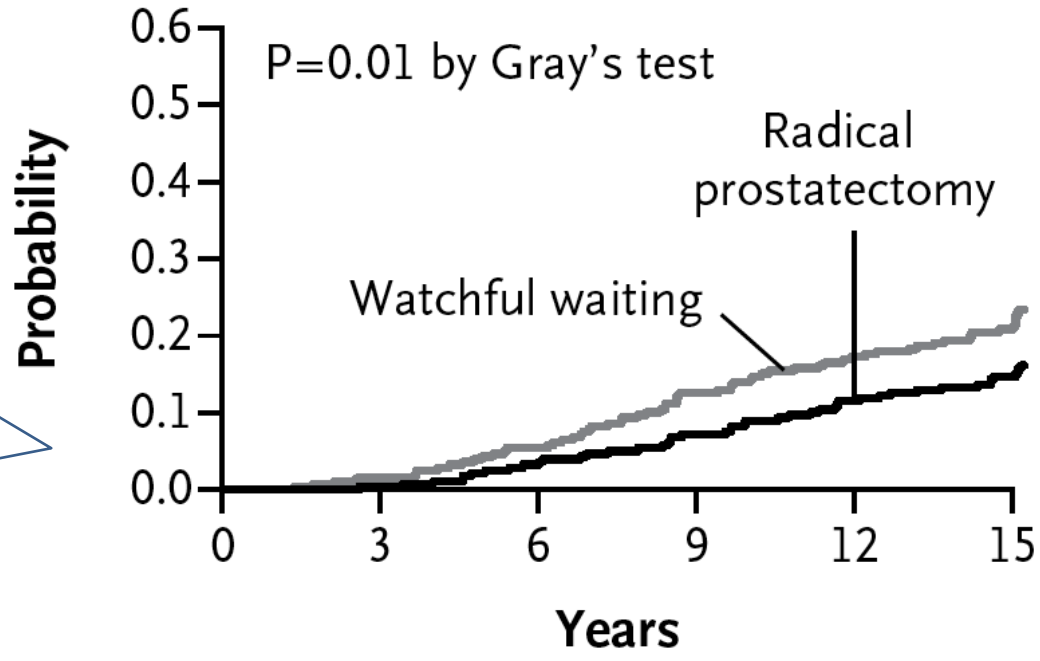

GRADE
P
• Population
I
• Intervention
C
• Comparison
O
• Outcomes
Used to first develop the health care question
Used to determine if the evidence found directly answers the health care question

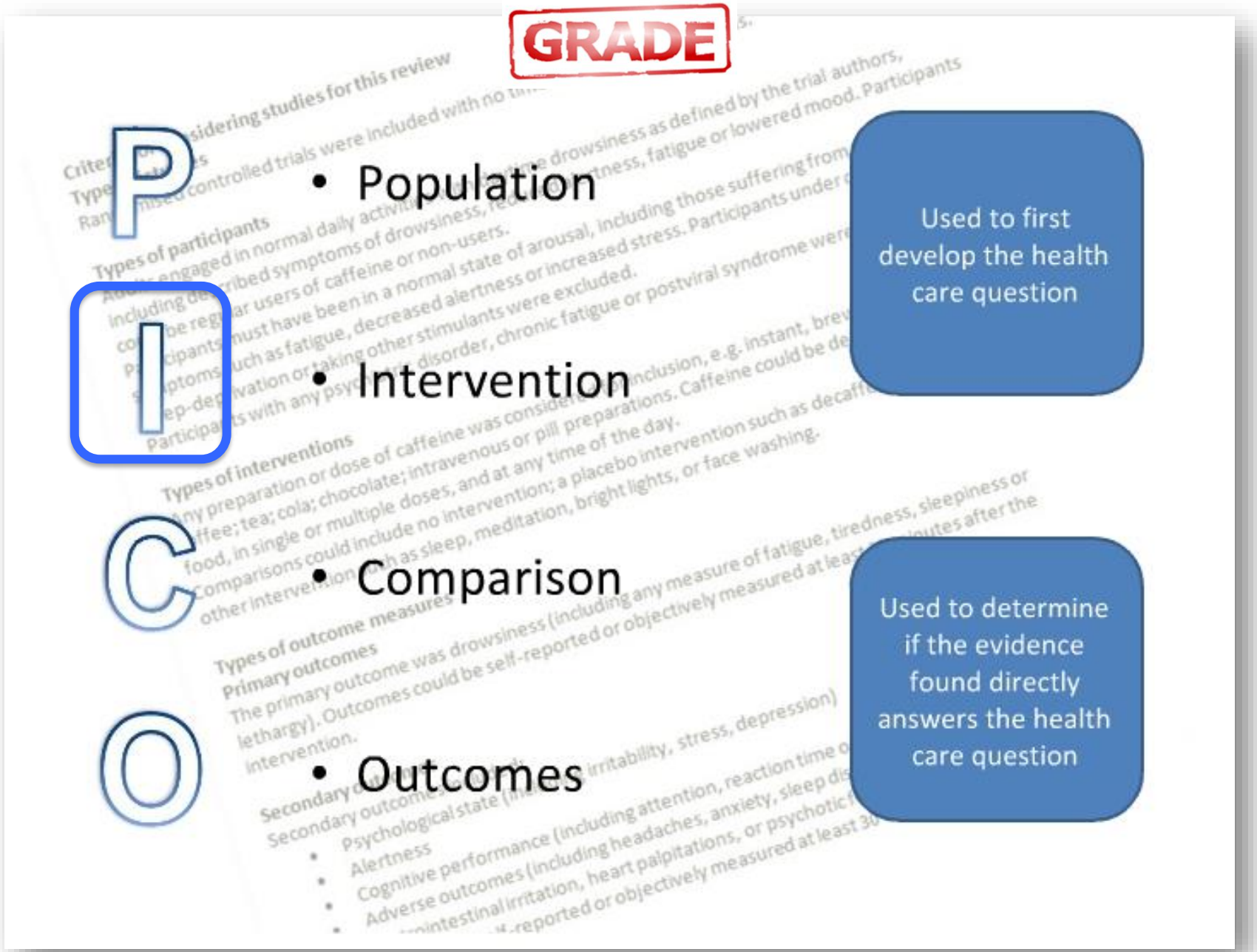
Radical Prostatectomy versus Watchful Waiting in Early Prostate Cancer

Anna Bill-Axelson, M.D., Ph.D., Lars Holmberg, M.D., Ph.D., Mirja Ruutu, M.D., Ph.D., Hans Garmo, Ph.D., Jennifer R. Stark, Sc.D., Christer Busch, M.D., Ph.D., Stig Nordling, M.D., Ph.D., Michael Häggman, M.D., Ph.D., Swen-Olof Andersson, M.D., Ph.D., Stefan Bratell, M.D., Ph.D., Anders Spångberg, M.D., Ph.D., Juni Palmgren, Ph.D., Gunnar Steineck, M.D., Ph.D., Hans-Olov Adami, M.D., Ph.D., and Jan-Erik Johansson, M.D., Ph.D., for the SPCG-4 Investigators*

N Engl J Med 2011;364:1708-17.

Pazienti in era pre-PSA




GRADE
P
• Population
I
• Intervention
C
• Comparison
O
• Outcomes
Used to first develop the health care question
Used to determine if the evidence found directly answers the health care question

Long-term oncologic outcomes of postoperative adjuvant versus salvage radiotherapy in prostate cancer: Systemic review and meta-analysis of 5-year and 10-year follow-up data

Ja Yoon Ku[1], Chan Ho Lee[1], Hong Koo Ha[1,2]

Korean J Urol 2015;56:735-741.

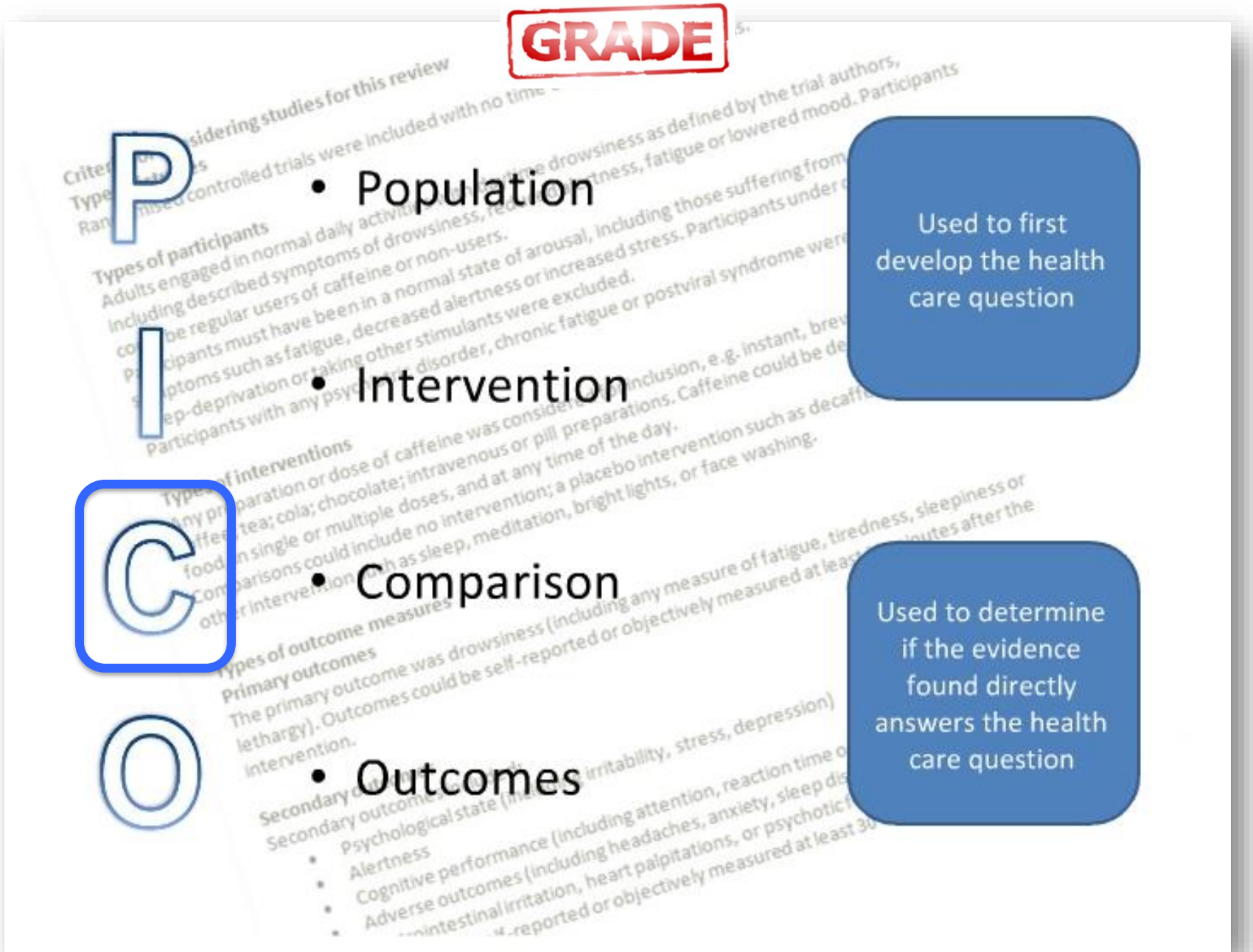
The present systemic review has the following limitations that must be taken into account.

The first limitation is that we heterogeneously recruit randomized controlled studies and retrospective studies: in retrospective studies, the initiation timing of radiotherapy is somewhat different in each study.

The second limitation is that there is a difference in dose and modality (2-dimensional, 3-dimensional, or intensity modulated radiation therapy) of radiation compared to recent practice, which may alter oncologic outcomes.

The third limitation [illegible] were different in eac[illegible] of long-term outcome be[illegible]

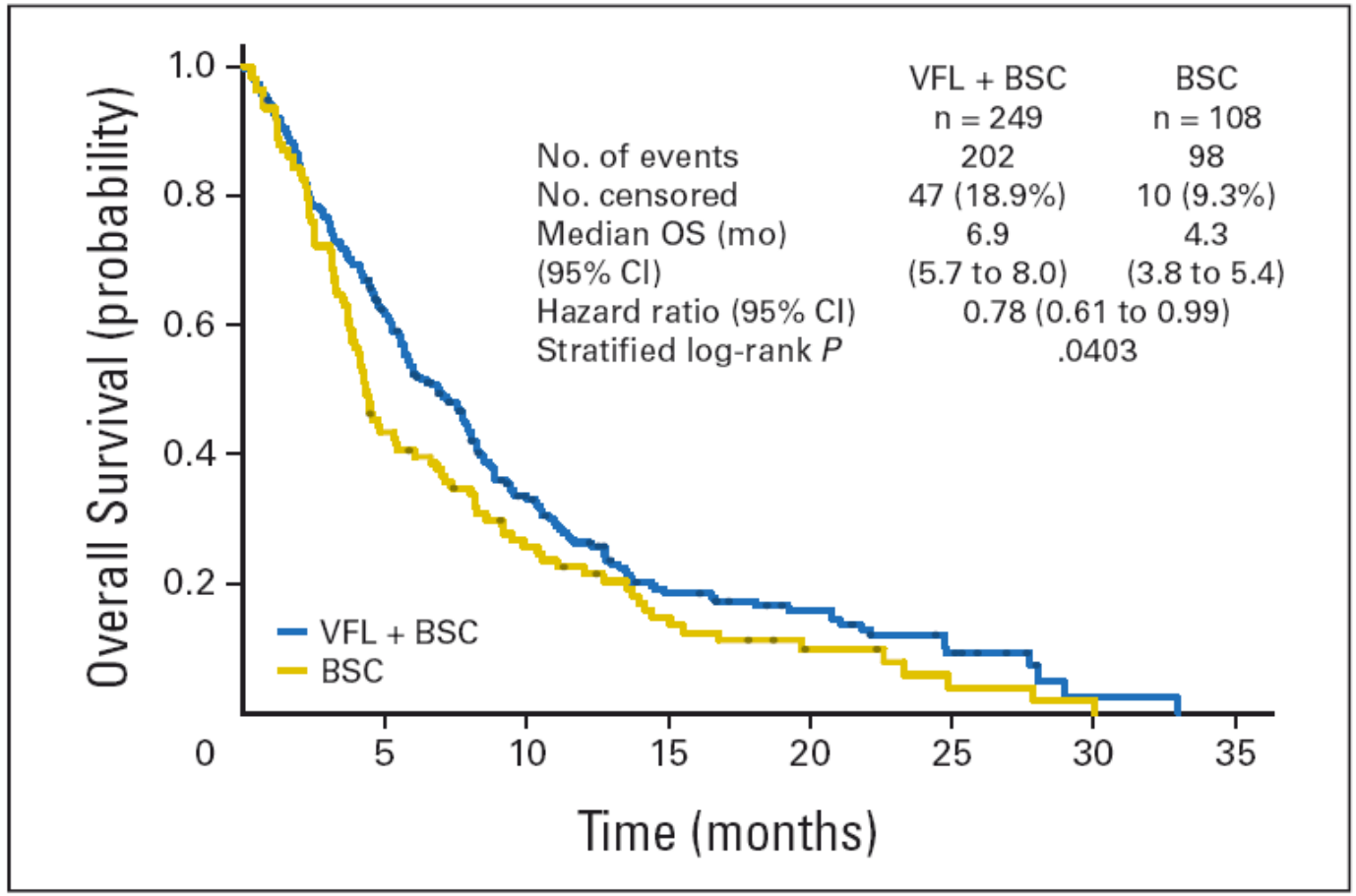
Indirectness per I. (di P.I.C.O.)

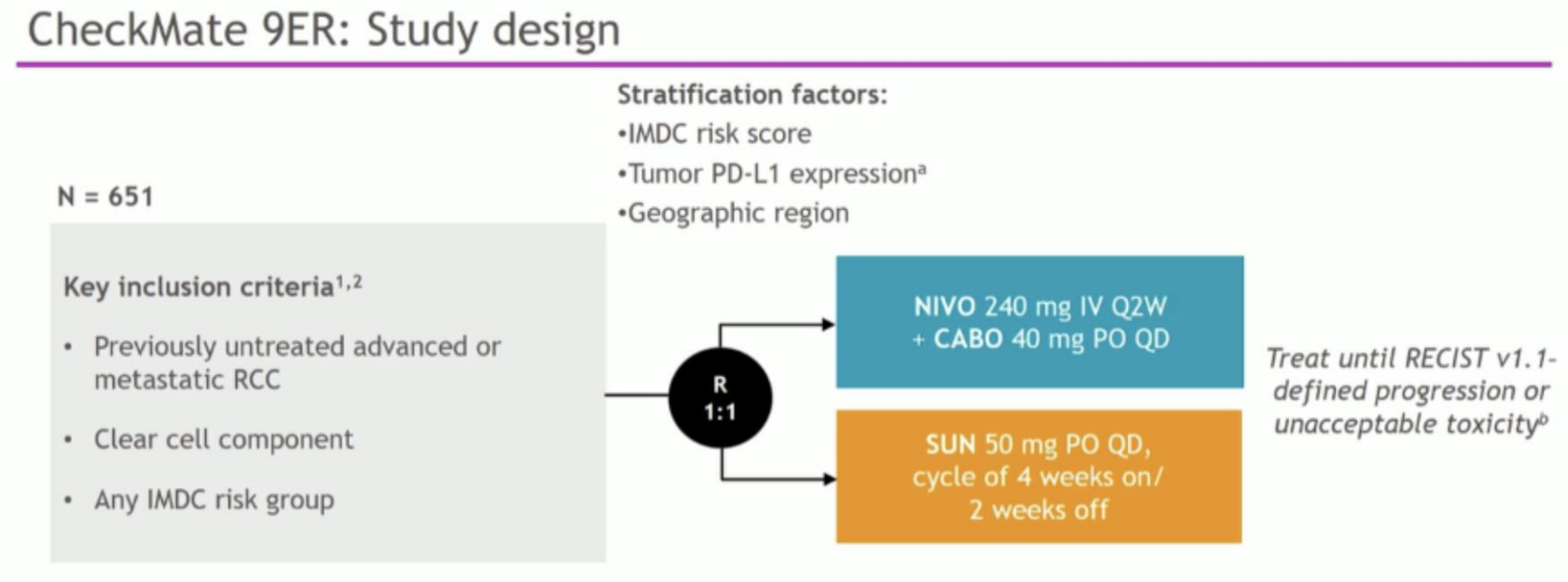

GRADE
P
I
C
O
• Population
• Intervention
• Comparison
• Outcomes
Used to first develop the health care question
Used to determine if the evidence found directly answers the health care question

Phase III Trial of Vinflunine Plus Best Supportive Care Compared With Best Supportive Care Alone After a Platinum-Containing Regimen in Patients With Advanced Transitional Cell Carcinoma of the Urothelial Tract

Joaquim Bellmunt, Christine Théodore, Tomasz Demkov, Boris Komyakov, Lisa Sengelov, Gedske Daugaard, Armelle Caty, Joan Carles, Agnieszka Jagiello-Gruszfeld, Oleg Karyakin, François-Michel Delgado, Patrick Hurteloup, Eric Winquist, Nassim Morsli, Yacine Salhi, Stéphane Culine, and Hans von der Maase

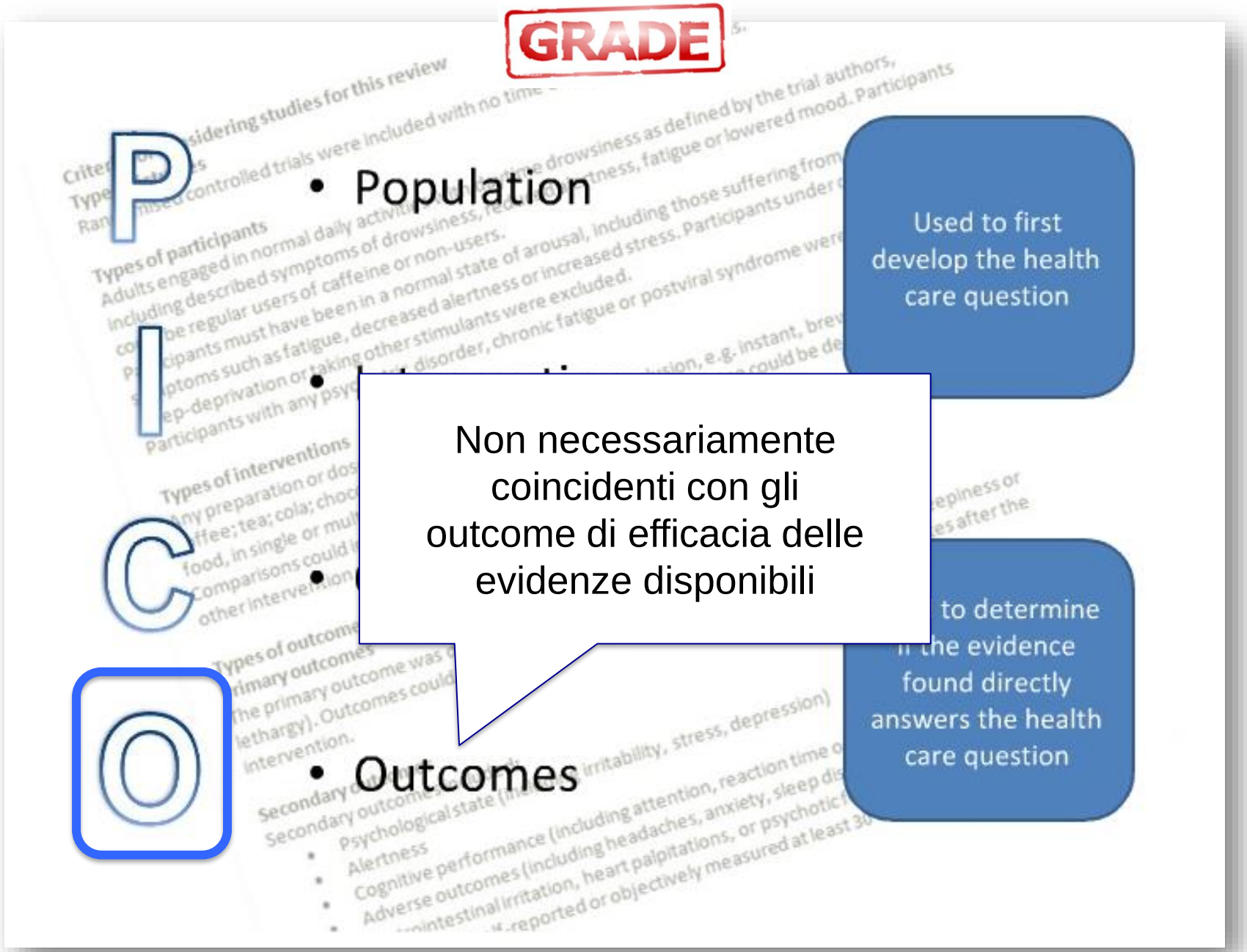
J Clin Oncol 27:4454-4461. © 2009 by American Society of Clinical Oncology



