La certezza delle Prove… (qualità delle evidenze)

Giovanni Pappagallo

Scuola di Metodologia della Ricerca Clinica
IRCCS "Sacro Cuore – Don Calabria"
Negrar di Valpolicella VR

La precisione della stima

Caso Vs Bias

CASO

↓

Errore Random

↓

Risultati Imprecisi

Errore in diminuzione con l'aumentare delle dimensioni del campione

BIAS

↓

Errore Sistematico

↓

Risultati Inesatti

Errore non influenzato dalle dimensioni del campione

Journal of Clinical Epidemiology 64 (2011) 1283–1293

GRADE guidelines 6. Rating the quality of evidence—imprecision

Gordon H. Guyatt[a,b,*], Andrew D. Oxman[c], Regina Kunz[d,e], Jan Brozek[a], Pablo Alonso-Coello[f], David Rind[g], PJ Devereaux[a], Victor M. Montori[h], Bo Freyschuss[i], Gunn Vist[c], Roman Jaeschke[b], John W. Williams Jr.[j], Mohammad Hassan Murad[h], David Sinclair[k], Yngve Falck-Ytter[l], Joerg Meerpohl[m,n], Craig Whittington[o], Kristian Thorlund[a], Jeff Andrews[p], Holger J. Schünemann[a,b]

Key Points

- GRADE's primary criterion for judging precision is to focus on the 95% confidence interval (CI) around the difference in effect between intervention and control for each outcome.

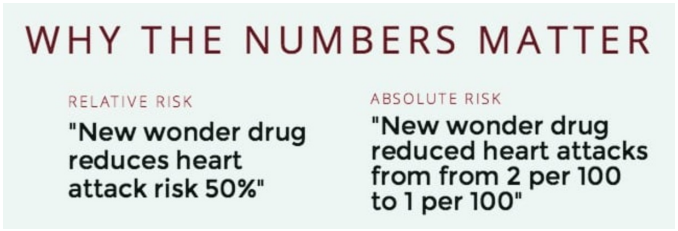
Journal of Clinical Epidemiology 64 (2011) 1283–1293

GRADE guidelines 6. Rating the quality of evidence—imprecision

Gordon H. Guyatt[a,b,*], Andrew D. Oxman[c], Regina Kunz[d,e], Jan Brozek[a], Pablo Alonso-Coello[f], David Rind[g], PJ Devereaux[a], Victor M. Montori[h], Bo Freyschuss[i], Gunn Vist[c], Roman Jaeschke[b], John W. Williams Jr.[j], Mohammad Hassan Murad[h], David Sinclair[k], Yngve Falck-Ytter[l], Joerg Meerpohl[m,n], Craig Whittington[o], Kristian Thorlund[a], Jeff Andrews[p], Holger J. Schünemann[a,b]

Key Points

- GRADE's primary criterion for judging precision is to focus on the 95% confidence interval (CI) around the difference in effect between intervention and control for each outcome.
- In general, the CIs to consider are those around the absolute, rather than the relative effect.



Journal of Clinical Epidemiology 64 (2011) 1283–1293

GRADE guidelines 6. Rating the quality of evidence—imprecision

Gordon H. Guyatt[a,b,*], Andrew D. Oxman[c], Regina Kunz[d,e], Jan Brozek[a], Pablo Alonso-Coello[f], David Rind[g], PJ Devereaux[a], Victor M. Montori[h], Bo Freyschuss[i], Gunn Vist[c], Roman Jaeschke[b], John W. Williams Jr.[j], Mohammad Hassan Murad[h], David Sinclair[k], Yngve Falck-Ytter[l], Joerg Meerpohl[m,n], Craig Whittington[o], Kristian Thorlund[a], Jeff Andrews[p], Holger J. Schünemann[a,b]