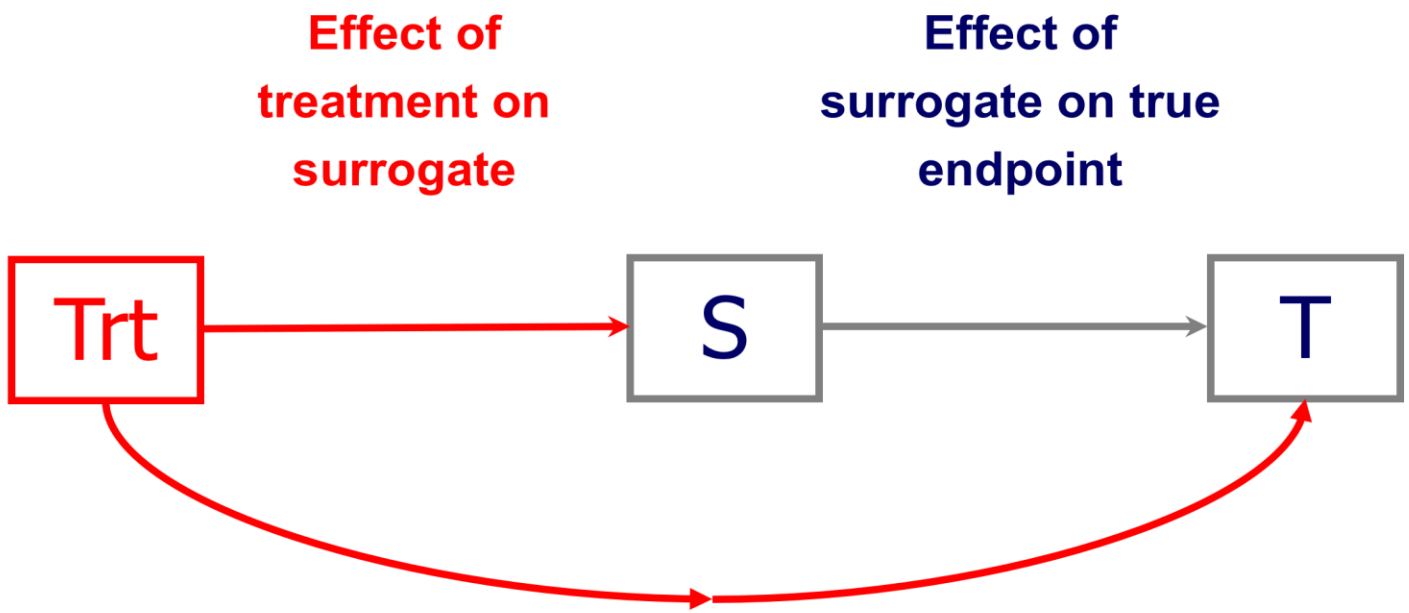


Pembrolizumab
Axitinib

Certezza globale delle prove	Raccomandazione clinica	Forza della raccomandazione clinica
BASSA	Nei pazienti affetti da carcinoma renale metastatico variante istologica a cellule chiare, rischio intermedio-alto sec. Heng, il trattamento di prima linea con pembrolizumab-axitinib dovrebbe essere preso in considerazione come approccio terapeutico di prima scelta	**Forte a Favore**


GRADE
P
I
C
O
• Population
• Outcomes
Used to first develop the health care question
Non necessariamente coincidenti con gli outcome di efficacia delle evidenze disponibili
to determine if the evidence found directly answers the health care question

VALIDATION OF SURROGATE ENDPOINTS: "FULL CAPTURE OF EFFECT"



Effect of treatment on true endpoint must be fully captured by surrogate

*Prentice, Statist Med 1989;**8**:431.*

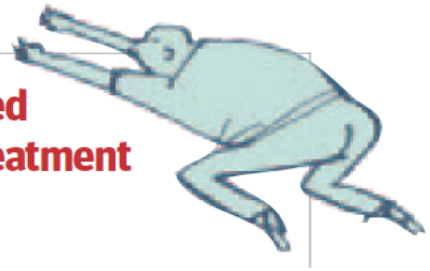
THE IDOLATRY OF THE SURROGATE

Easier to measure surrogate outcomes are often used instead of patient important outcomes such as death, quality of life, or functional capacity when assessing treatments. **John Yudkin, Kasia Lipska, and Victor Montori** argue that our obsession with surrogates is damaging patient care

BMJ | 14 JANUARY 2012 | VOLUME 344

Since they respond sooner than outcomes that are important to patients, surrogates are better suited as end points in clinical trials that need to be completed quickly and at low cost

Surrogate markers are not intrinsically flawed. When interpreted appropriately, they can be helpful in risk stratification and in treatment



In order to fully engage our patients in treatment decisions, we must understand how therapies affect outcomes that are important to them. Surrogate end-points will not provide us with these answers.

SACRO CUORE DON CALABRIA

ECI CANCER CENTRE

I R C C S

SCUOLA DI METODOLOGIA DELLA RICERCA CLINICA

2025 - 11ª EDIZIONE

MODULI SPECIALISTICI - S4

VENERDÌ 16 - SABATO 17 MAGGIO 2025

NEGRAR DI VALPOLICELLA (VR)

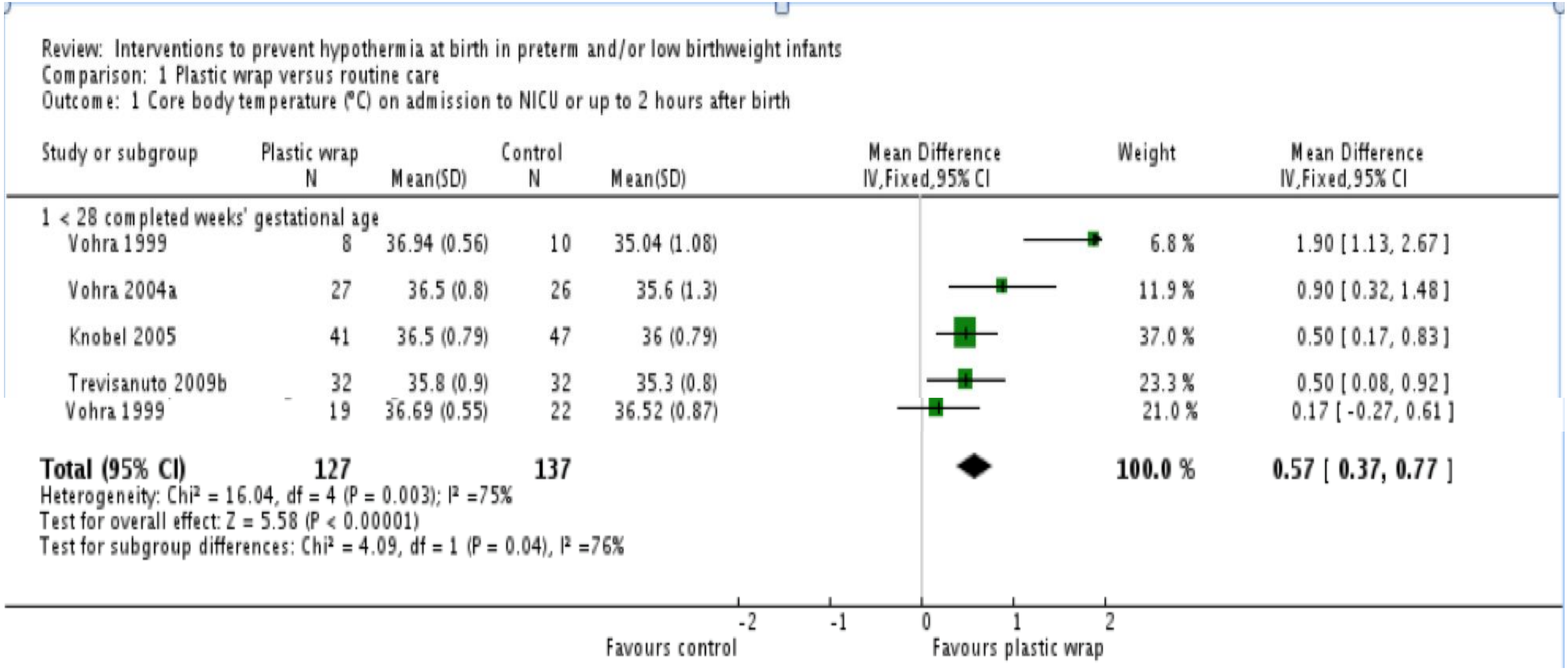
Centro Formazione IRCCS "Sacro Cuore - Don Calabria"

GRADE

La valutazione della qualità delle prove (4): *inconsistency, publication bias*

(Michela Cinquini)

Esempio di Metaview



What is heterogeneity?

Differences between studies with respect to:

Clinical heterogeneity (clinical diversity)

- *Participants*

  e.g. conditions under investigation, eligibility criteria for trials, geographical variation

- *Interventions*

  e.g. intensity / dose / duration, sub-type of drug, mode of administration, experience of practitioners, nature of the control (placebo/none/standard care)

- *Outcomes*

  e.g. definition of an event, follow-up duration, ways of measuring outcomes, cut-off points on scales

What is heterogeneity?

Differences between studies with respect to:

Methodological heterogeneity (methodological diversity)

- *Design*

  e.g. randomised vs non-randomised,
  crossover vs parallel group vs cluster randomised,
  pre-test and long follow up

- *Conduct*

  e.g. allocation concealment, blinding etc,
  approach to analysis, imputation methods for
  missing data

What is heterogeneity?

What do we do if there ***is*** statistical heterogeneity?

Variation in the *true effects* underlying the studies

...which may manifest itself in more observed variation than expected by chance alone

May be due to clinical diversity (different treatment effects) or methodological diversity (different biases)

Inconsistency is important only when it reduces confidence in results in relation to a particular decision.

Come si misura questa eterogeneità?

Confidence interval overlapping Eyeball test

Cochran's Q: to assess whether observed differences in results are compatible with change alone

χ^2 distribution; low power (small number of studies, small sample size)

p=<0.10 (heterogeneity)

I^2 quantifing heterogeneity (describes the percentage of variation across studies that is due to heterogeneity rather than chance)

0-40% might not be important

30-60% may represent moderate heterogeneity

50-90% may represent substantial heterogeneity

75-100% considerable heterogeneity

Tau….

Unexplained heterogeneity

Differenza fra effetto grande e piccolo.

Non importante se anche l''effetto piccolo è clinicamente significativo.

Rilevante se ci sono differenze clinicamente rilevanti (impatto sul paziente) fra effetto piccolo e effetto grande

Downgrade: dipende

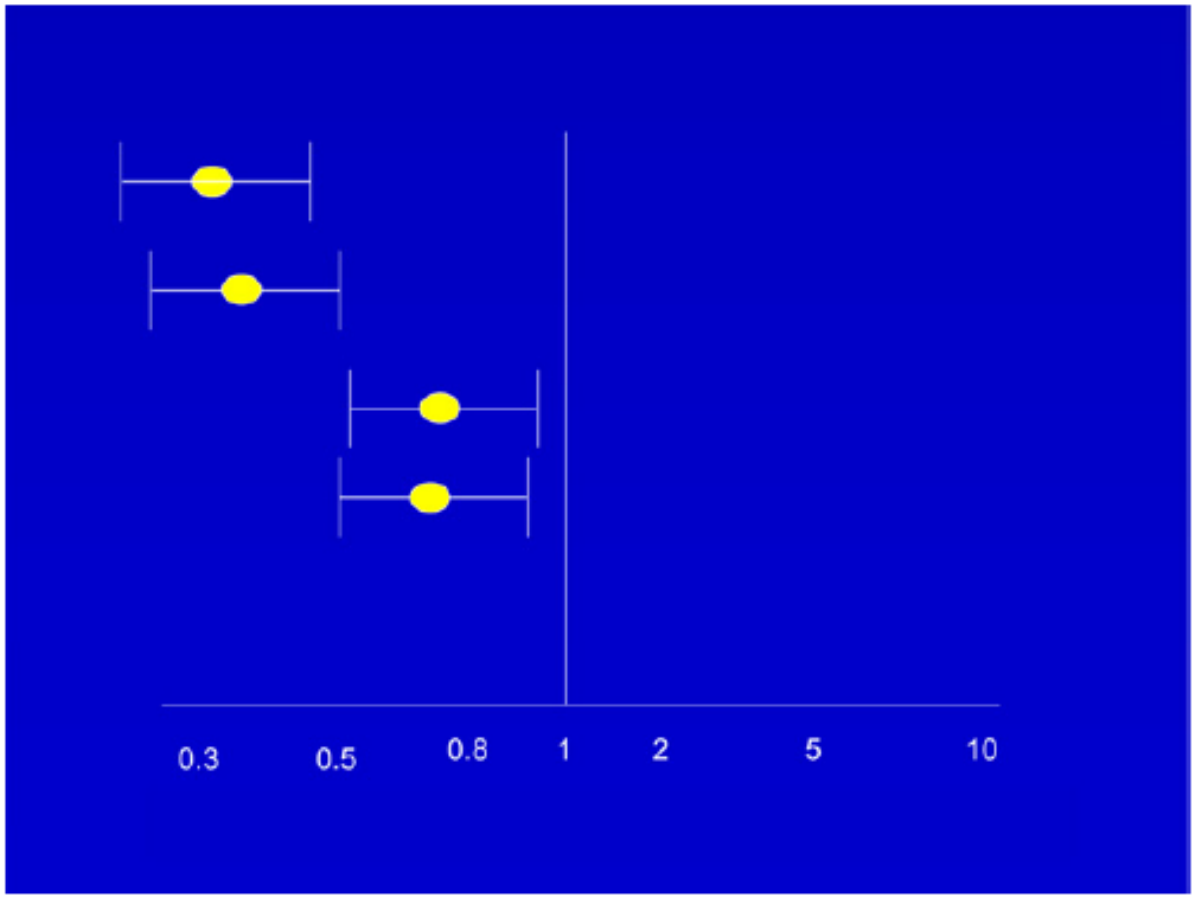


Fig. 2. Substantial heterogeneity, but of questionable importance.

Unexplained heterogeneity

La grandezza della variabilità è la stessa ma in questo caso due studi vanno in una direzione e due in un'altra.

Inconsistency importante

Pooled estimate di non effetto ma con grande eterogeneità

Downgrade: sì

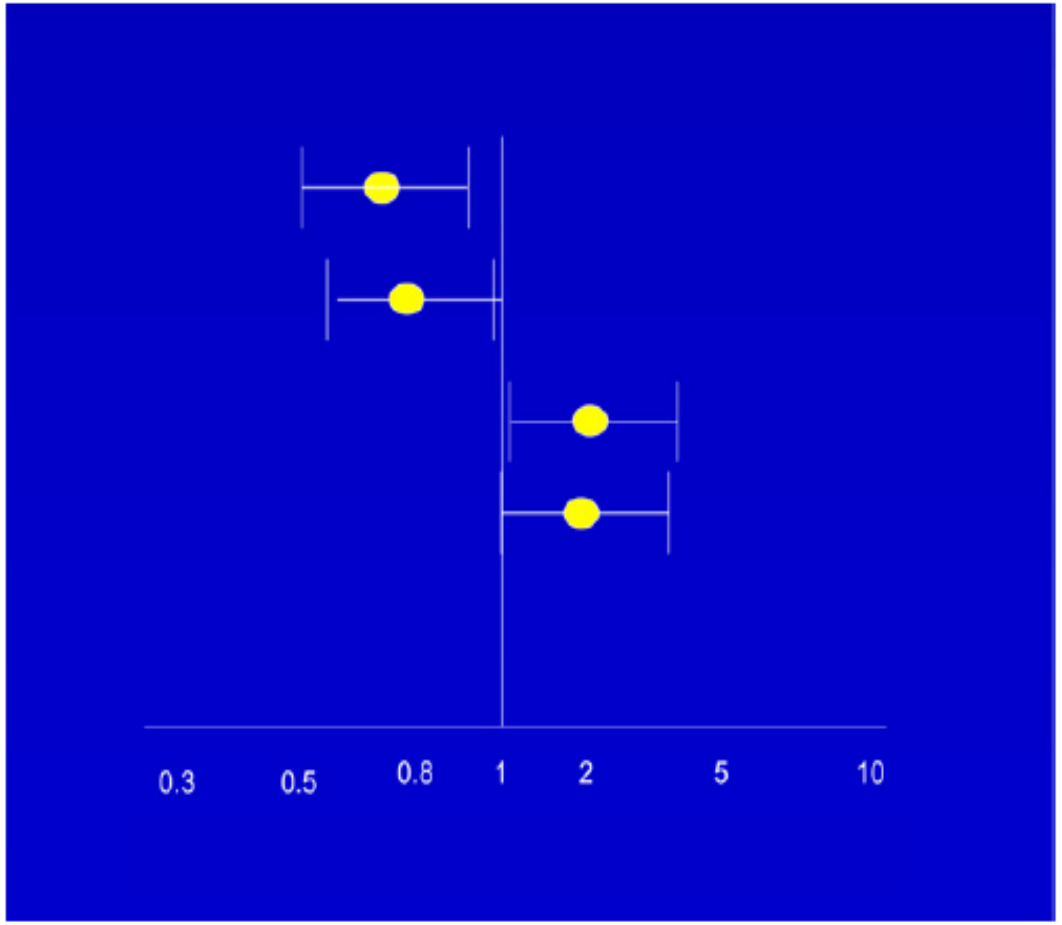


Fig. 3. Substantial heterogeneity, of unequivocal importance.

Unexplained heterogeneity

Pooled estimate di non effetto, come prima, ma in questo caso le differenze fra gli studi sono piccole , tutti concludono per differenze piccole e non significative

Downgrade: no

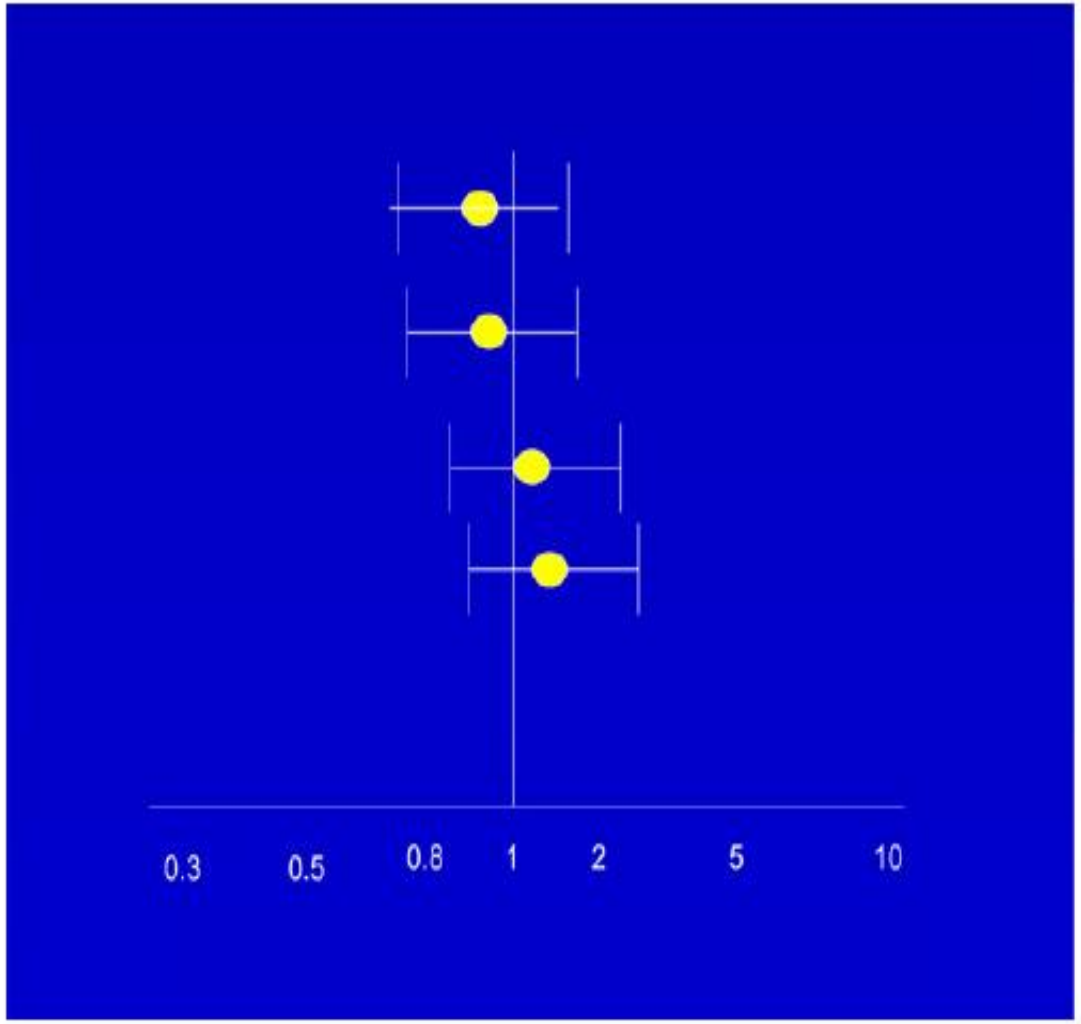


Fig. 1. Differences in direction, but minimal heterogeneity.

Publication bias

- Selective reporting bias: outcome dichiarato nel protocollo ma risultati non riportati: uno dei domini del risk of bias
- Publication bias: mancata pubblicazione di un intero studio

✓English language bias

✓Time lag bias

✓Citation bias

- Sovrastima dell'effetto del trattamento

Publication bias in context

Publication bias and other related biases can be summarised as statistically significant, 'positive' results being:

- more likely to be published (publication bias)
- more likely to be published rapidly (time lag bias)
- more likely to be published in English (language bias)
- more likely to be published more than once (multiple publication bias)
- more likely to be cited by others (citation bias)

[The Cochrane Collaboration]

Publication bias: come valutarne il rischio?

Sezione metodi della revisione:

Cercati studi non pubblicati?

Solo studi in lingua inglese?

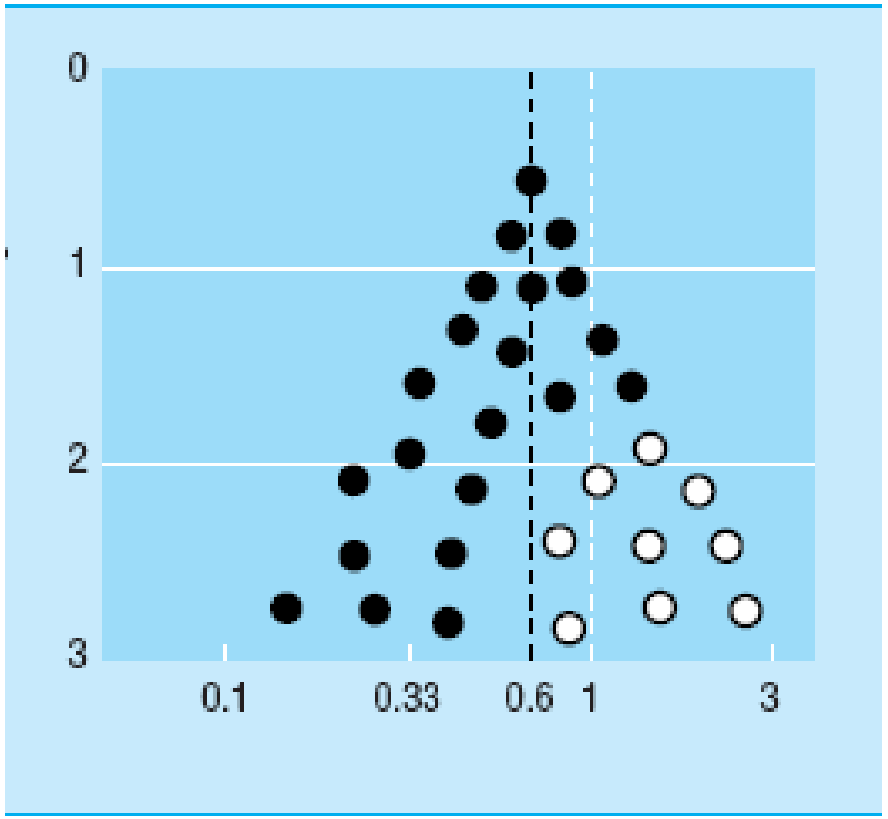
Revisione su nuova terapia introdotta da poco?

Trials piccoli e sponsorizzati da industria farmaceutica?

Funnel plot?

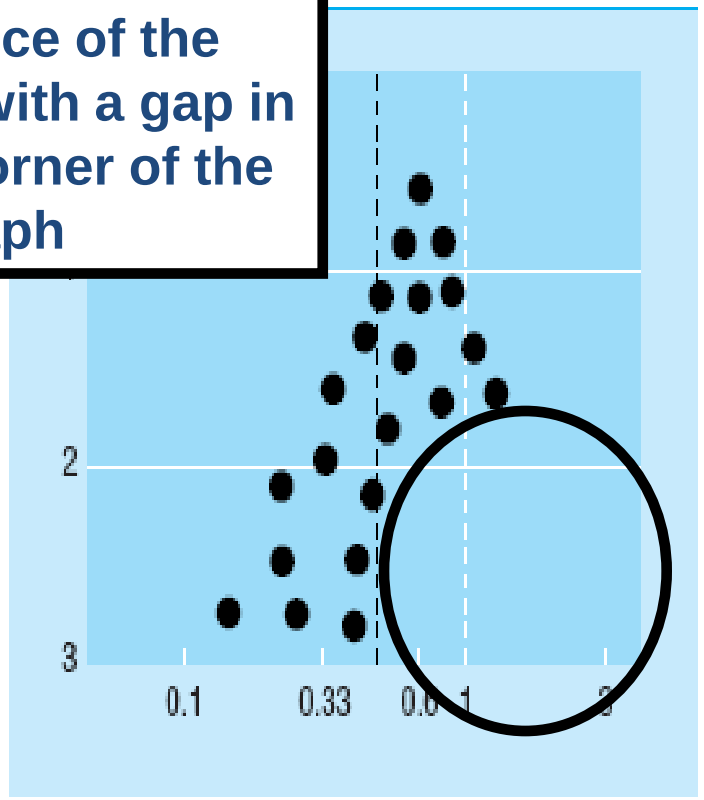
Studi osservazionali? Rischio maggiore perché non obbligatoria registrazione

Publication of All Trials

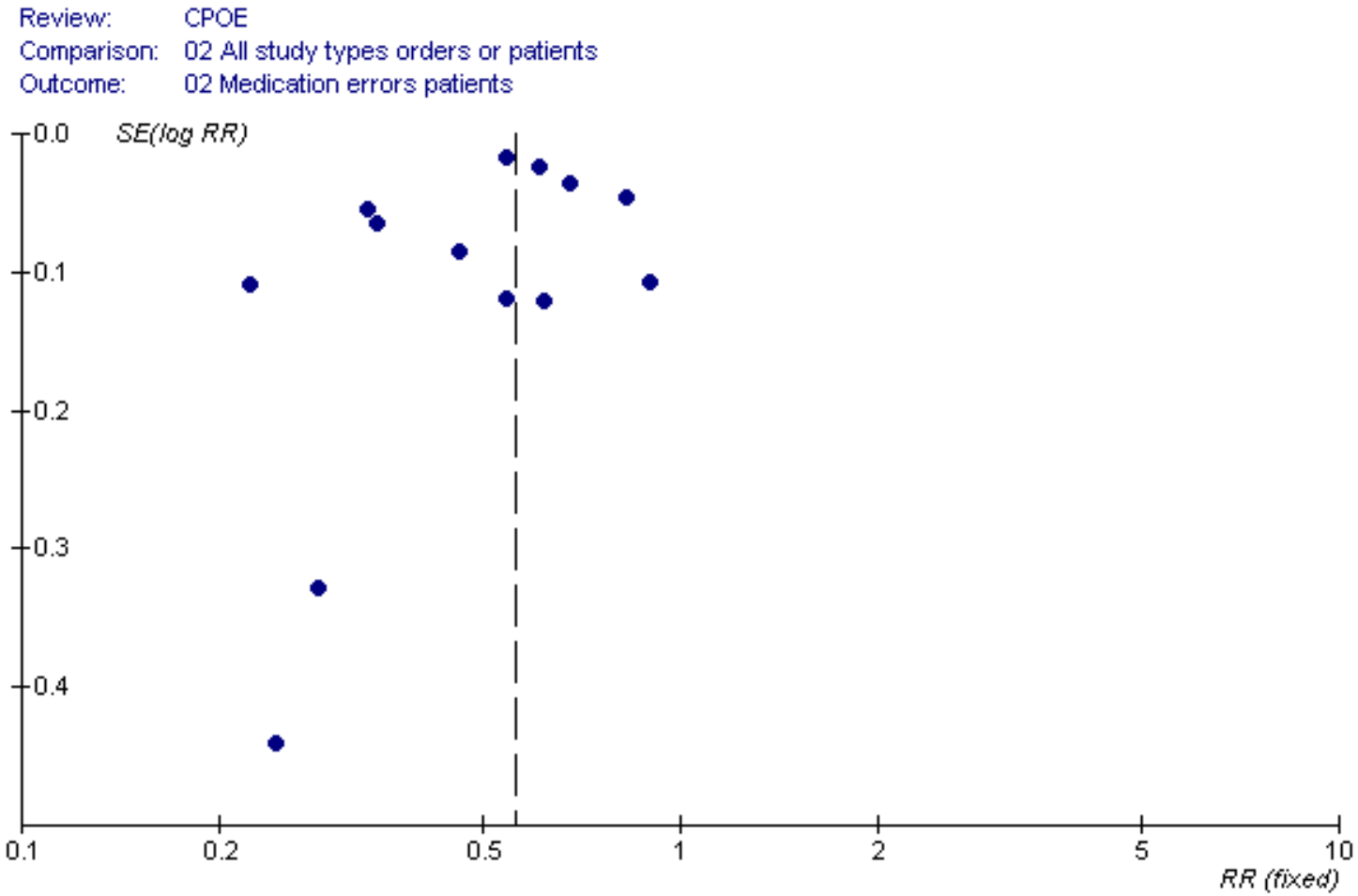


Publication Bias

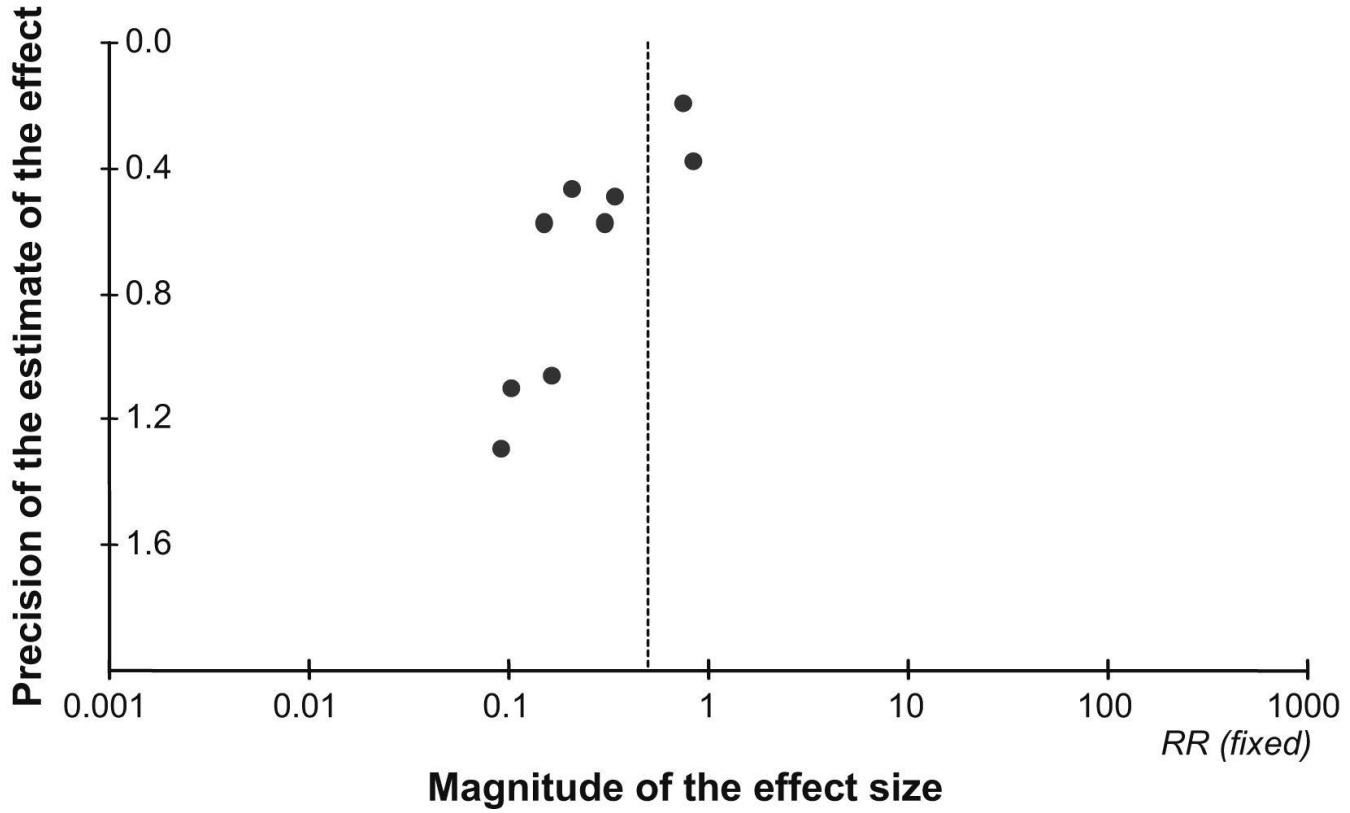
Asymmetrical appearance of the funnel plot with a gap in a bottom corner of the graph



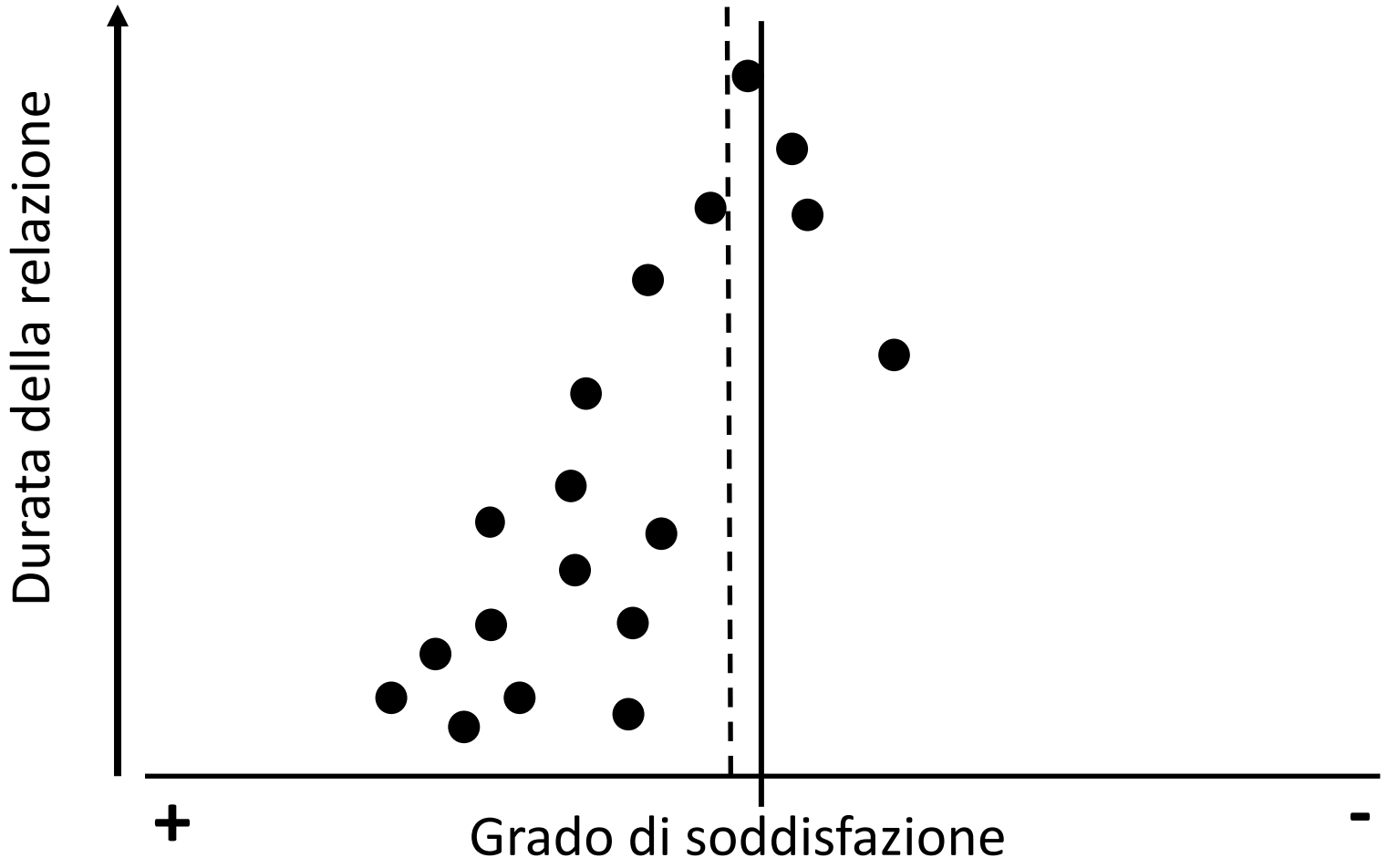
Study 2: Effects of CPOE on medication errors / ADEs



Funnel plot of studies of flavonoids for ameliorating symptoms in patients with hemorrhoids



Nella vita affettivo-relazionale di GLP c'è un evidente bias di "pubblicazione"



G.L. Pappagallo, personal experience, 2023

SCUOLA DI METODOLOGIA
DELLA RICERCA CLINICA

2024 - 10ª EDIZIONE

MODULI SPECIALISTICI - S2

VENERDÌ 12 - SABATO 13 APRILE 2024

NEGRAR DI VALPOLICELLA (VR)

Centro Formazione IRCCS "Sacro Cuore - Don Calabria"

GRADE

La valutazione della qualità delle prove (5): gli studi non randomizzati

(Michela Cinquini)

Observational studies

Quality starts as low

Then it can be further downgraded for one of the reasons cited above (risk of bias, inconsistency, imprecision, indirectness, publication bias)

Then it can be upgraded ;

The circumstances under which we may wish to rate up the quality of evidence for intervention studies will likely occur infrequently and are primarily relevant to observational studies.

Indeed, although it is theoretically possible to rate up results from randomized control trials (RCTs), we have yet to find a compelling example of such an instance.

What can raise quality?

1. **large magnitude of effect** can upgrade (**RRR 50%/RR 2**)
   - very large two levels **(**RRR 80%/RR 5) ; modeling studies suggests that confounding (from nonrandom allocation) alone is unlikely to explain associations with a relative risk (RR) greater than 2 (or less than 0.5), and very unlikely to explain associations with an RR greater than 5 (or less than 0.2)
   - Es: relationship between infant sleeping position and sudden infant death syndrome (SIDS) found an odds ratio (OR) of 4.1 (95% confidence interval [CI]: 3.1, 5.5) of SIDS occurring with front vs. back sleeping positions

Large magnitude of effect

Further consideration for rating up:

- rapidity of treatment response, and the previous underlying trajectory of the condition

E.g.: we feel confident that hip replacement has a large effect on reducing pain and functional limitations in severe osteoarthritis not only because of the size of the treatment response, but because the natural history of hip osteoarthritis is a progressive deterioration that surgery rapidly and uniformly reverses

- indirect evidence that provides further support for large treatment effects.

E.g. : the effectiveness of antibiotic prophylaxis in a variety of other situations supports observational studies that suggest that antibiotic prophylaxis results in an 89% RR reduction in meningococcal disease in contacts of patients who have suffered the illness

What can raise quality?

2. dose response relation

e.g.: childhood lymphoblastic leukemia: risk for CNS malignancies 15 years after cranial irradiation

- no radiation: 1% (95% CI 0% to 2.1%)
- 12 Gy: 1.6% (95% CI 0% to 3.4%)
- 18 Gy: 3.3% (95% CI 0.9% to 5.6%)

E.g.: systematic review of observational studies on cyclooxygenase-2 inhibitors on cardiovascular events:

- RR 1.33 (95% CI: 1.00, 1.79) with doses less than 25 mg/d
- RR 2.19 (95% CI: 1.64, 2.91) with doses more than 25 mg/d

Residual confounding

- **3.** all **plausible residual confounding** may be working to reduce the demonstrated effect or increase the effect if no effect was observed (underestimate of the treatment effect)

- Es: effect of condom use on HIV infection among men who have sex with men RR: 0.34 [0.21, 0.54] (RRR: 66%) in favor of condom use compared with no condom use. Condom users were more likely to have more partners (but studies did not adjust for this confounding factor in their analyses). Considering the number of partners would, if anything, strengthen the effect estimate in favor of condom use.

Residual confounding

- Es: SR di studi osservazionali: mortalità per tutte le cause in ospedali privati for profit vs ospedali privati not – for profit: RR 1.020, 95% CI 1.003-1.038 : mortalità leggermente superiore in for profit
- Fattori confondenti per cui non hanno aggiustato: livello di reddito dei pazienti e disponibilità risorse ospedali (> in for profit) che dovrebbe essere associato a < mortalità. Quindi in caso di aggiustamento l'effetto (RR) sarebbe stato maggiore

SCUOLA DI METODOLOGIA
DELLA RICERCA CLINICA

2024 - 10ª EDIZIONE

MODULI SPECIALISTICI - S2

VENERDÌ 12 - SABATO 13 APRILE 2024

NEGRAR DI VALPOLICELLA (VR)

Centro Formazione IRCCS "Sacro Cuore - Don Calabria"

GRADE

Evidence to Decision Framework e formulazione della Raccomandazione

(Ivan Moschetti)

The Evidence-to-Decision framework

GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction

Pablo Alonso-Coello,[1,2] Holger J Schünemann,[2,3] Jenny Moberg,[4] Romina Brignardello-Petersen,[2,5] Elie A Akl,[2,6] Marina Davoli,[7] Shaun Treweek,[8] Reem A Mustafa,[2,9] Gabriel Rada,[10,11,12] Sarah Rosenbaum,[4] Angela Morelli,[4] Gordon H Guyatt,[2,3] Andrew D Oxman[4] the GRADE Working Group

BMJ 2016;353:i2016

GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines

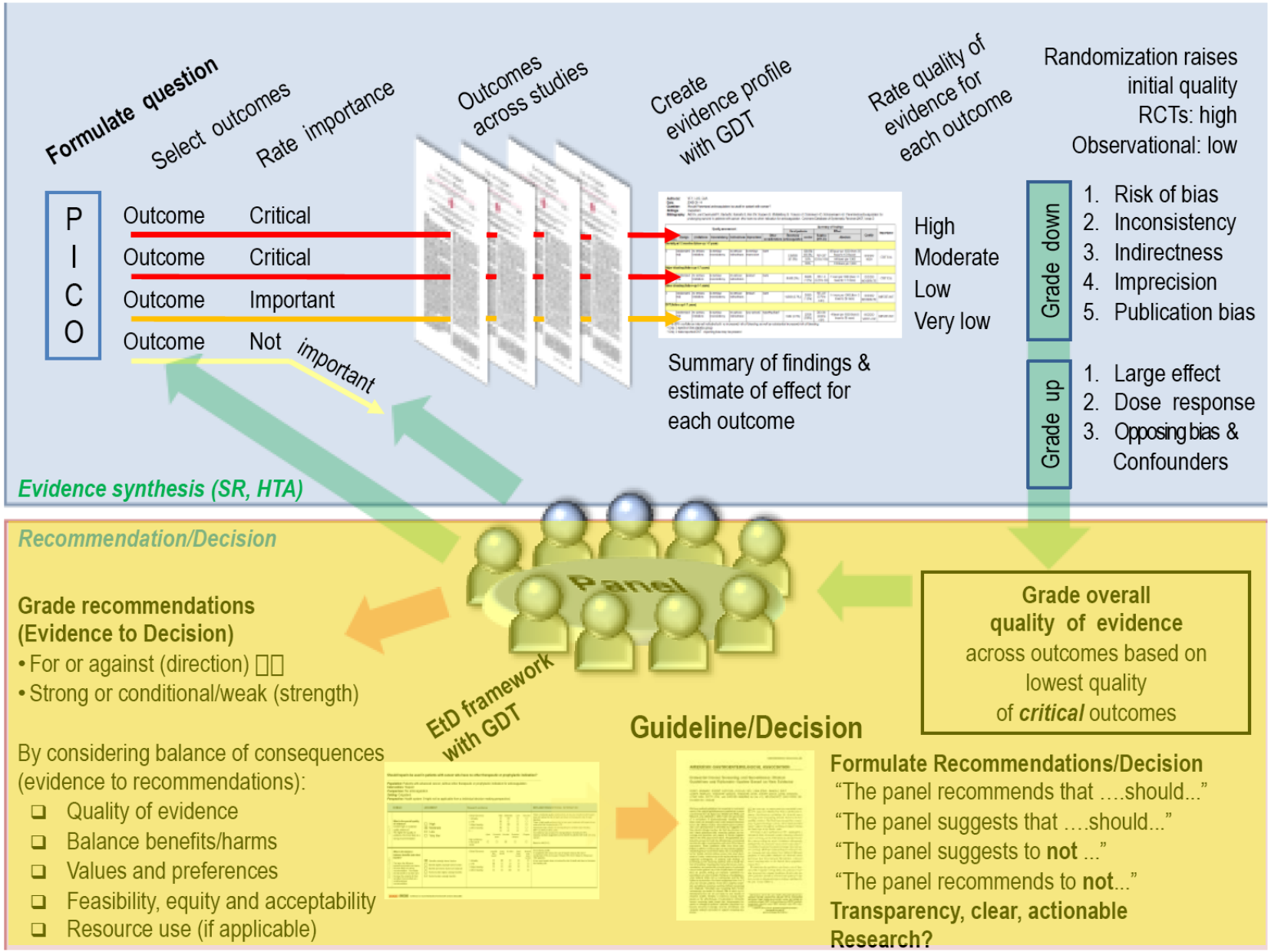
Pablo Alonso-Coello,[1,2] Andrew D Oxman,[3] Jenny Moberg,[3] Romina Brignardello-Petersen,[2,4] Elie A Akl,[2,5] Marina Davoli,[6] Shaun Treweek,[7] Reem A Mustafa,[2,8] Per O Vandvik,[3] Joerg Meerpohl,[9] Gordon H Guyatt,[2,10] Holger J Schünemann,[2,10] the GRADE Working Group

BMJ 2016;353:i2089

SUMMARY POINTS

- Explicit and transparent systems for decision making can help to ensure that all important criteria are considered and that decisions are informed by the best available research evidence
- The purpose of Evidence to Decision (EtD) frameworks is to help people use evidence in a structured and transparent way to inform decisions in the context of clinical recommendations, coverage decisions, and health system or public health recommendations and decisions
- EtD frameworks inform users about the judgments that were made and the evidence supporting those judgments by making the basis for decisions transparent to target audiences

The GRADE process in developing guidelines



GRADEpro GDT

Should Patisiran vs. [placebo] be used in Hereditary Transthyretin Amyloidosis (mNIS+7 analysis)?

Bottom panel | Explanations

1 **Problem**
Is the problem a priority?

2 **Desirable Effects**
How substantial are the desirable anticipated effects?

3 **Undesirable Effects**
How substantial are the undesirable anticipated effects?