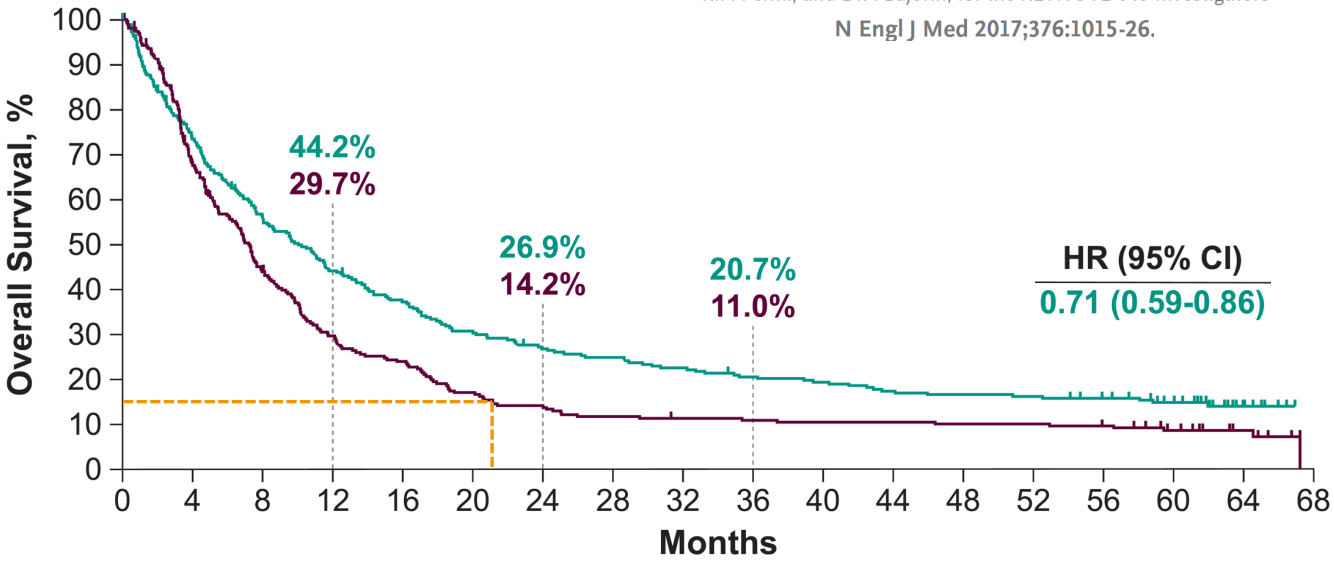
Key Points

- GRADE's primary criterion for judging precision is to focus on the 95% confidence interval (CI) around the difference in effect between intervention and control for each outcome.
- In general, the CIs to consider are those around the absolute, rather than the relative effect.
- If a recommendation or clinical course of action would differ if the upper versus the lower boundary of the CI represented the truth, consider the rating down for imprecision.

Pembrolizumab as Second-Line Therapy for Advanced Urothelial Carcinoma

J. Bellmunt, R. de Wit, D.J. Vaughn, Y. Fradet, J.-L. Lee, L. Fong, N.J. Vogelzang, M.A. Climent, D.P. Petrylak, T.K. Choueiri, A. Necchi, W. Gerritsen, H. Gurney, D.I. Quinn, S. Culine, C.N. Sternberg, Y. Mai, C.H. Poehlein, R.F. Perini, and D.F. Bajorin, for the KEYNOTE-045 Investigators*

N Engl J Med 2017;376:1015-26.



LC95% consistenti con una rilevante efficacia (↓18 eventi decesso / 100 paz.) e con un beneficio marginale (↓5 eventi decesso / 100 paz.)

NON seria imprecisione

- Median f.u.: 21.1 months
- Baseline risk* at median f.u.: 85%
- **Risk Difference**: 11 events lower / 100 pts (**95%CI: 18 lower to 5 lower**)
- NNT = 9