4 **Certainty of evidence**
What is the overall certainty of the evidence of effects?

5 **Values**
Is there important uncertainty about or variability in how much people value the main outcomes?

6 **Balance of effects**
Does the balance between desirable and undesirable effects favor the intervention or the comparison?

7 **Resources required**
How large are the resource requirements (costs)?

8 **Certainty of evidence of required resources**
What is the certainty of the evidence of resource requirements (costs)?

9 **Cost effectiveness**
Does the cost-effectiveness of the intervention favor the intervention or the comparison?

10 **Equity**
What would be the impact on health equity?

11 **Acceptability**
Is the intervention acceptable to key stakeholders?

12 **Feasibility**
Is the intervention feasible to implement?

GRADEpro GDT

Should Patisiran vs. [placebo] be used in Hereditary Transthyretin Amyloidosis (mNIS+7 analysis)?

Bottom panel | Explanations

CRITERIA	SUMMARY OF JUDGEMENTS							IMPORTANCE FOR DECISION
PROBLEM	No	Probably no	Probably yes	**Yes**		Varies	Don't know	**HIGH**
DESIRABLE EFFECTS	Trivial	Small	Moderate	**Large**		Varies	Don't know	**HIGH**
UNDESIRABLE EFFECTS	Large	Moderate	**Small**	Trivial		Varies	Don't know	**HIGH**
CERTAINTY OF EVIDENCE	Very low	Low	Moderate	**High**		No included studies		**MODERATE**
VALUES	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	**No important uncertainty or variability**				**HIGH**
BALANCE OF EFFECTS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	**Favors the intervention**	Varies	Don't know	**HIGH**
RESOURCES REQUIRED	**Large costs**	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	Don't know	**LOW**
CERTAINTY OF EVIDENCE OF REQUIRED RESOURCES	Very low	Low	Moderate	High		No included studies		**LOW**
COST EFFECTIVENESS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	Varies	No included studies	**LOW**
EQUITY	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know	**MODERATE**
ACCEPTABILITY	No	Probably no	**Probably yes**	Yes		Varies	Don't know	**LOW**
FEASIBILITY	No	Probably no	**Probably yes**	Yes		Varies	Don't know	**MODERATE**

Valutazione della qualità globale delle prove

Dopo la valutazione della qualità per i singoli outcome importanti effettuata dall'ERT (Sezione 4.5) si deve formulare il giudizio complessivo di qualità. Il metodo GRADE suggerisce di procedere considerando soltanto gli outcome critici per la formulazione della raccomandazione relativa al quesito clinico.

Valutazione della qualità globale delle prove

Dopo la valutazione della qualità per i singoli outcome importanti effettuata dall'ERT (Sezione 4.5) si deve formulare il giudizio complessivo di qualità. Il metodo GRADE suggerisce di procedere considerando soltanto gli outcome critici per la formulazione della raccomandazione relativa al quesito clinico. Se la qualità è diversa fra i singoli outcome critici, il metodo indica la seguente linea di comportamento:

- se i risultati vanno in direzioni opposte (es. il trattamento oggetto della raccomandazione è migliore in termini di efficacia ma peggiore per quanto riguarda gli effetti indesiderati), la qualità globale viene attribuita basandosi sulla valutazione peggiore ossia assumendo come più rappresentativo l'outcome che ha ottenuto la più bassa valutazione di qualità;
- se i risultati vanno nella stessa direzione per tutti gli outcome (beneficio o danno), viene assunta come qualità globale delle prove la qualità di un singolo outcome critico che da solo basterebbe per formulare compiutamente la raccomandazione;

Valutazione della qualità globale delle prove

Dopo la valutazione della qualità per i singoli outcome importanti effettuata dall'ERT (Sezione 4.5) si deve formulare il giudizio complessivo di qualità. Il metodo GRADE suggerisce di procedere considerando soltanto gli outcome critici per la formulazione della raccomandazione relativa al quesito clinico. Se la qualità è diversa fra i singoli outcome critici, il metodo indica la seguente linea di comportamento:

- se i risultati vanno in direzioni opposte (es. il trattamento oggetto della raccomandazione è migliore in termini di efficacia ma peggiore per quanto riguarda gli effetti indesiderati), la qualità globale viene attribuita basandosi sulla valutazione peggiore ossia assumendo come più rappresentativo l'outcome che ha ottenuto la più bassa valutazione di qualità;
- se i risultati vanno nella stessa direzione per tutti gli outcome (beneficio o danno), viene assunta come qualità globale delle prove la qualità di un singolo outcome critico che da solo basterebbe per formulare compiutamente la raccomandazione;

Defining Strength of recommendation

- The extent to which a guideline panel is confident that desirable effects of an intervention outweigh undesirable effects, or vice versa, across the range of patients for whom the recommendation is intended.

Strength of recommendation

- A recommendation can have one of 2 strength:
  - **Strong** : panel is confident that the desirable effects of adherence to the recommendation outweigh the undesirable effects (or vice versa).
  - **Conditional** : panel concludes that the desirable effects of adherence to the recommendation probably outweigh the undesirable effects (or vice versa), but is not confident.

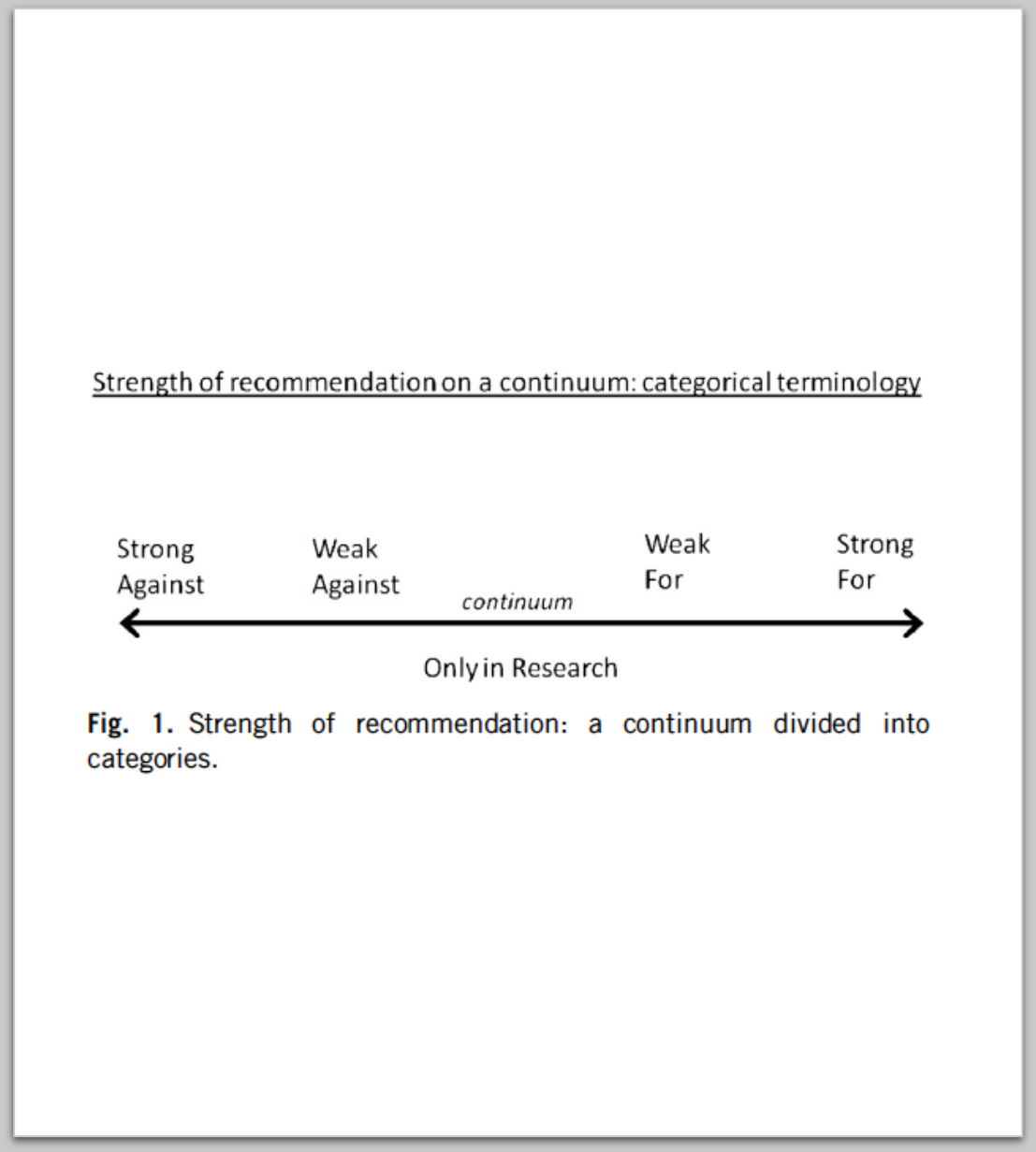


Fig. 1. Strength of recommendation: a continuum divided into categories.

Implications of strong and conditional recommendations

	Strong recommendation	Conditional recommendation
Patients	Most people in your situation would want the recommended course of action and only a small proportion would not	The majority of people in your situation would want the recommended course of action, but many would not
Clinicians		
Policy makers		

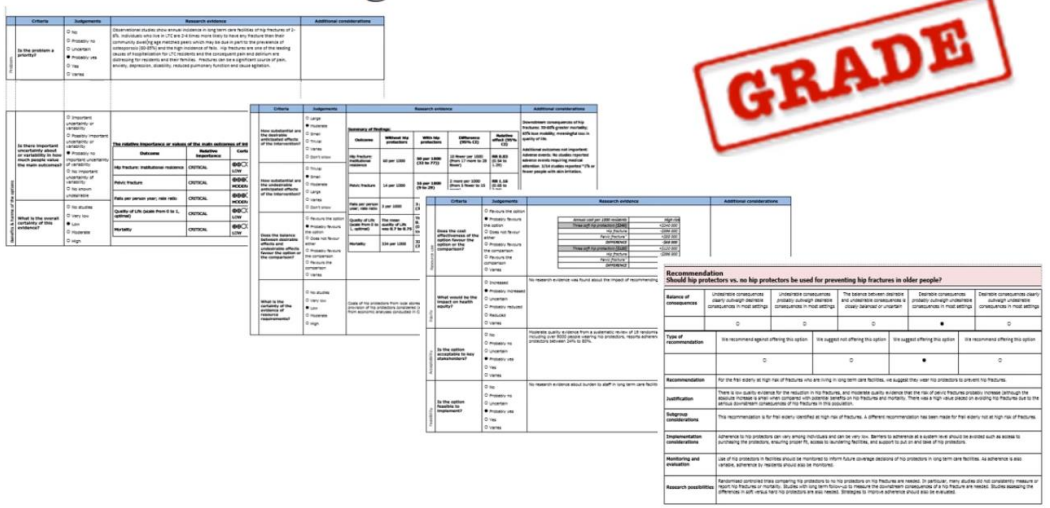
Implications of strong and conditional recommendations

	Strong recommendation	Conditional recommendation
Patients	Most people in your situation would want the recommended course of action and only a small proportion would not	The majority of people in your situation would want the recommended course of action, but many would not
Clinicians	Most patients should receive the recommended course of action	Be prepared to help patients to make a decision that is consistent with their own values
Policy makers		

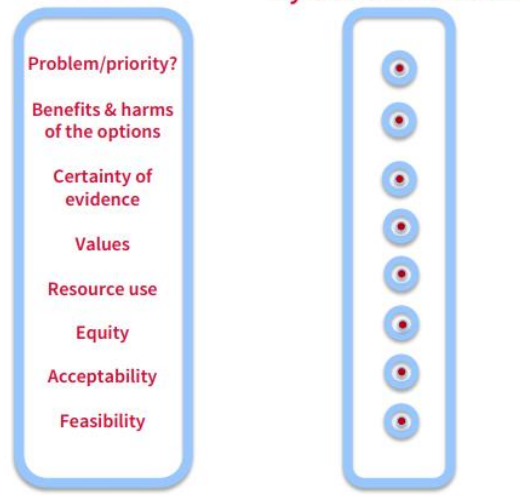
Implications of strong and conditional recommendations

	Strong recommendation	Conditional recommendation
Patients	Most people in your situation would want the recommended course of action and only a small proportion would not	The majority of people in your situation would want the recommended course of action, but many would not
Clinicians	Most patients should receive the recommended course of action	Be prepared to help patients to make a decision that is consistent with their own values
Policy makers	The recommendation can be adapted as a policy in most situations	There is a need for substantial debate and involvement of stakeholders

Evidence to Decision Table:
To facilitate making recommendations



GRADE



SACRO CUORE DON CALABRIA I R C C S

ECI CANCER CENTRE

SCUOLA DI METODOLOGIA DELLA RICERCA CLINICA

2024 - 10ª EDIZIONE

MODULI SPECIALISTICI - S2

VENERDÌ 12 - SABATO 13 APRILE 2024

NEGRAR DI VALPOLICELLA (VR)

Centro Formazione IRCCS "Sacro Cuore - Don Calabria"

GRADE

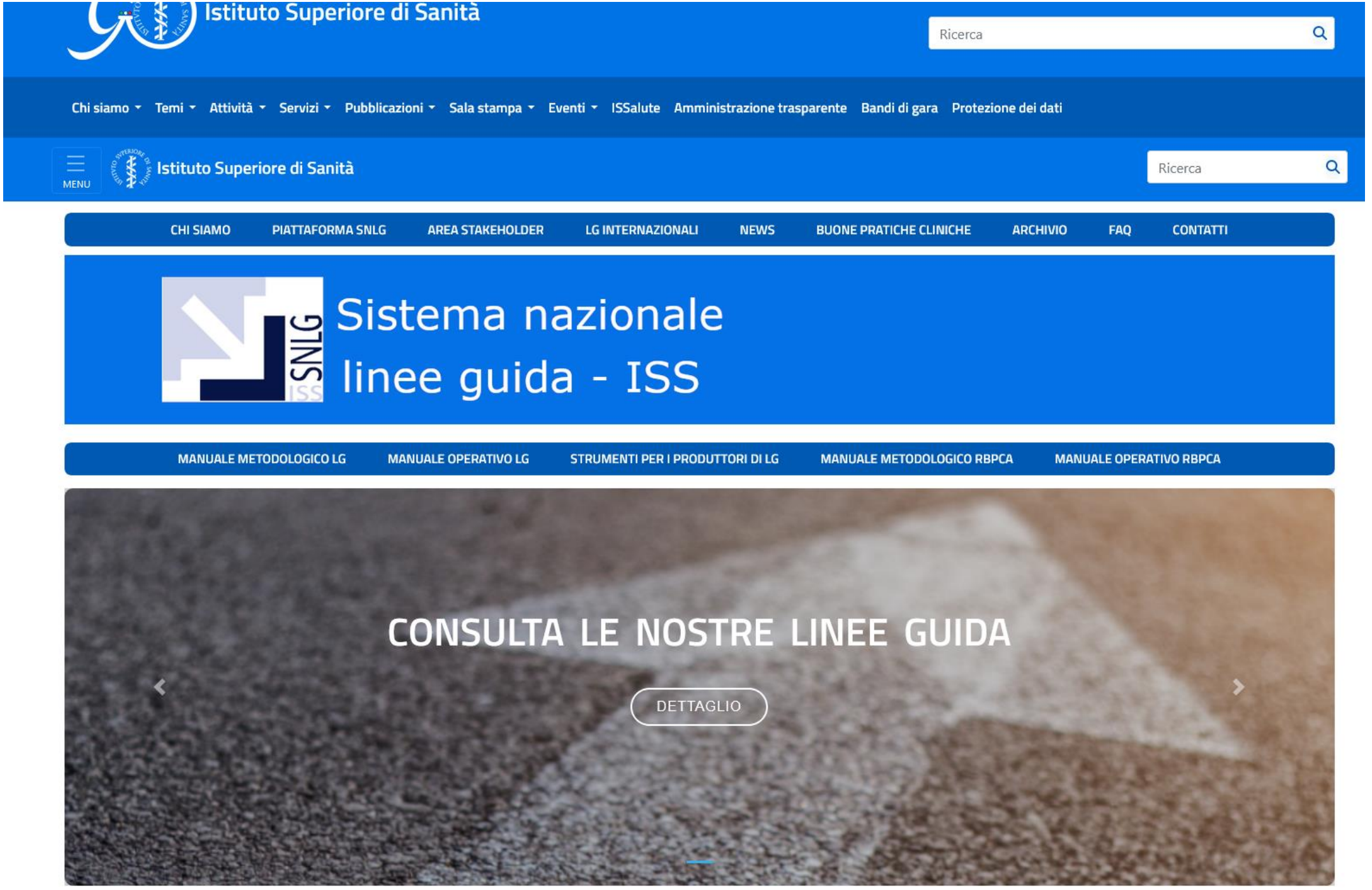
Metodiche di valutazione di una Linea Guida e normativa vigente

(Michela Cinquini)

Accreditamento delle linee guida

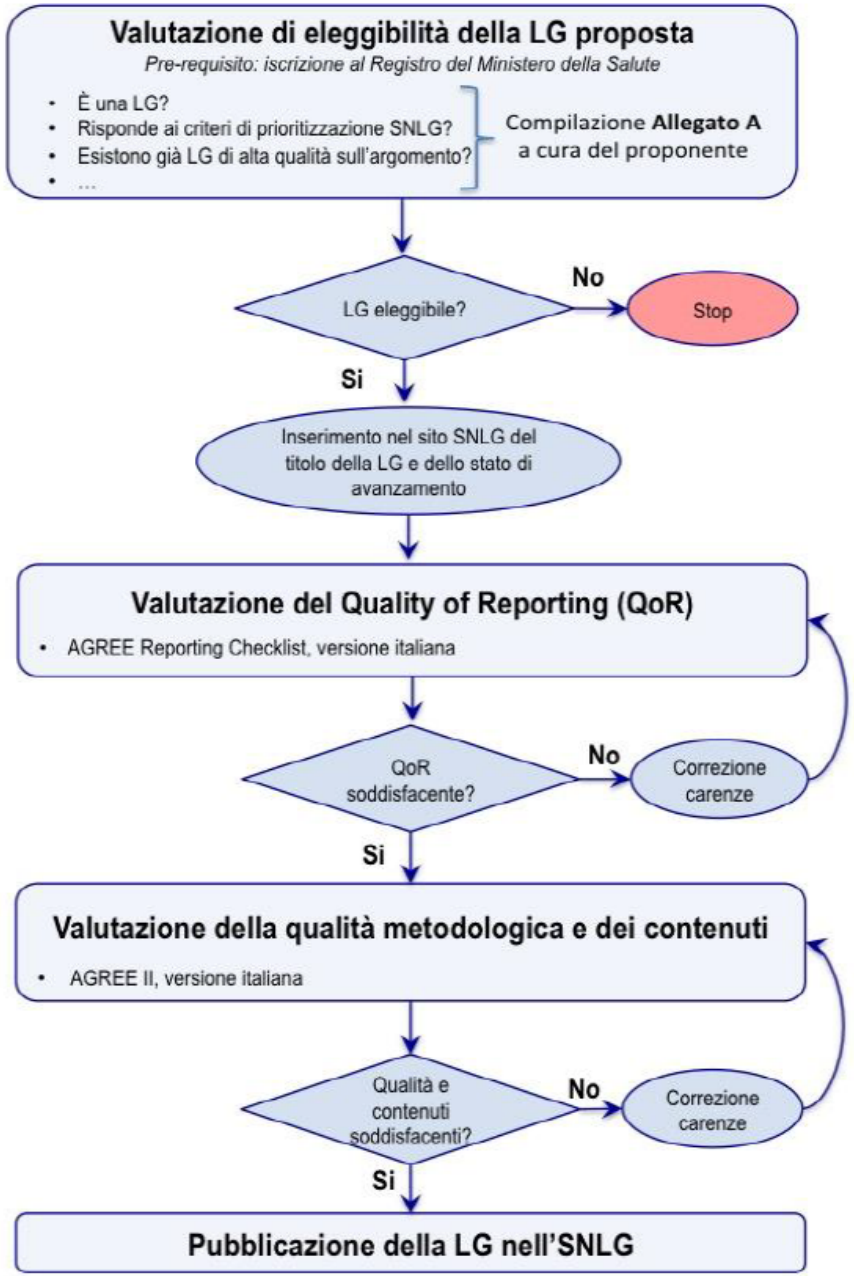
l'Italia si munisce quindi di un sistema di accreditamento, monitoraggio ed aggiornamento delle linee guida (2017)

- Attività di **verifica su linee guida** che, ai sensi del comma l, sono «**elaborate da**:
  - **enti** ed **istituzioni** pubblici e privati nonché dalle **società scientifiche** e dalle **associazioni tecnico-scientifiche** delle professioni sanitarie iscritte in **apposito elenco** istituito e regolamentato con decreto del Ministro della salute,
    - da aggiornare con cadenza biennale


Istituto Superiore di Sanità
Ricerca
Chi siamo
Temi
Attività
Servizi
Pubblicazioni
Sala stampa
Eventi
ISSalute
Amministrazione trasparente
Bandi di gara
Protezione dei dati
MENU
Istituto Superiore di Sanità
Ricerca
CHI SIAMO
PIATTAFORMA SNLG
AREA STAKEHOLDER
LG INTERNAZIONALI
NEWS
BUONE PRATICHE CLINICHE
ARCHIVIO
FAQ
CONTATTI
SNLG
ISS
Sistema nazionale
linee guida - ISS
MANUALE METODOLOGICO LG
MANUALE OPERATIVO LG
STRUMENTI PER I PRODUTTORI DI LG
MANUALE METODOLOGICO RBPCA
MANUALE OPERATIVO RBPCA
CONSULTA LE NOSTRE LINEE GUIDA
DETTAGLIO

Processo di valutazione delle LG proposte da soggetti ex art.5 L. n.24/17 per la pubblicazione nell'SNLG

Courtesy of Primiano Iannone



Le richieste di valutazione vanno inviate online attraverso la **piattaforma SNLG**

Workflow sviluppato dal CNEC con il supporto del centro collaboratore **GIMBE**

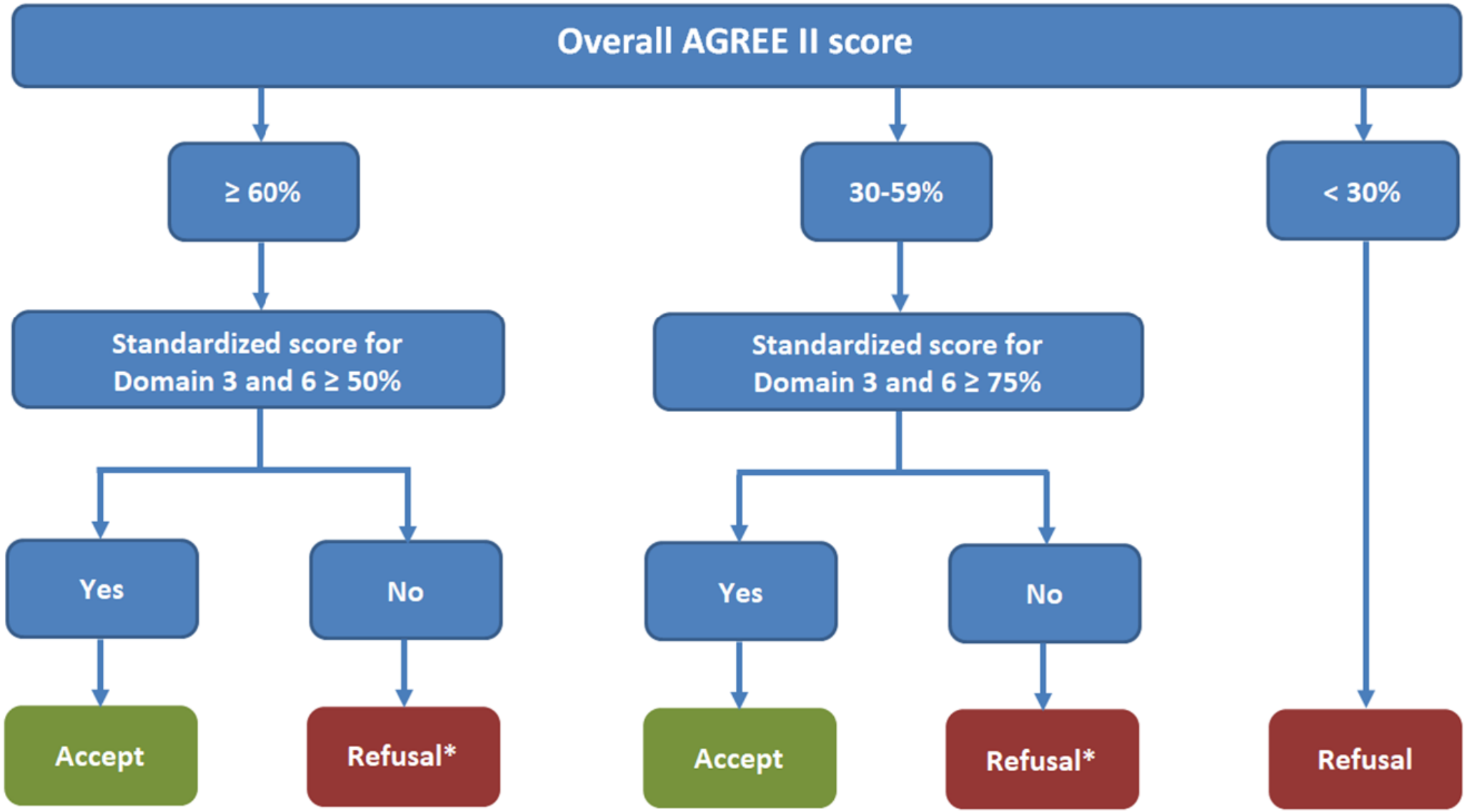
AGREE II e AGREE Reporting Checklist

APPRAISAL OF GUIDELINES FOR RESEARCH & EVALUATION

Dimensione 1	OBIETTIVI E AMBITI DI APPLICAZIONE (1-3)
Dimensione 2	COINVOLGIMENTO DEGLI STAKEHOLDER (4-6)
Dimensione 3	RIGORE METODOLOGICO (7-14)
Dimensione 4	CHIAREZZA ESPOSITIVA (15-17)
Dimensione 5	APPLICABILITÀ (18-21)
Dimensione 6	INDIPENDENZA EDITORIALE (22-23)

benefici potenziali LG proporzionali alla loro qualità → **strumento di valutazione standardizzato**

Courtesy of Primiano Iannone



* Accept with reserve if no other CPG on specific topic is available

Courtesy of Primiano Iannone

Pubblicato 21/03/2020 - Modificato 17/09/2024

Le Linee Guida (LG) di pratica clinica sono uno strumento di supporto decisionale finalizzato a consentire che, fra opzioni alternative, sia adottata quella che offre un migliore bilancio fra benefici ed effetti indesiderati, tenendo conto della esplicita e sistematica valutazione delle prove disponibili, commisurandola alle circostanze peculiari del caso concreto e condividendola-laddove possibile- con il paziente o i *caregivers*. Conoscere e adottare giudiziosamente le raccomandazioni cliniche contenute nelle migliori LG rappresenta un obiettivo etico non solo del singolo professionista ma dei sistemi sanitari nel loro complesso, per massimizzare la probabilità di esiti di salute favorevoli attraverso pratiche cliniche consistenti con le migliori evidenze disponibili, e promuovere la sicurezza, l'equità, l'efficienza e l'appropriatezza- in una parola, la qualità- delle cure. Per queste ragioni numerosi sistemi sanitari hanno riconosciuto da tempo la necessità di raccolte nazionali di LG di riferimento.

In questa sezione sono riportate le **LG del *S*istema Nazionale Linee Guida (SNLG)** in *fase di sviluppo*, in *fase di valutazione* e *concluse*, elaborate a cura dei soggetti di cui all'art. 5 comma 1 della legge n° 24/2017: enti e istituzioni pubbliche e private e società scientifiche e associazioni tecnico-scientifiche delle professioni sanitarie iscritte in apposito elenco istituito e regolamentato con DM 2 agosto 2017.

L'SNLG, istituito con DM 27 febbraio 2018, costituisce **l'unico punto di accesso** per professionisti sanitari, manager, decisori, utenti e caregiver a un corpus di LG per la pratica clinica e la salute pubblica ordinato secondo criteri di rilevanza, non ridondanza e coerenza interna.

Qui di seguito sono riportate le tre tipologie di documenti consultabili:

- **Linee guida concluse**: LG che ha superato la valutazione con AGREE II e se ne raccomanda l'utilizzo.
- **Linee guida in fase di valutazione** con lo strumento AGREE II.
- **Linee guida in progress**: in fase di sviluppo da parte dei soggetti di cui all'art. 5 comma 1 della legge n° 24/2017.

Cerca LG concluse | Cerca LG in progress | Cerca LG in valutazione

Linee Guida SNLG

Linee guida SNLG

Manuale metodologico per la produzione di linee guida di pratica clinica

v. 2.0 dicembre 2024

Versione 2024

Sviluppato dal CNEC in collaborazione con il GRADE Working Group

Allegati

Allegato 1 - AGREE quality of reporting checklist

Allegato 2 - AGREE II versione italiana

Allegato 3 - Dichiarazione Impegno Membri Panel LG

Allegato 4 - Dichiarazione Impegno altri soggetti LG

Allegato 5 - Modulo conflitto di interessi compilabile - versione dicembre 2024

Allegato 6 - Format per il reporting di linee guida complete

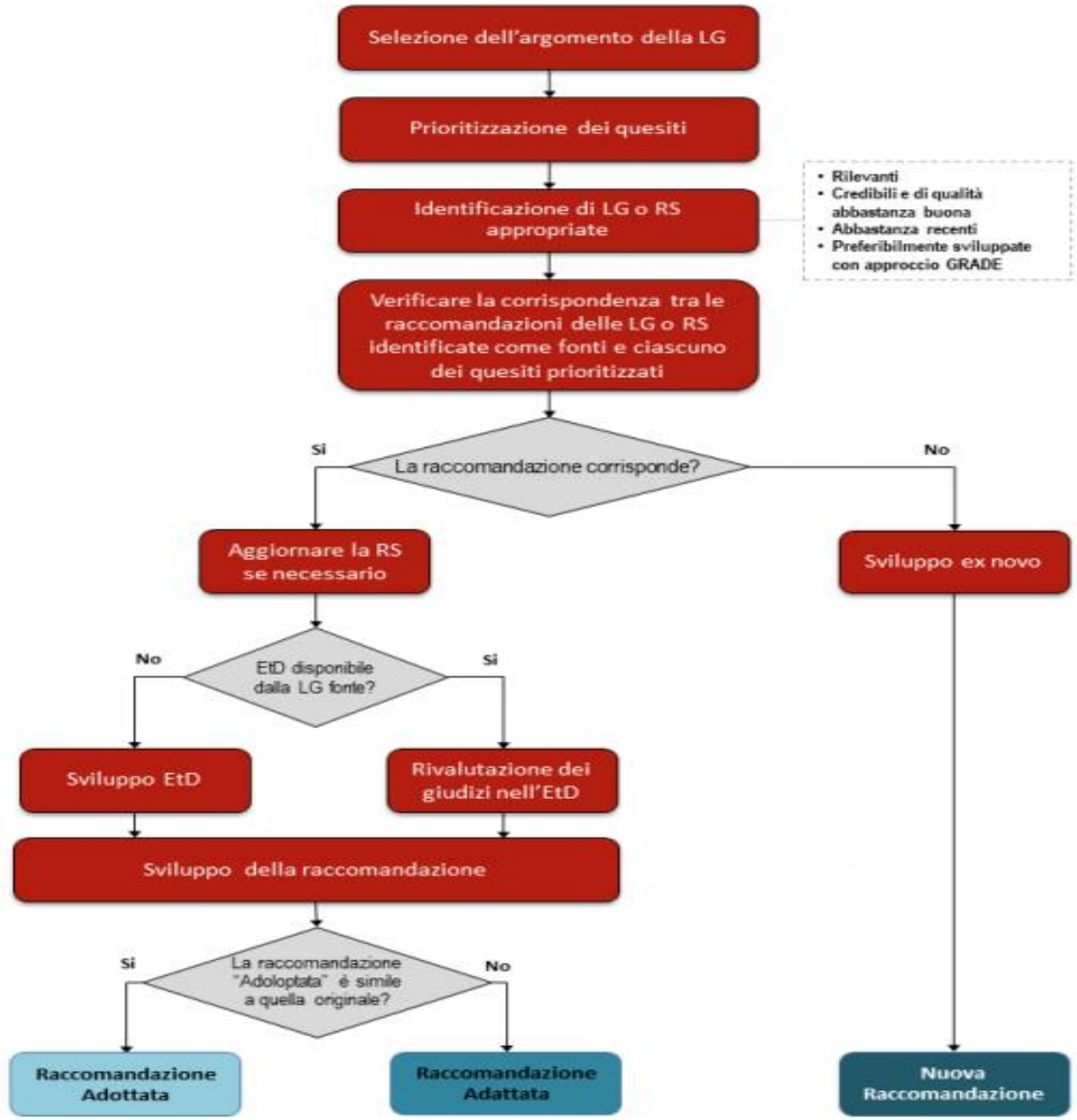


Figura 3 – Tradotta da: *Journal of Clinical Epidemiology* 2017 81, 101-110DOI: (10.1016/j.jclinepi.2016.09.009)

Procedure di invio e valutazione di Linee Guida per la pubblicazione nel SNLG

Manuale operativo

CENTRO NAZIONALE
ECCELLENZA CLINICA,
QUALITÀ E SICUREZZA DELLE CURE

Versione 3.1 – marzo 2023

Indicazioni metodologiche per la stesura di raccomandazioni per le buone pratiche clinico assistenziali

v. 1.1 dicembre 2024

Procedure di invio e valutazione di raccomandazioni per le buone pratiche clinico-assistenziali

v. 1.3 marzo 2025

SACRO CUORE DON CALABRIA I R C C S

ECI CANCER CENTRE

SCUOLA DI METODOLOGIA DELLA RICERCA CLINICA

2024 - 10ª EDIZIONE

MODULI SPECIALISTICI - S2

VENERDÌ 12 - SABATO 13 APRILE 2024

NEGRAR DI VALPOLICELLA (VR)

Centro Formazione IRCCS "Sacro Cuore - Don Calabria"

Lettura: come valutare il reale vantaggio clinico di un nuovo approccio terapeutico

(Massimo Di Maio)

Dallo studio clinico alla pratica clinica

Analisi critica delle evidenze

Come valutare il reale vantaggio clinico di un nuovo approccio terapeutico

Dati post-registrativi

Panel 1: Problems that might limit interpretation of randomised controlled trials

Some randomised controlled trials might:

- Ask questions of commercial rather than clinical interest
- Be based on inadequate preclinical and early clinical studies
- Use surrogate endpoints that do not reflect patient benefit (ie, duration or quality of survival)
- Fail to assess or inadequately assess health-related quality of life or patient-reported outcomes even though the goals of treatment are palliative
- Show statistically significant but clinically irrelevant results
- Be analysed and reported prematurely
- Underestimate the toxicity of new treatments
- Be subject to biased reporting, both in the primary publication and by the media
- Select patients who do not represent those seen in everyday practice

Ian Tannock et al,
Relevance of randomised controlled trials in oncology,
Lancet Oncology 2016; 16: e560-567

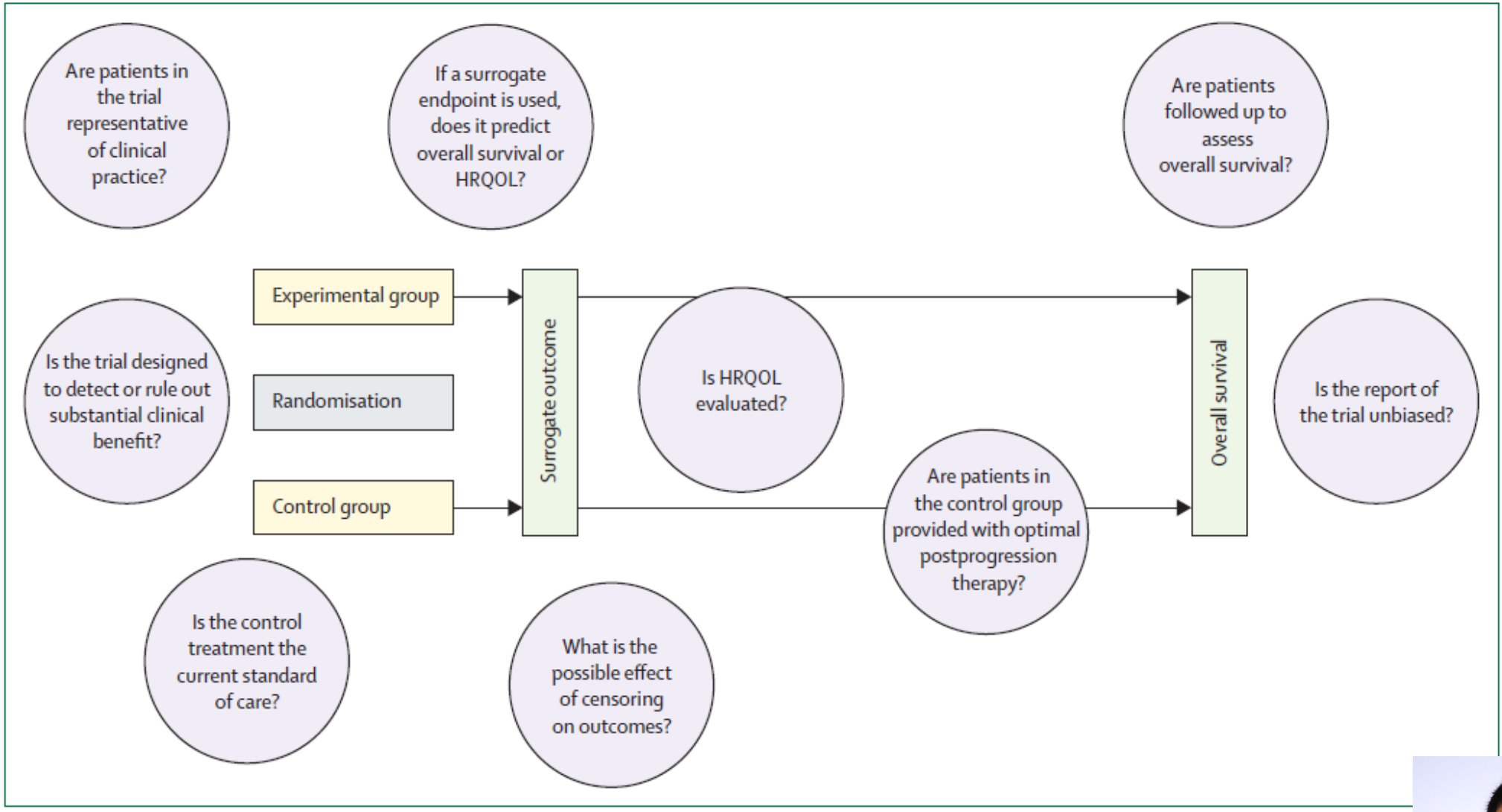


Figure: **Questions relating to the design and reporting of randomised controlled trials**
Figure adapted from Gyawali and colleagues.[5] HRQOL=health-related quality of life.

Gyawali B et al. Common Sense Oncology principles for the design, analysis, and reporting of phase 3 randomised clinical trials. *Lancet Oncol* 2025; 26: e80–89

Studi clinici: validità dei risultati

Internal validity

Is the research conducted correctly?

L'adeguatezza del braccio di controllo

European Journal of Cancer 189 (2023) 112920

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original research

Analysis of the adequacy of control arms in oncology randomised clinical trials published between 2017 and 2021: a meta-research study

Alessandro Rossi [a], Giacomo Aimar [b,c], Marco Audisio [d], Maristella Bungaro [e], Andrea Caglio [f], Raimondo Di Liello [g], Teresa Gamba [f], Piera Gargiulo [h], Eleonora Ghisoni [i], Pasquale Lombardi [j], Laura Marandino [k], Annapaola Mariniello [l,m], Chiara Paratore [d], Maria Lucia Reale [n], Federica Trastu [f], Valentina Tuninetti [f], Fabio Turco [m,o], Alessandra Fabi [a], Francesco Perrone [h], Massimo Di Maio [p,*,1,2]

Rossi A, et al. Eur J Cancer. 2023 Aug;189:112920.

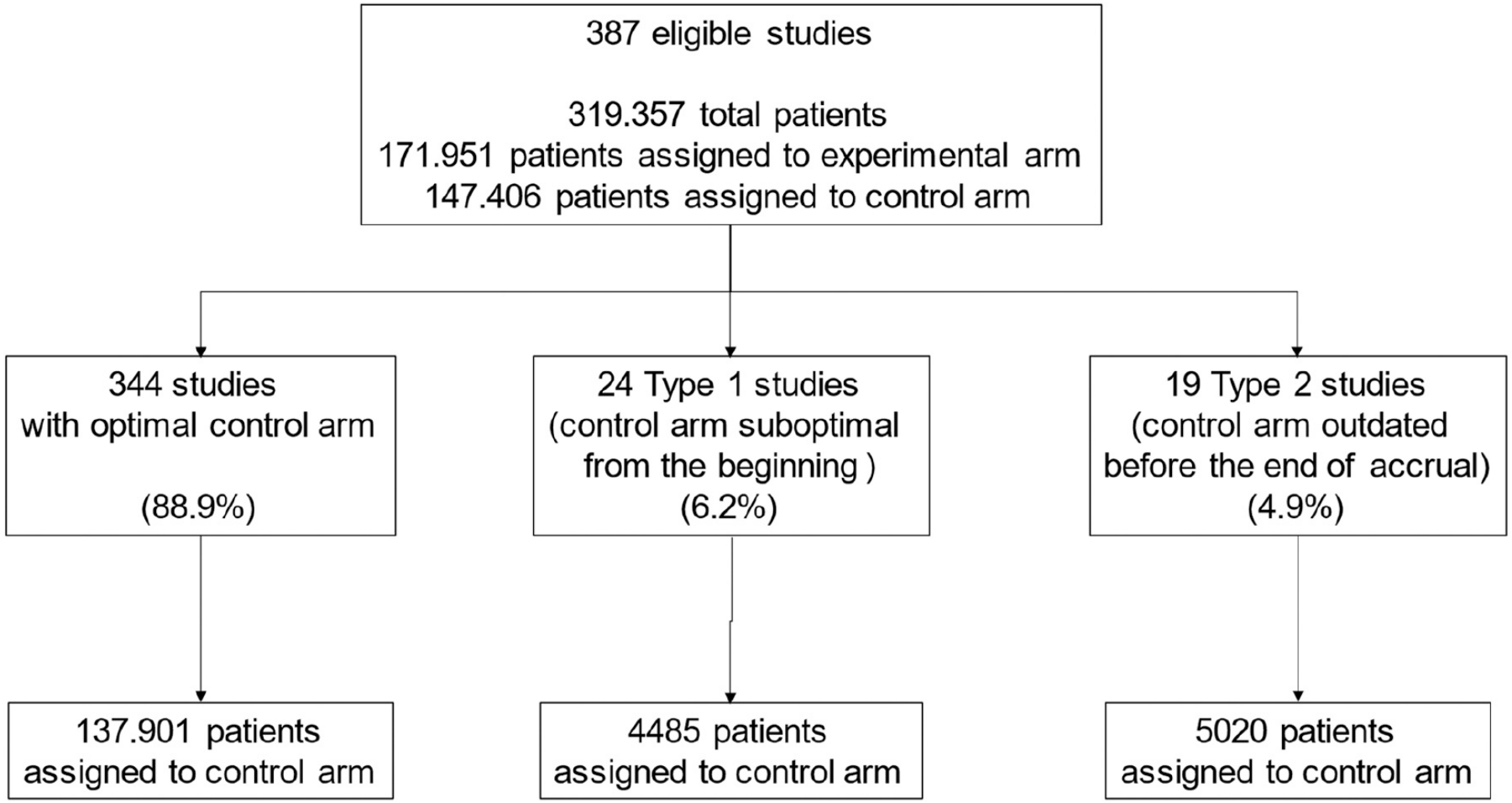
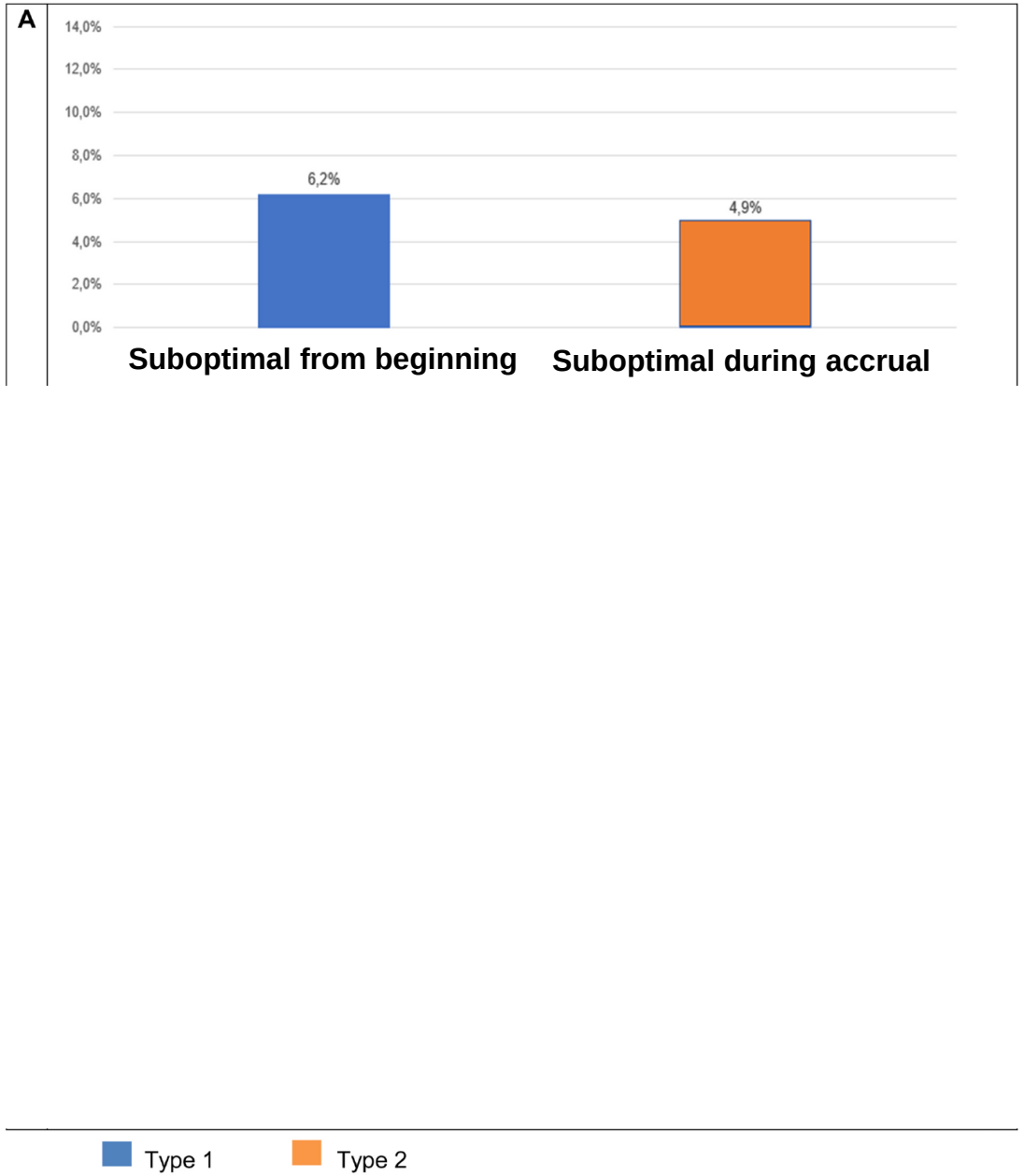
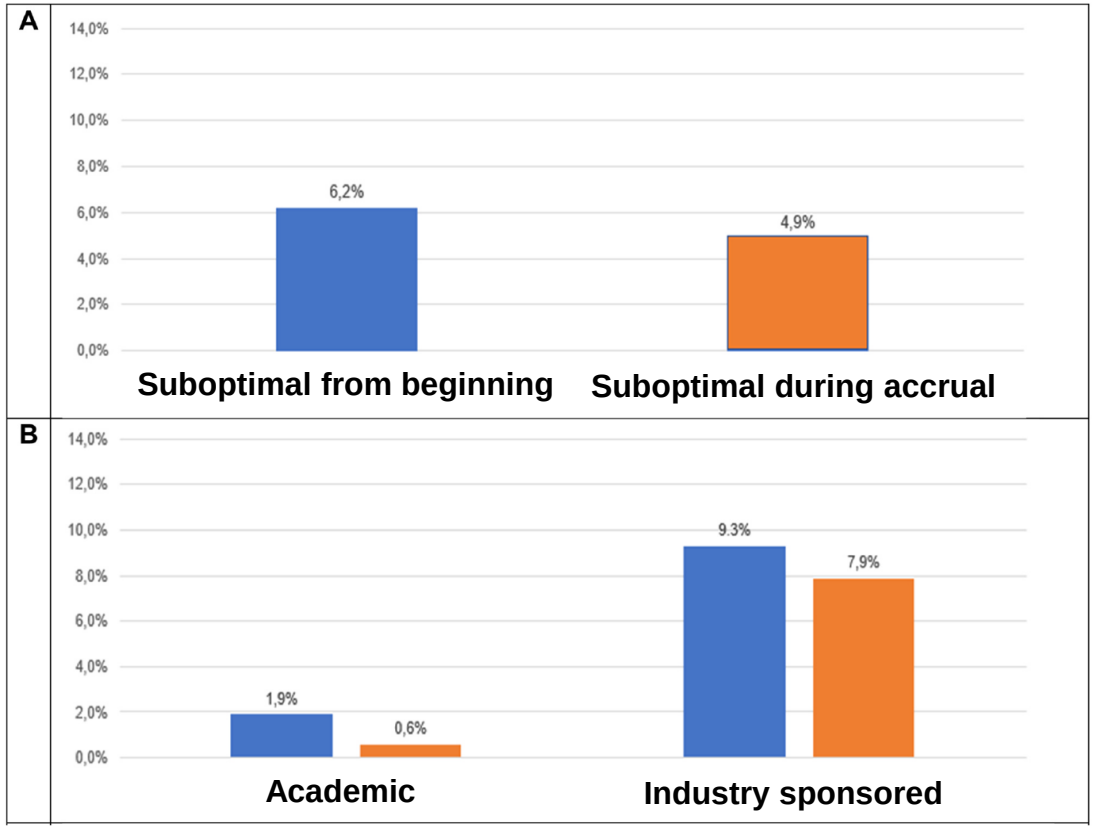