** J Clin Epidemiol 118 (2020) 124-131*

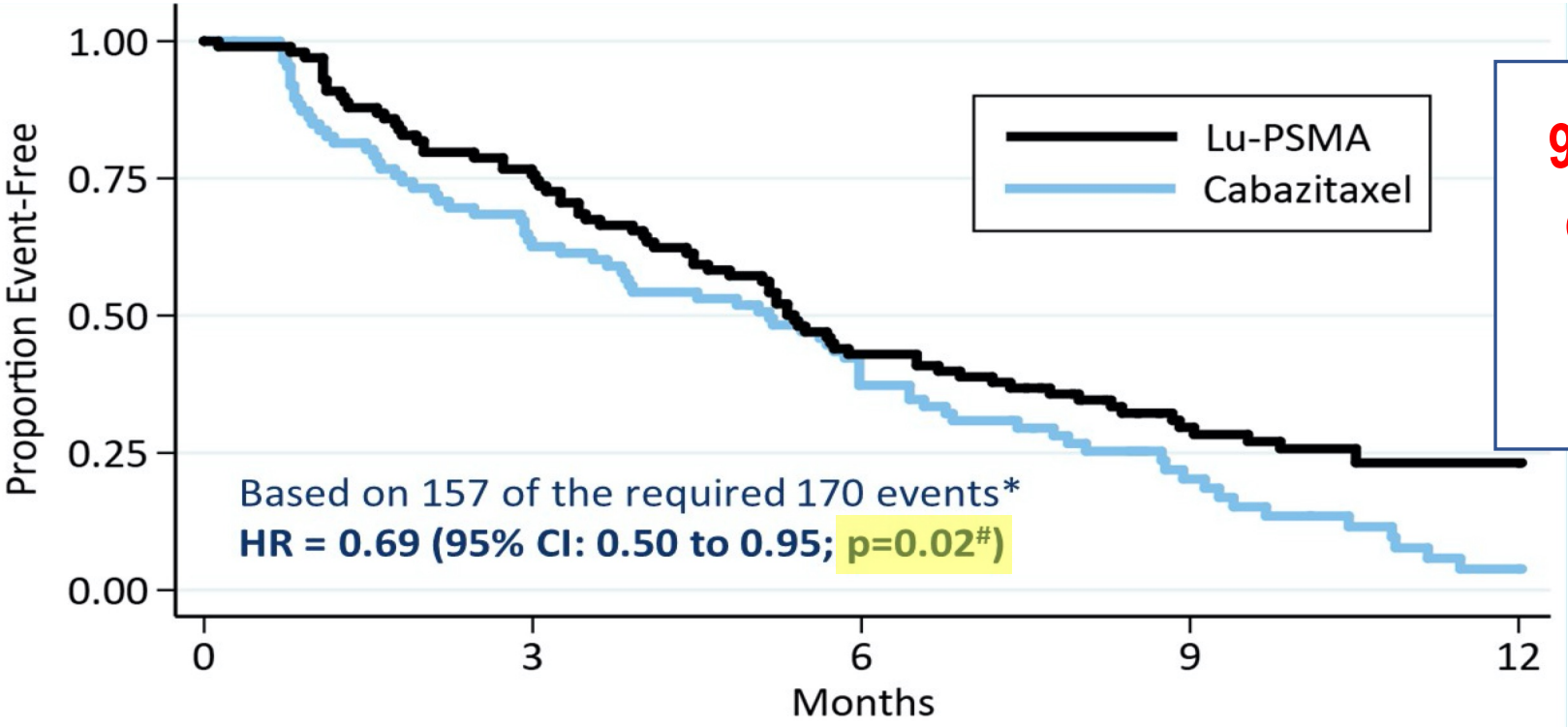
PRESENTED AT: 2020 ASCO ANNUAL MEETING #ASCO20 #TheraP

A randomised phase II trial of ^{177}Lu-PSMA-617 (Lu-PSMA) theranostic versus cabazitaxel in metastatic castration resistant prostate cancer (mCRPC) progressing after docetaxel: Initial results

TheraP (ANZUP 1603)

Michael Hofman, Louise Emmett, Shahneen Sandhu, Amir Iravani, Anthony Joshua, Jeffrey Goh, David Pattison, Hsiang Tan, Ian Kirkwood, Siobhan Ng, Roslyn Francis, Craig Gedye, Natalie Rutherford, Alison Zhang, Margaret McJannett, Martin Stockler, John Violet, Scott Williams, Andrew Martin, Ian Davis

- median follow-up: 13.3 months
- baseline risk: 95.1%
- absolute risk: 8 events fewer (95%CI: 1 fewer to 17 fewer)



95%CI of absolute effect consistent with opposite interpretations

NON Imprecisione Clinica anche in presenza di valore di P ≥0.05

M.I.D. EORTC QLQ-C30 GHS: **10 punti**

Scale	Cabazitaxel	Lu-PSMA	Diff.
	Pred. Mean (SE) {95% CI}	Pred. Mean (SE) {95% CI}	Pred. Mean (SE) {95% CI} [p-value]
Global health status / QoL	60.4 (1.8) {56.9 to 63.9}	63.4 (1.6) {60.3 to 66.5}	3.0 (2.3) {-1.6 to 7.5} [0.202]

IC95% dell'effetto assoluto compreso nel *range* di non rilevanza clinica (<10 punti)