Fig. 2. Number of studies and number of patients assigned to control arm for each subgroup (studies with optimal control arm, studies with control arm suboptimal from the beginning, studies with control arm outdated before the end of the accrual).

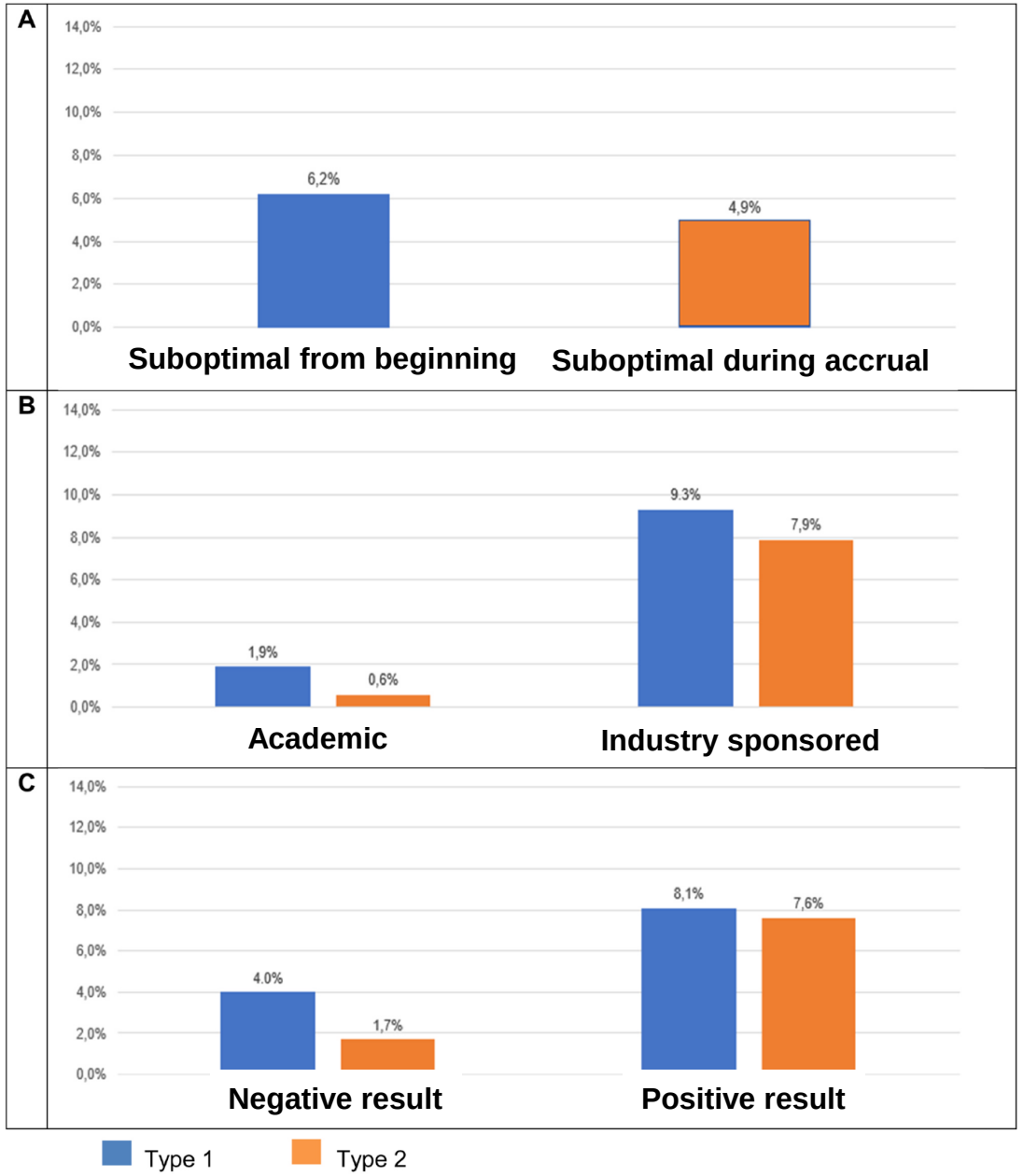
Rossi A, et al. Eur J Cancer. 2023 Aug;189:112920.



Rossi A, et al. Eur J Cancer. 2023 Aug;189:112920.



Rossi A, et al. Eur J Cancer. 2023 Aug;189:112920.



Rossi A, et al. Eur J Cancer. 2023 Aug;189:112920.

Panel 1: Problems that might limit interpretation of randomised controlled trials

Some randomised controlled trials might:

- Ask questions of commercial rather than clinical interest
- Be based on inadequate preclinical and early clinical studies
- Use surrogate endpoints that do not reflect patient benefit (ie, duration or quality of survival)
- Fail to assess or inadequately assess health-related quality of life or patient-reported outcomes even though the goals of treatment are palliative
- Show statistically significant but clinically irrelevant results
- Be analysed and reported prematurely
- Underestimate the toxicity of new treatments
- Be subject to biased reporting, both in the primary publication and by the media
- Select patients who do not represent those seen in everyday practice

Ian Tannock et al,
Relevance of randomised controlled trials in oncology,
Lancet Oncology 2016; 16: e560-567

Debate: This House Believes That Overall Survival is the Only Endpoint for Drug Approval
Stockholm, 27 September 2011

Overall survival is NOT the only endpoint for drug approval

Massimo Di Maio, MD
Medical oncologist
Clinical Trials Unit
National Cancer Institute, Napoli, Italy
dimaiomax@libero.it

A primary rationale for using PFS as an endpoint in cancer trials is that it could be considered as a **clinical benefit endpoint in itself**, provided that:

- **treatment effect is sufficiently large**
- **not only instrumental but clinical benefit**

Panel 1: Problems that might limit interpretation of randomised controlled trials

Some randomised controlled trials might:

- Ask questions of commercial rather than clinical interest
- Be based on inadequate preclinical and early clinical studies
- Use surrogate endpoints that do not reflect patient benefit (ie, duration or quality of survival)
- Fail to assess or inadequately assess health-related quality of life or patient-reported outcomes even though the goals of treatment are palliative
- Show statistically significant but clinically irrelevant results
- Be analysed and reported prematurely
- Underestimate the toxicity of new treatments
- Be subject to biased reporting, both in the primary publication and by the media
- Select patients who do not represent those seen in everyday practice

Ian Tannock et al,
Relevance of randomised controlled trials in oncology,
Lancet Oncology 2016; 16: e560-567

Underrating and underreporting of QoL and PROs in oncology

2012-2016

- 446 phase III trials
- 47.1% QoL not included among endpoints
- 38.1% QoL results collected but not presented in primary publications

Marandino et al, Ann Oncol 2018

Underrating and underreporting of QoL and PROs in oncology

2012-2016

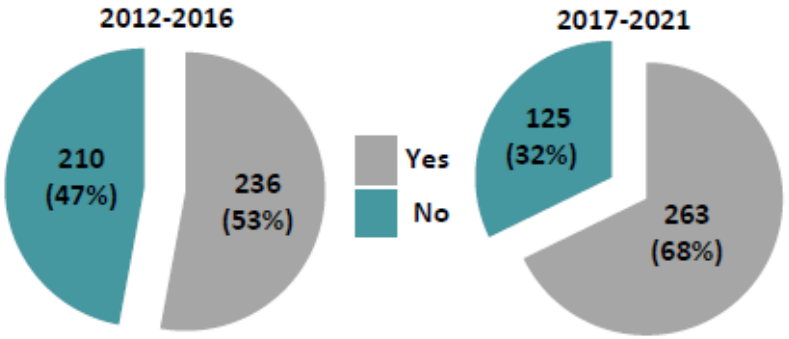
- 446 phase III trials
- 47.1% QoL not included among endpoints
- 38.1% QoL results collected but not presented in primary publications

Marandino et al, Ann Oncol 2018

2017-2021

388 phase III trials

Fig.1 Inclusion of QoL among study endpoints



Marandino et al, BMJ Oncology 2023;2:e000021

Underrating and underreporting of QoL and PROs in oncology

2012-2016

- 446 phase III trials
- 47.1% QoL not included among endpoints
- 38.1% QoL results collected but not presented in primary publications

Marandino et al, Ann Oncol 2018

2017-2021

388 phase III trials

Fig.1 Inclusion of QoL among study endpoints

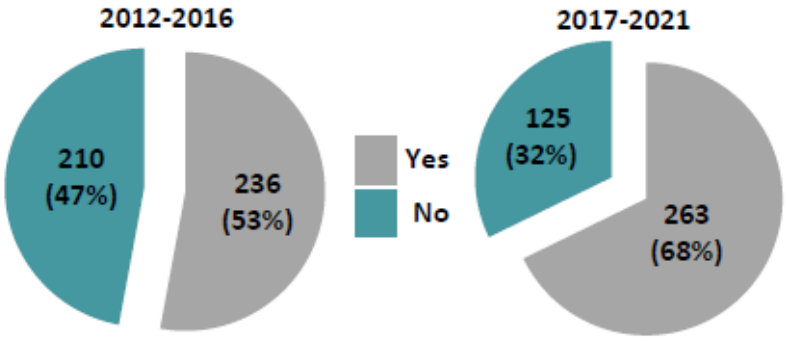
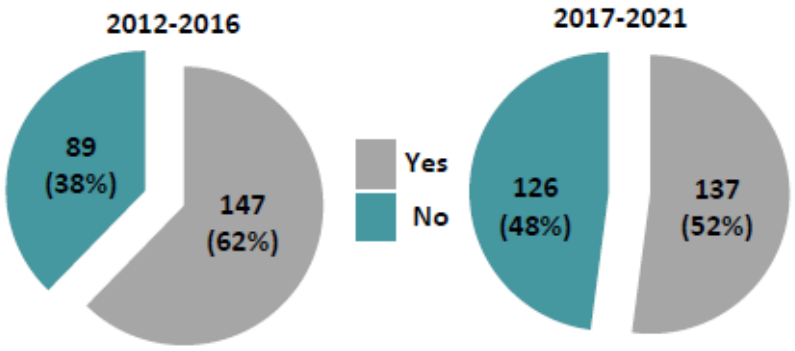


Fig.2 QoL results in primary study publication



Marandino et al, BMJ Oncology 2023;2:e000021

External validity

Are the results applicable to the real world?

Internal validity

Is the research conducted correctly?

Panel 1: Problems that might limit interpretation of randomised controlled trials

Some randomised controlled trials might:

- Ask questions of commercial rather than clinical interest
- Be based on inadequate preclinical and early clinical studies
- Use surrogate endpoints that do not reflect patient benefit (ie, duration or quality of survival)
- Fail to assess or inadequately assess health-related quality of life or patient-reported outcomes even though the goals of treatment are palliative
- Show statistically significant but clinically irrelevant results
- Be analysed and reported prematurely
- Underestimate the toxicity of new treatments
- Be subject to biased reporting, both in the primary publication and by the media
- Select patients who do not represent those seen in everyday practice

Ian Tannock et al,
Relevance of randomised controlled trials in oncology,
Lancet Oncology 2016; 16: e560-567

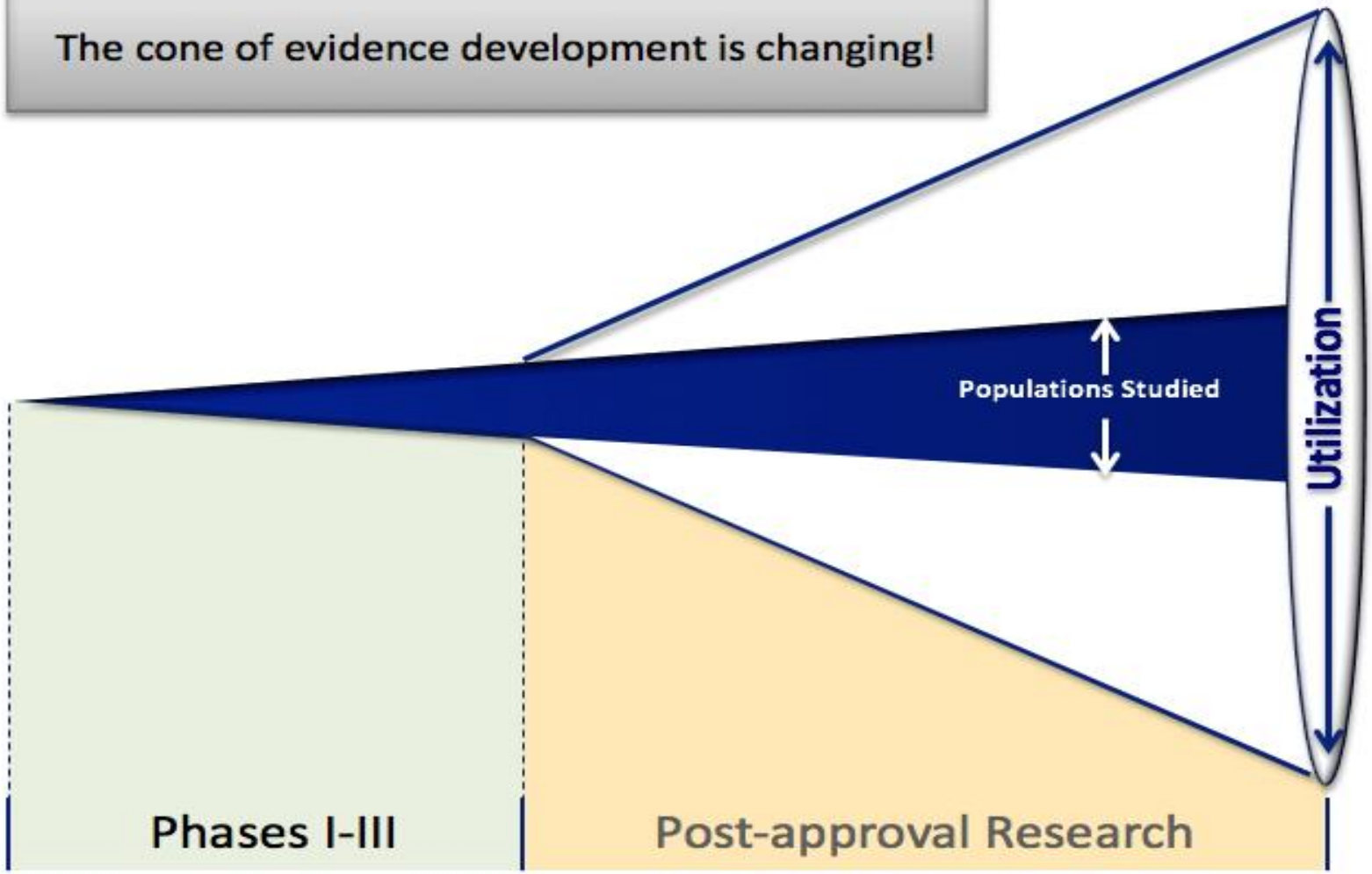
FDA analysis of Investigational New Drug Applications in 2015

- ~290 commercial IND submissions from 2015
  - 4% included pediatric patients
  - 60% required ECOG Performance status of 0-1
  - 77% excluded known, active or symptomatic CNS or brain metastases (47% allowed treated or stable brain metastases)
  - 84% excluded patients with known or active HIV (with only 2% allowing patients to enroll with adequate CD4 counts)
  - 74% excluded patients with history (or current) cardiovascular disease or risk (including angina pectoris, uncontrolled HTN, MI, CHF, arrhythmia)

Jin et al. JCO 2017

https://www.fda.gov/media/110332/download

The cone of evidence development is changing!



Variances in populations utilizing technology versus the populations studied

- Differing age groups (elderly, pediatrics)
- Race, ethnicity & gender variances
- Unstudied co-morbid conditions
- Differing concomitant drugs (including OTC)
- Lifestyle variances including smoking, dietary habits
- Differences in disease severity
- Varying levels of compliance

UNIVERSITÀ DI TORINO

Comparison of baseline patient characteristics in Italian oncology drug monitoring registries and clinical trials: a real-world cross-sectional study

Maria Lucia Iacovino,[a,j] Simone Celant,[b,j] Luca Tomassini,[b] Laura Arenare,[a] Andrea Caglio,[c] Andrea Canciello,[a] Flavio Salerno,[c] Pier Paolo Olimpieri,[b] Susanna Di Segni,[b] Antonella Sferrazza,[b] Maria Carmela Piccirillo,[a] Giordano Domenico Beretta,[d] Carmine Pinto,[e] Livio Blasi,[f] Saverio Cinieri,[g] Luigi Cavanna,[h] Massimo Di Maio,[i] Pierluigi Russo,[b,k] and Francesco Perrone[a,k,*]

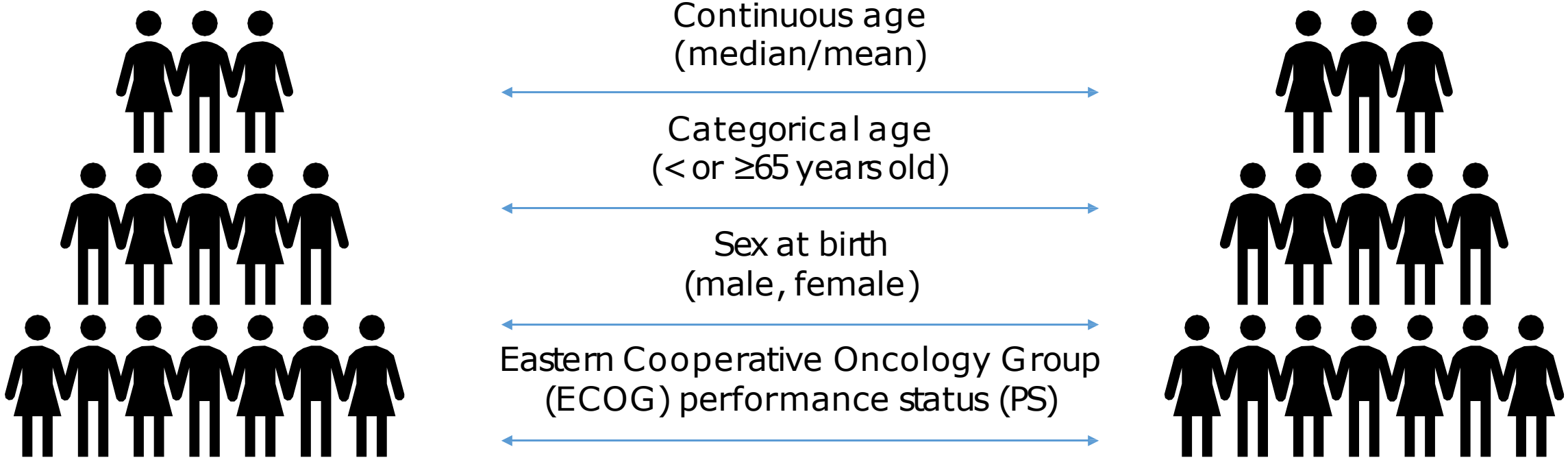
[a]National Cancer Institute, IRCCS Fondazione G.Pascale, Naples, Italy
[b]Italian Medicines Agency, Rome, Italy
[c]Department of Oncology, University, Ordine Mauriziano Hospital Umberto I, Turin, Italy
[d]Medical Oncology, Hospital, Pescara, Italy
[e]Medical Oncology, Comprehensive Cancer Centre, AUSL-IRCCS di Reggio Emilia, Italy
[f]Medical Oncology, Civic Hospital Cristina Benfratelli, Palermo, Italy
[g]Medical Oncology and Breast Unit, Perrino Hospital, Brindisi, Italy
[h]Medical Oncology and Hematology, Civil Hospital, Piacenza, Italy
[i]Department of Oncology, University of Turin, AOU Città della Salute e della Scienza di Torino, Turin, Italy

Iacovino ML,. Lancet Reg Health Eur. 2024 Apr 22;41:100912. doi: 10.1016/j.lanepe.2024.100912.

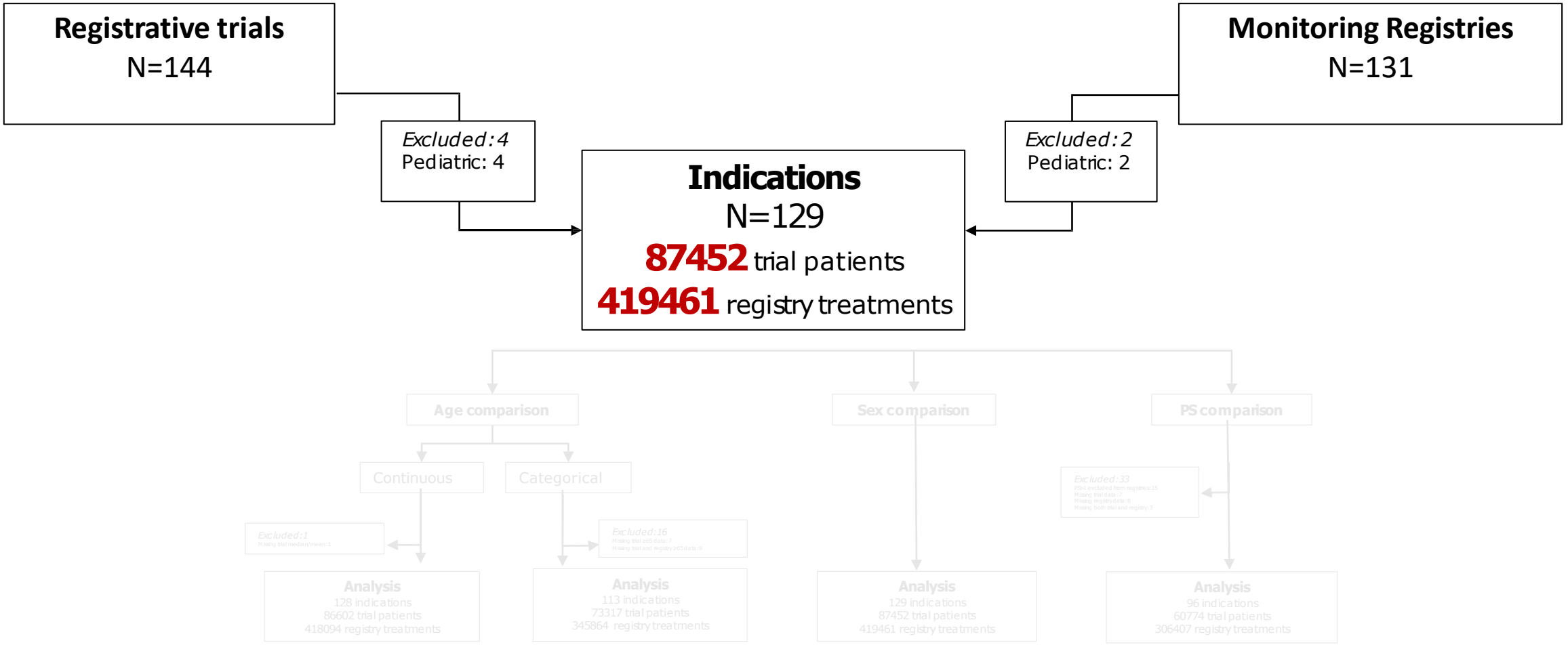
Registrative Clinical Trials from European Public Assessment Reports

AIFA cancer monitoring registries from 02/01/2013 to 22/04/2023

Baseline characteristics



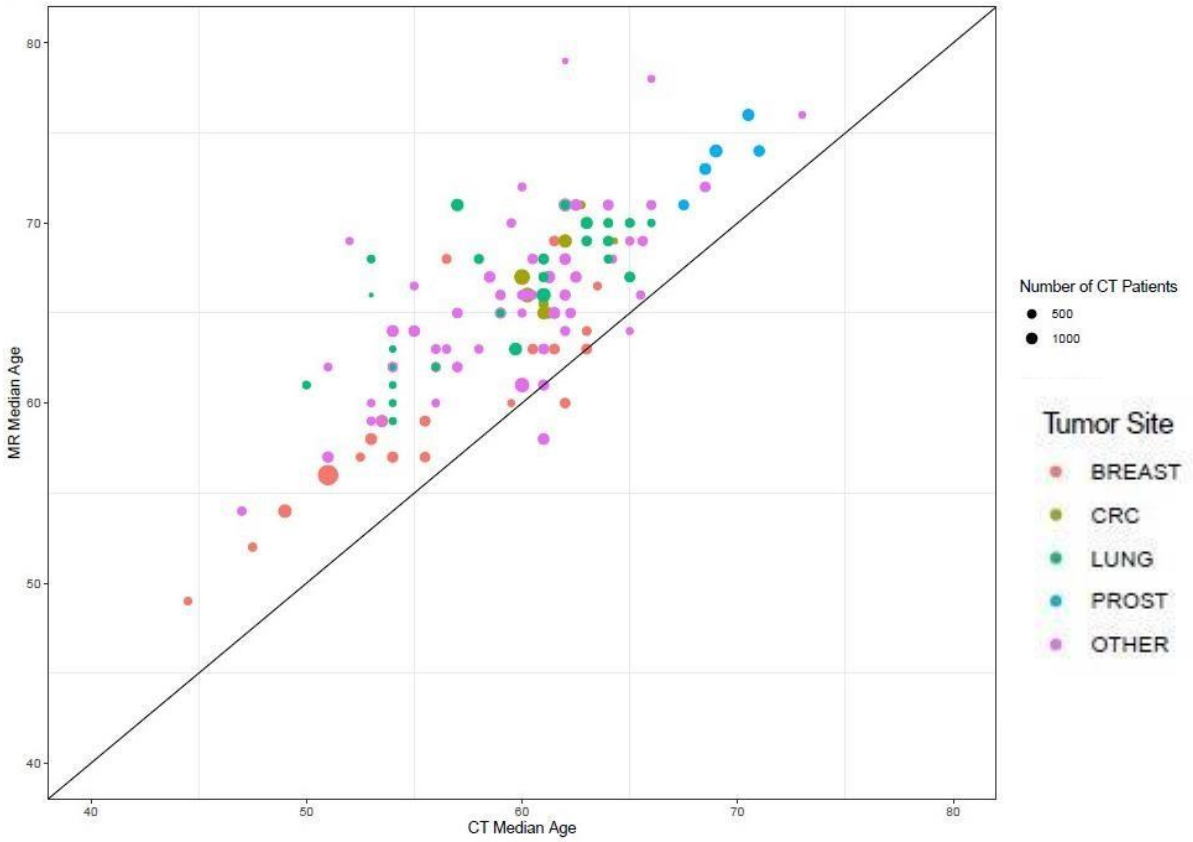
Iacovino ML,. Lancet Reg Health Eur. 2024 Apr 22;41:100912. doi: 10.1016/j.lanepe.2024.100912.



Iacovino ML,. Lancet Reg Health Eur. 2024 Apr 22;41:100912. doi: 10.1016/j.lanepe.2024.100912.

Results: continuous age

	Monitoring Registries	Clinical Trials	ΔMR-CT
Median age, weighted mean (SD)	66.2 (4.6)	59.8 (5.2)	**+5.3 (2.9)**

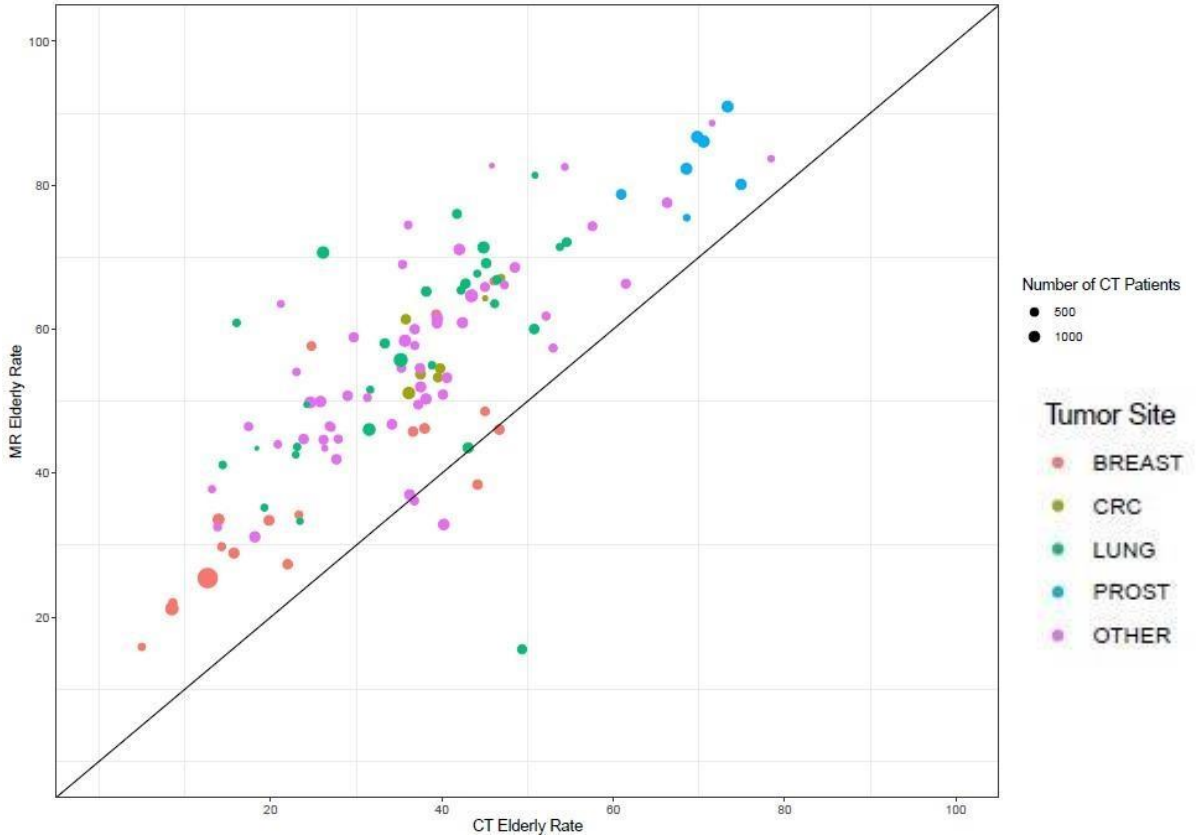


Tumor site	Δ median age
Breast	+4.1 (2.4)
CRC	+5.7 (1.3)
Lung	+3.5 (3.2)
Prostate	+4.5 (0.9)
Other	
- **GU**	+5.3 (2.3)
- **GYN**	+3.0 (2.6)
- **H&N**	+7.2 (2.4)
- **Melanoma**	+6.6 (4.0)
- **NET**	+3.3 (2.7)
- **Skin**	+9.6 (4.8)
- **Sarcoma**	+6.1 (0.7)
- **UP-GI**	+5.9 (2.6)

Iacovino ML,. Lancet Reg Health Eur. 2024 Apr 22;41:100912. doi: 10.1016/j.lanepe.2024.100912.

Results: categorical age

	Monitoring Registries	Clinical Trials	ΔMR-CT
Elderly rate (%), weighted mean (SD)	56.5 (15.8)	36.8 (16.3)	**+16.8 (10.1)**

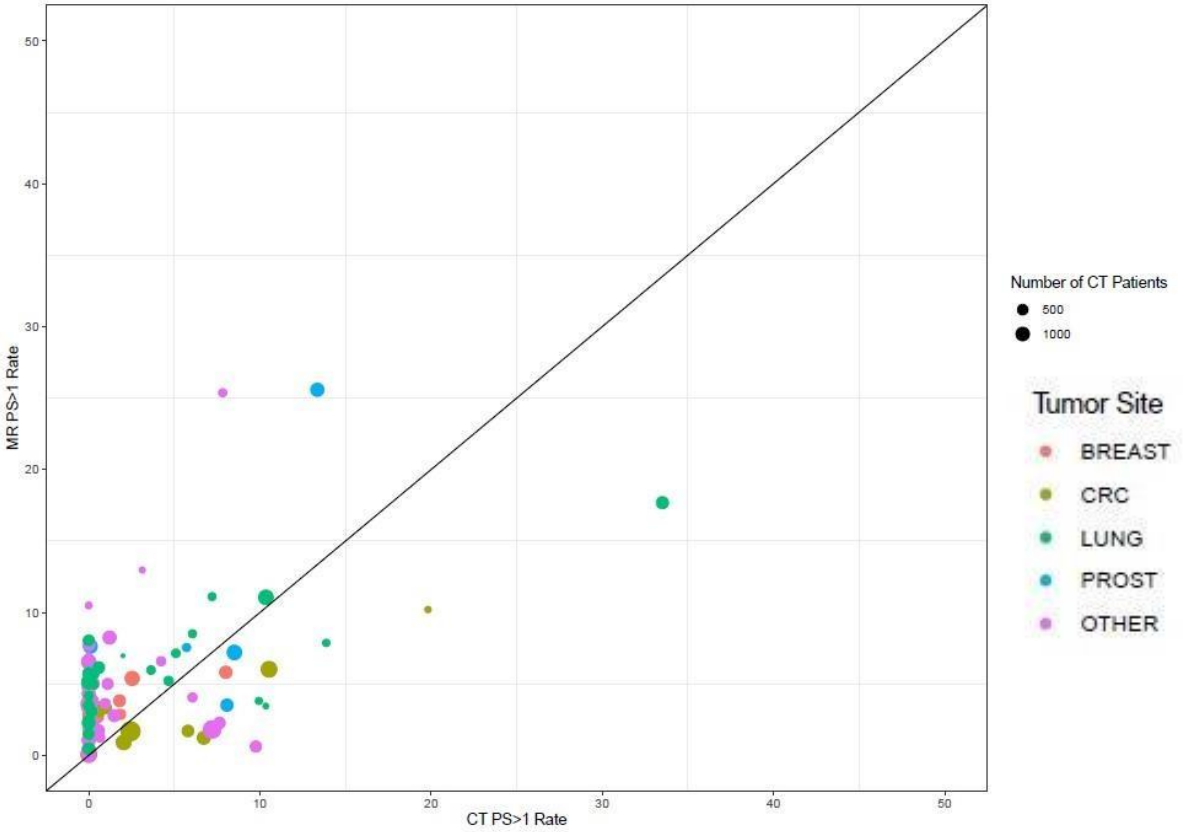


Tumor site	Δ elderly rate
Breast	+12.2 (7.1)
CRC	+17.1 (4.0)
Lung	+20.3 (14.7)
Prostate	+14.3 (4.3)
Other	
- **GU**	+18.0 (6.7)
- **GYN**	+12.0 (8.4)
- **H&N**	+24.4 (3.8)
- **Melanoma**	+17.2 (10.0)
- **NET**	+18.7 (1.0)
- **Skin**	+20.1 (11.5)
- **Sarcoma**	+24.6 (-)
- **UP-GI**	+19.4 (9.3)

Iacovino ML,. Lancet Reg Health Eur. 2024 Apr 22;41:100912. doi: 10.1016/j.lanepe.2024.100912.

Results – performance status

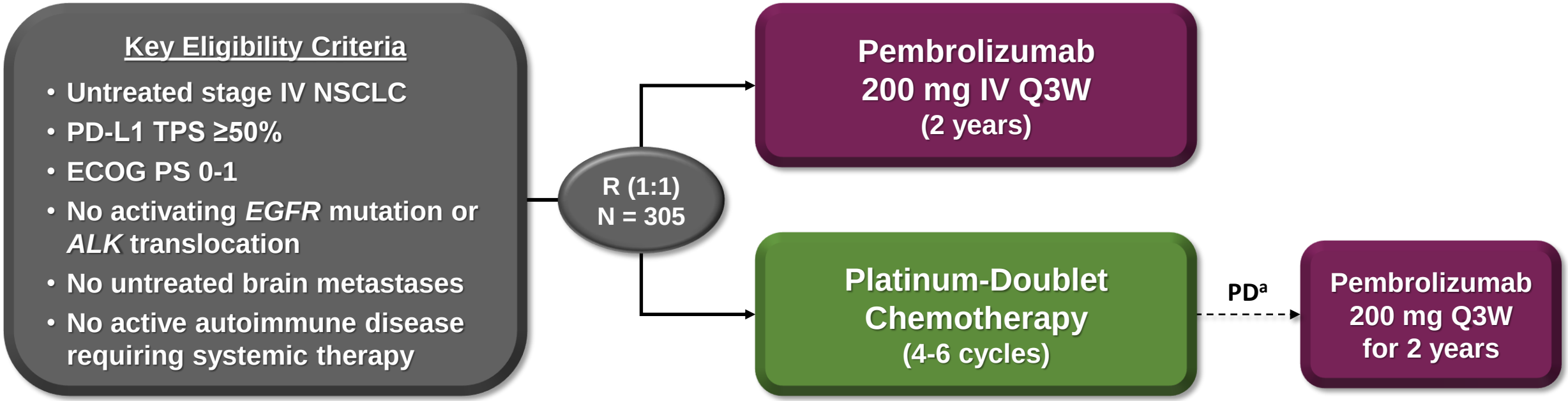
	Monitoring Registries	Clinical Trials	ΔMR-CT
PS >1 rate (%), weighted mean (SD)	4.2 (3.6)	3.1 (5.0)	**+1.3 (4.4)**



Tumor site	Δ PS>1 rate
Breast	+1.1 (2.7)
CRC	-2.1 (2.5)
Lung	+1.7 (5.2)
Prostate	+3.6 (6.4)
Other	
- **GU**	+4.0 (2.1)
- **GYN**	-2.2 (2.9)
- **H&N**	+4.1 (3.1)
- **Melanoma**	+1.9 (2.9)
- **NET**	NA
- **Skin**	+10.3 (4.6)
- **Sarcoma**	+4.7 (1.3)
- **UP-GI**	-2.3 (4.9)

Iacovino ML,. Lancet Reg Health Eur. 2024 Apr 22;41:100912. doi: 10.1016/j.lanepe.2024.100912.

Pembrolizumab as first-line treatment of advanced NSCLC: KEYNOTE-024



Key End Points

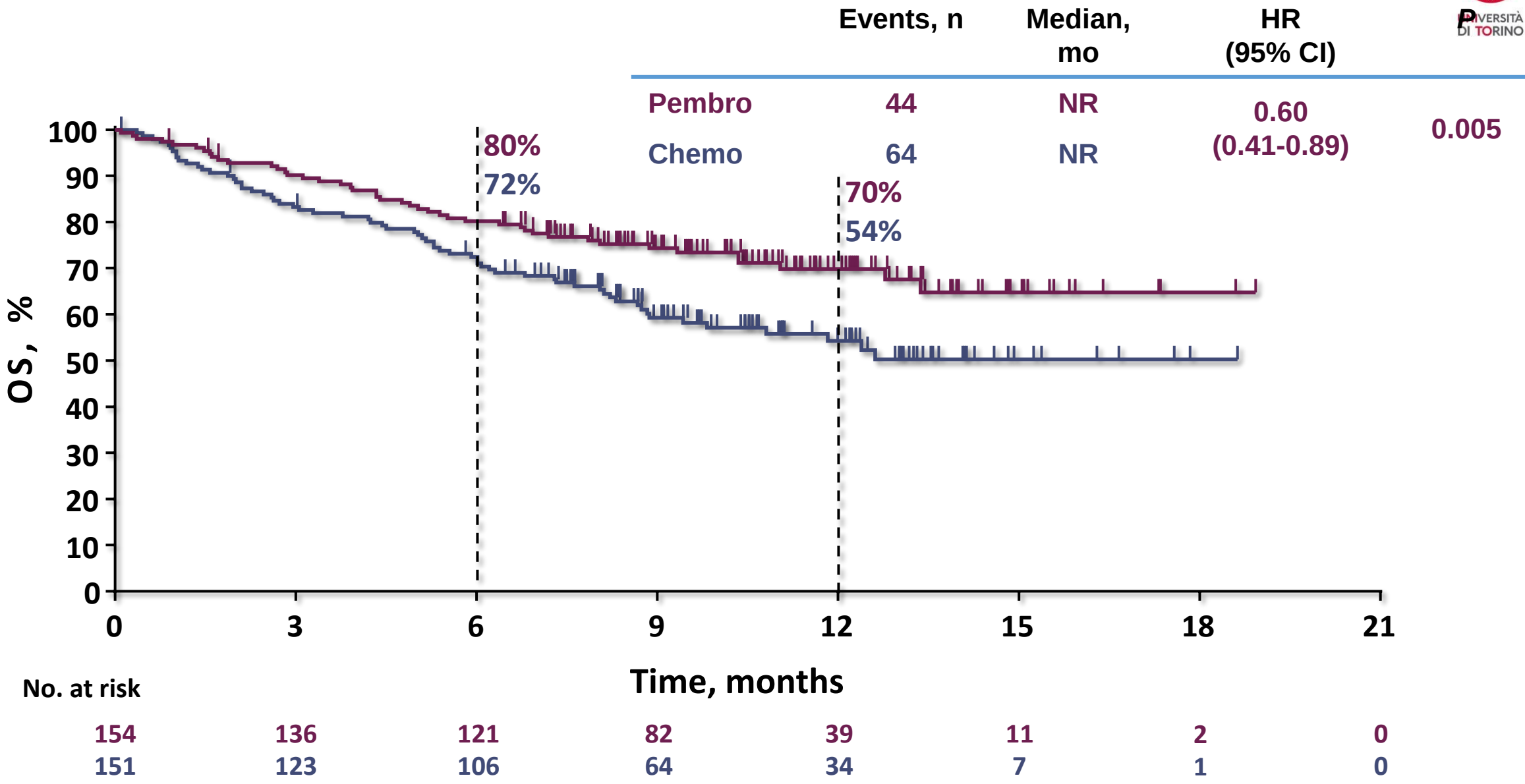
Primary: PFS (RECIST v1.1 per blinded, independent central review)

Secondary: OS, ORR, safety

Exploratory: DOR

Reck M et al, N Engl J Med 2016 Nov 10;375(19):1823-1833.

KEYNOTE 024: overall survival



European Journal of Cancer 130 (2020) 155–167

Available online at www.sciencedirect.com

ScienceDirect

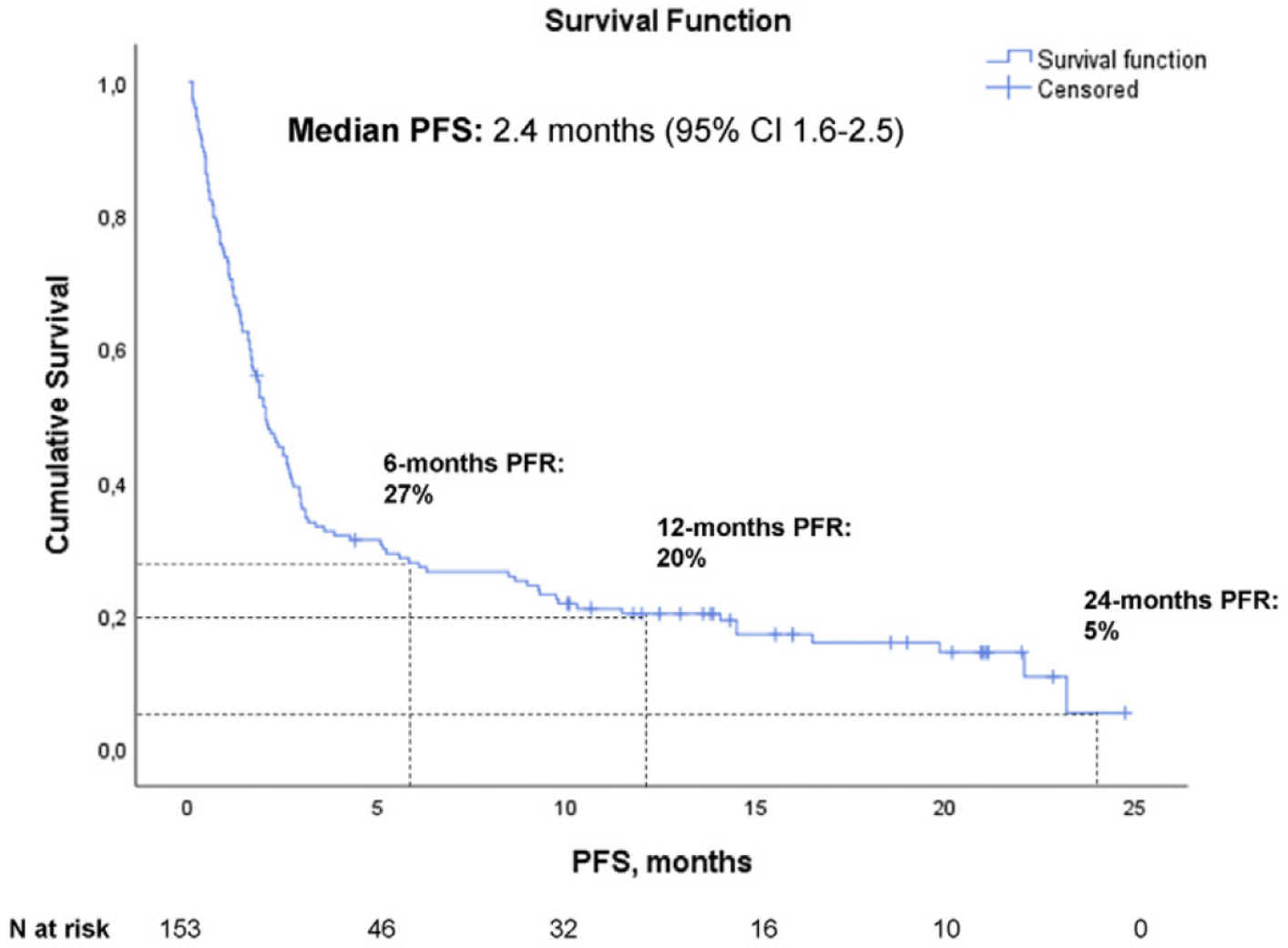
journal homepage: www.ejcancer.com

Original Research

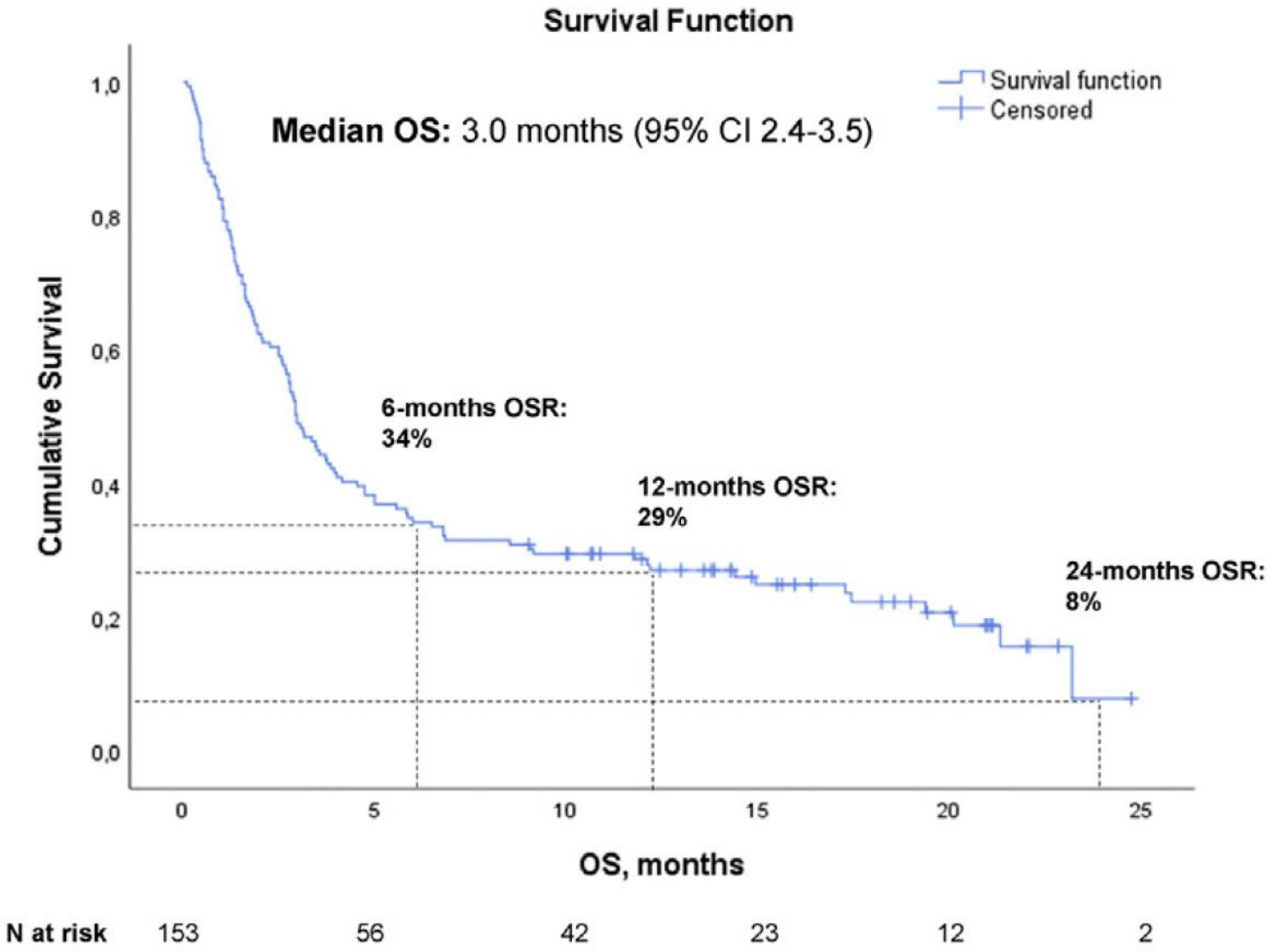
First-line pembrolizumab in advanced non–small cell lung cancer patients with poor performance status

Francesco Facchinetti [a,b,*], Giulia Mazzaschi [a], Fausto Barbieri [c], Francesco Passiglia [d], Francesca Mazzoni [e], Rossana Berardi [f], Claudia Proto [g], Fabiana Letizia Cecere [h], Sara Pilotto [i], Vieri Scotti [j], Sabrina Rossi [k], Alessandro Del Conte [l], Emanuele Vita [m], Chiara Bennati [n], Andrea Ardizzoni [o], Giulio Cerea [p], Maria Rita Migliorino [q], Elisa Sala [r], Andrea Camerini [s], Alessandra Bearz [l], Elisa De Carlo [l], Francesca Zanelli [t], Giorgia Guaitoli [c], Marina Chiara Garassino [g], Lucia Pia Ciccone [j], Giulia Sartori [i], Luca Toschi [k], Filippo Gustavo Dall'Olio [o], Lorenza Landi [n], Elio Gregory Pizzutilo [p,u], Gabriele Bartoli [q], Cinzia Baldessari [c], Silvia Novello [d], Emilio Bria [m], Diego Luigi Cortinovis [r], Giulio Rossi [v], Antonio Rossi [w], Giuseppe Luigi Banna [x], Roberta Camisa [a], Massimo Di Maio [y], Marcello Tiseo [a,z]

Facchinetti F, …, Di Maio M, Tiseo M. Eur J Cancer. 2020 May;130:155-167.



Facchinetti F, …, Di Maio M, Tiseo M. Eur J Cancer. 2020 May;130:155-167.



Facchinetti F, …, Di Maio M, Tiseo M. Eur J Cancer. 2020 May;130:155-167.

Clinicians' Attitudes and Survival Outcomes for Patients with NSCLC and Poor Performance Status in the Immunotherapy Era: PICASO study (GOIRC-04-2020)

UNIVERSITÀ DI TORINO

ters. **Figure 2. PFS in the overall population (A), according to treatment received (B), ar burden to the poor PS (C).** Median follow-up 9.4 months (reverse Kaplan Meier).

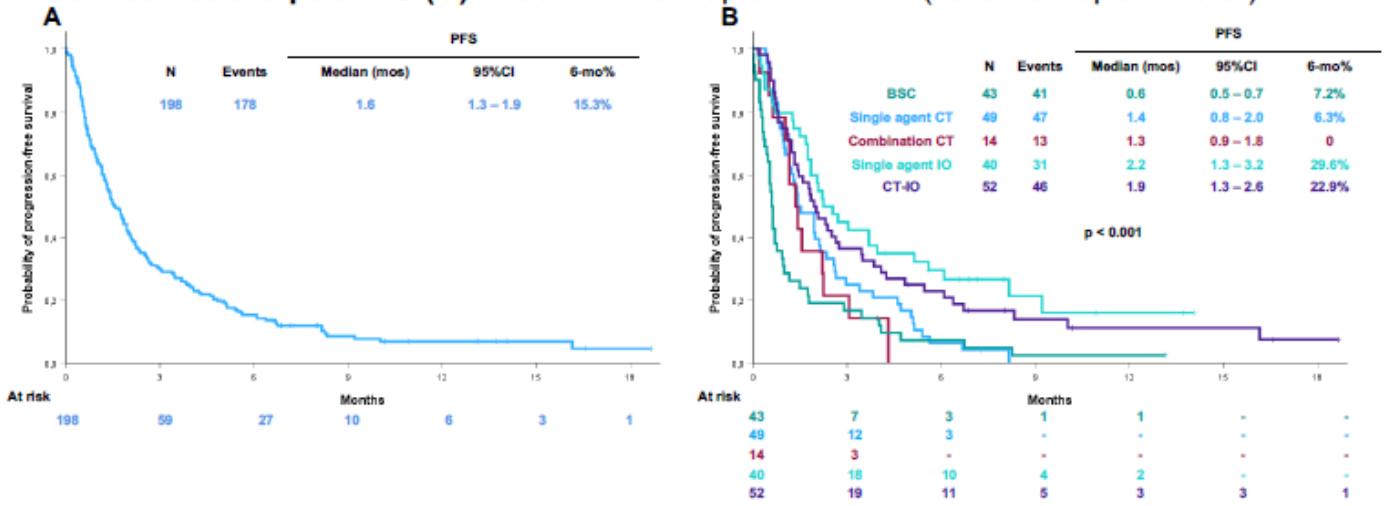
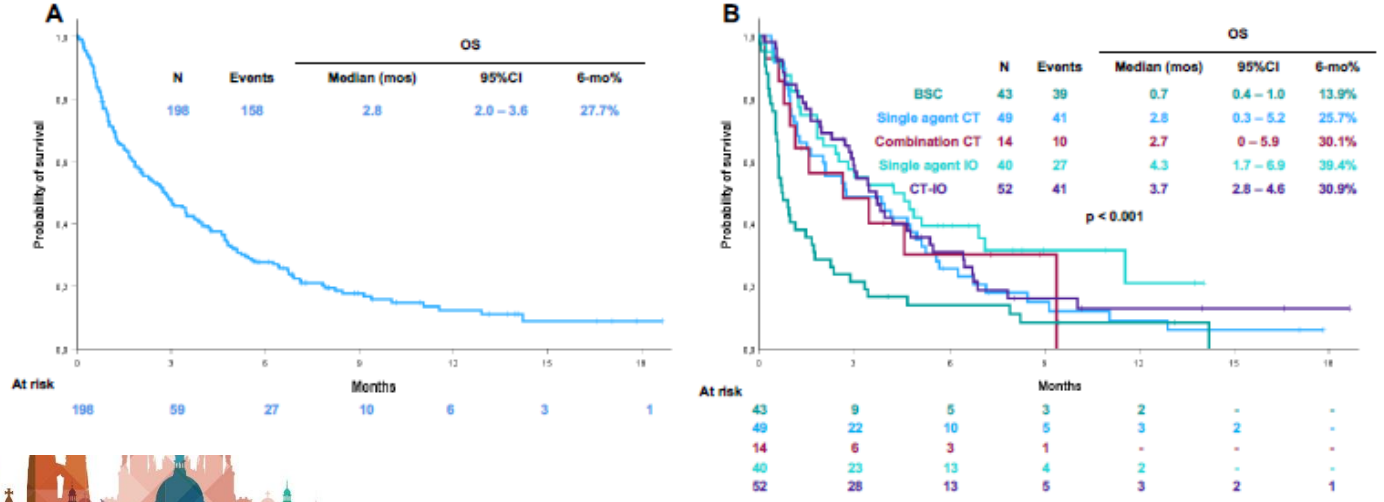


Figure 3. OS in the overall population (A), according to treatment received (B), and burden to the poor PS (C). Median follow-up 9.4 months (reverse Kaplan Meier).



BARCELONA 2024 ESMO congress

Facchinetti F, … Di Maio M, Tiseo M. Lung Cancer 2025

The Oncologist®

Commentary

Real-World Evidence in Oncology: Opportunities and Limitations

Massimo Di Maio,[a] Francesco Perrone,[b] Pierfranco Conte[c]

[a]Department of Oncology, University of Turin; Ordine Mauriziano Hospital, Torino, Italy; [b]Clinical Trial Unit, National Cancer Institute, IRCCS Fondazione Pascale, Napoli, Italy; [c]Department of Surgery, Oncology and Gastroenterology, University of Padova and Oncologia Medica 2, Istituto Oncologico Veneto, I.R.C.C.S., Padova, Italy

Disclosures of potential conflicts of interest may be found at the end of this article.

Key Words. Real-world evidence • Clinical trials • Cancer treatments

Traditionally, randomized controlled clinical trials (RCTs) have been considered the highest level of evidence to define the efficacy of treatments, before their adoption in clinical practice. However, in oncology, like in other fields of medicine, the analysis of real-world evidence (RWE) to answer clinical and

and treat in daily clinical practice [4, 5]. This is due to stringent eligibility criteria, such as good performance status and absence of clinically relevant concomitant diseases. For instance, the analysis of the Investigational New Drugs applications submitted in 2015 to the U.S. Food and Drug Administration, for

Di Maio, Perrone, Conte. The Oncologist 2020

ESMO daily REPORTER

An ESMO study reveals that poor quality of real-world data from oncology studies is still an issue

21 Oct 2023 | Targeted Therapy | Therapy | Clinical Research | Cancer Research | ESMO Congress 2023

ESMO Congress 2023 taking place in Madrid, Spain (20-24 October)

Scarcity of prospective data was reported as well as a limited use of real-world evidence as definitive evidence to regulatory approval decisions of targeted therapies in Europe

https://dailyreporter.esmo.org/esmo-congress-2023/digital-oncology/an-esmo-study-reveals-that-poor-quality-of-real-world-data-from-oncology-studies-is-still-an-issue

UNIVERSITÀ DI TORINO

European Journal of Cancer 155 (2021) 136–144

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

Real-world outcomes associated with new cancer medicines approved by the Food and Drug Administration and European Medicines Agency: A retrospective cohort study

Jemma M. Boyle [a,1], Gemma Hegarty [b,1], Christopher Frampton [c], Elizabeth Harvey-Jones [d], Joanna Dodkins [d], Katharina Beyer [e], Gincy George [e], Richard Sullivan [d,f], Christopher Booth [g,2], Ajay Aggarwal [a,d,f,*,2]

Boyle et al, European Journal of Cancer 2021; 153:136-144

Table 1
Characteristics of the RWD studies (n = 293 for 45 drug indications) identified for FDA and EMA approved indications, including RWD studies included in the survival analyses (n = 224 for 37 drug indications).

	Observational RWD Studies		Observational RWD studies included in survival analyses	
	No.	**%**	**No.**	**%**
Country				
Italy	52	17.7	44	19.6
Japan	30	10.2	22	9.8
China	28	9.6	22	9.8
USA	26	8.9	16	7.1
France	22	7.5	18	8.0
Spain	15	5.1	12	5.4
Canada	14	4.8	9	4.0
South Korea	13	4.4	12	5.4
UK	8	2.7	6	2.7
Netherlands	8	2.7	8	3.6
Poland	7	2.4	5	2.2
Multi-country	14	4.8	9	4.0
Other individual country	56	19.1	41	18.3
Study Type				
Prospective	50	17.1	39	17.4
Retrospective	242	82.6	184	82.1
Unknown	1	0·3	1	0·4
Multicentre				
Yes	180	61.4	135	60.3
No	112	38.2	88	39.3
Unknown	1	0·3	1	0·4

Boyle et al, European Journal of Cancer 2021; 153:136-144

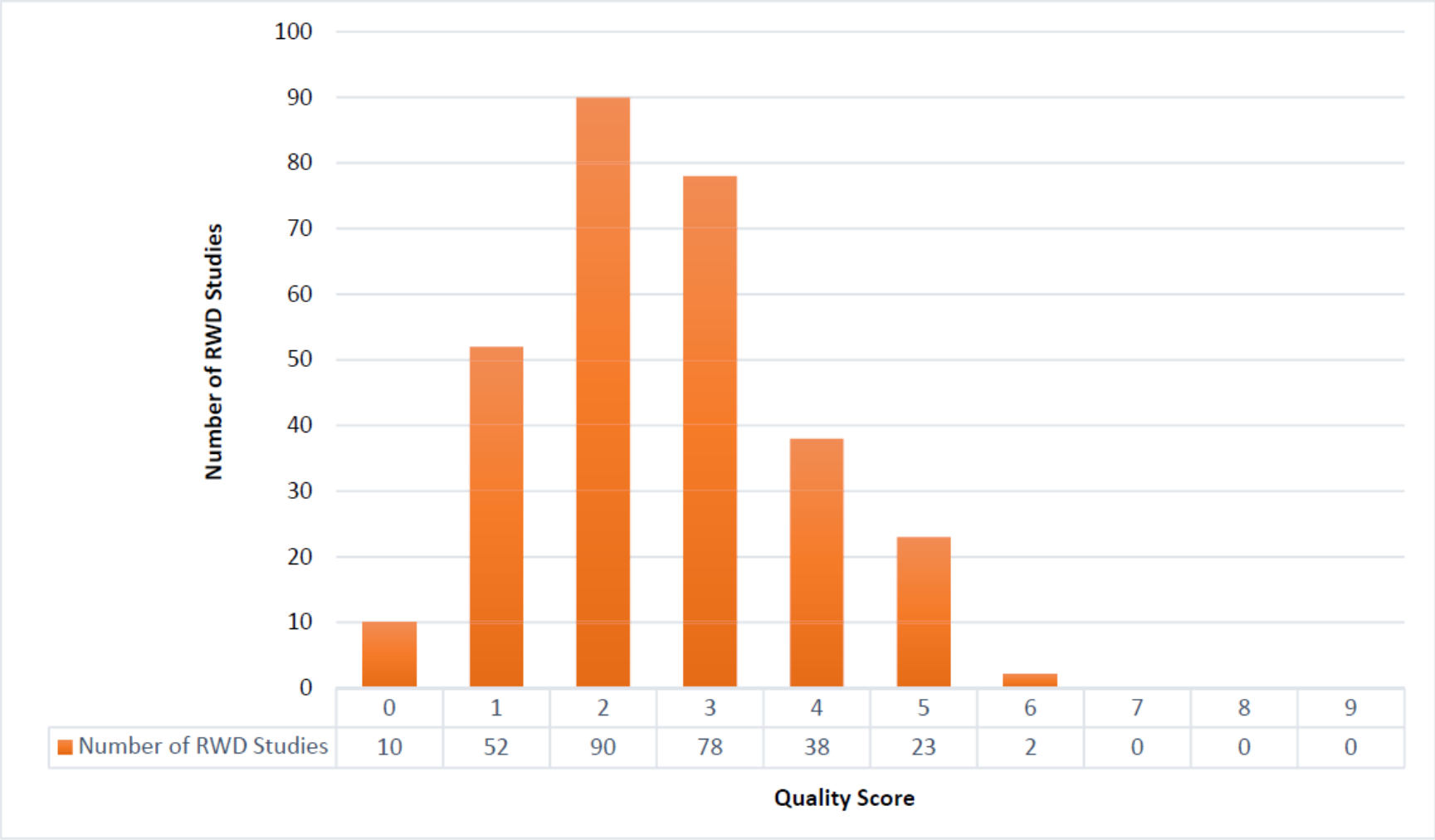
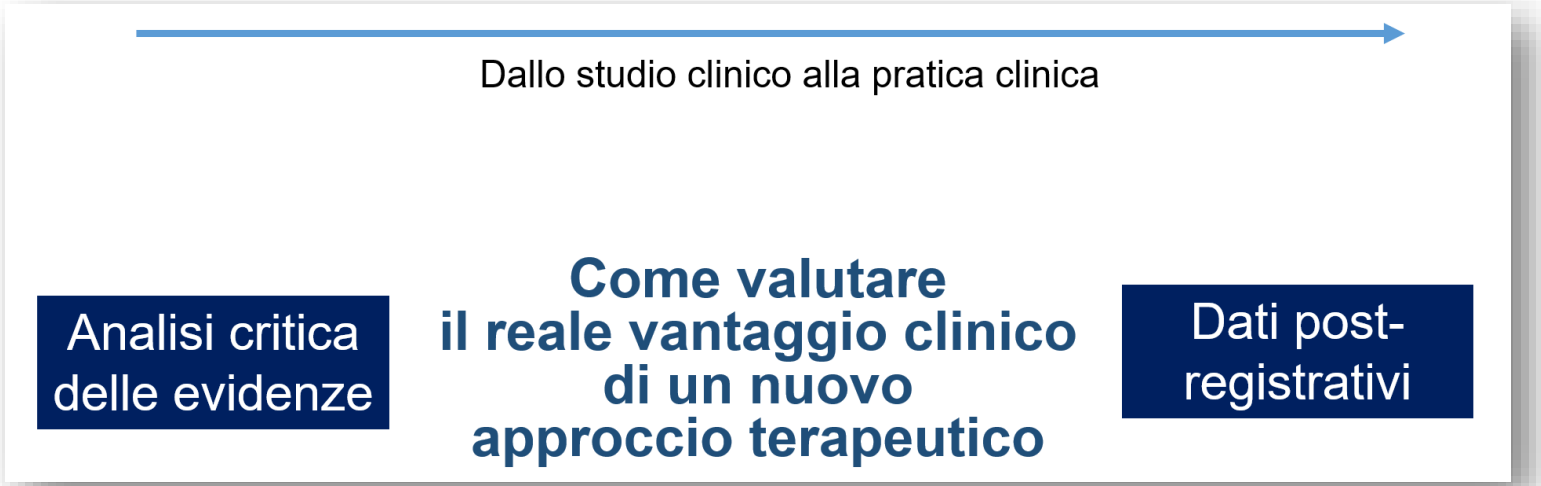


Fig. 1. Histogram of the distribution of total scores for RWD studies appraised using the Newcastle Ottawa Scale.

Boyle et al, European Journal of Cancer 2021; 153:136-144

Take home messages



- Gli studi clinici che hanno portato all'approvazione di un trattamento vanno letti criticamente, provando a valutare la validità interna dello studio e i limiti della sua applicabilità

- Le evidenze post-registrative possono aiutare a definire il reale valore del trattamento nel «real world», integrando i risultati degli studi e rispondendo ad alcuni quesiti senza risposta al momento dell'approvazione