NON Imprecisione

Lancet 2021; 397: 797–804


XXXII CONGRESSO NAZIONALE AIRO
XXXIII CONGRESSO NAZIONALE AIRB
XII CONGRESSO NAZIONALE AIRO GIOVANI
AIRO2022
Radioterapia di precisione per un'oncologia innovativa e sostenibile
BOLOGNA, 25-27 NOVEMBRE
PALAZZO DEI CONGRESSI

“C’è un bias”…

Caso Vs Bias

CASO

↓

Errore Random

↓

Risultati Imprecisi

Errore in diminuzione con l'aumentare delle dimensioni del campione

BIAS

↓

Errore Sistematico

↓

Risultati Inesatti

Errore non influenzato dalle dimensioni del campione

Il Bias

Distorsione dell'effetto di un intervento diagnostico-terapeutico per effetto di una componente estranea al caso:

- **selezione dei soggetti da assegnare a uno specifico trattamento, rimozione di alcuni pazienti dall'analisi, analisi non preordinata su sottogruppi di pazienti (*selection bias*)**

  ***rimedio:* randomizzazione (e stratificazione)**

- conoscenza da parte del paziente/medico/valutatore del trattamento in atto (*performance/detection bias*)

  rimedio: mascheramento

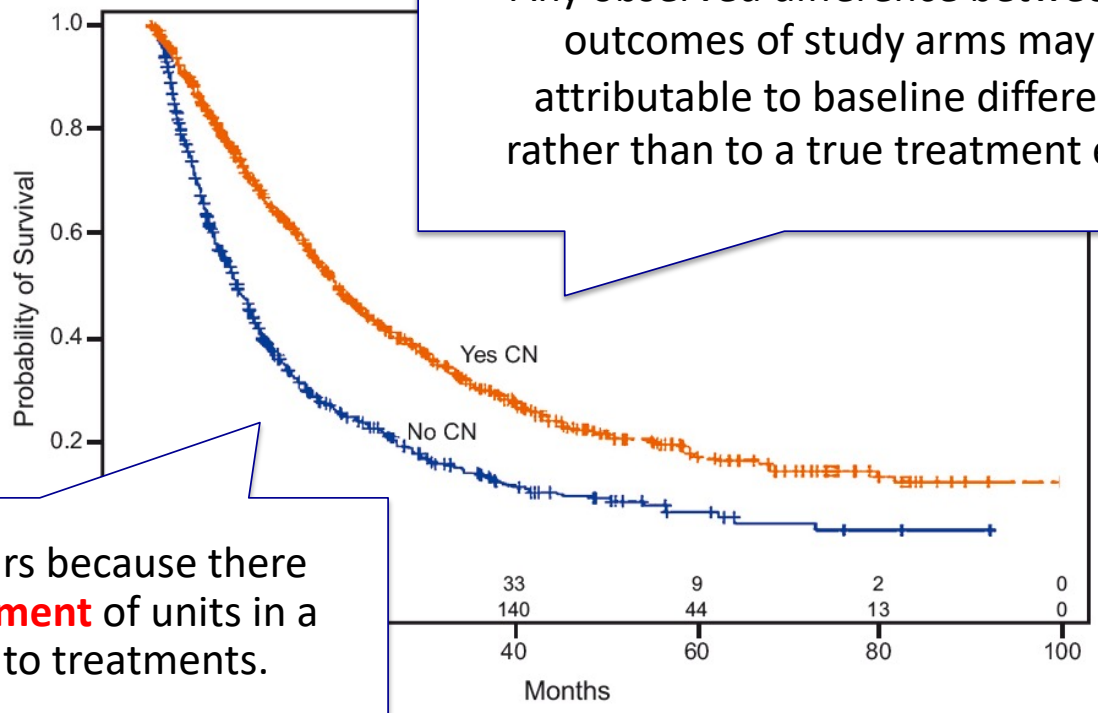
- selezione dei risultati da comunicare a mezzo presentazione/pubblicazione (*reporting bias*)

  rimedio: database degli studi clinici in corso

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

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

EUROPEAN UROLOGY 66 (2014) 7



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

A possible **bias** occurs because there is **no random assignment** of units in a target population to treatments.

Reconciling the Use of Cytoreductive Nephrectomy in the Targeted Therapy Era

Stephen H. Culp *

EUROPEAN UROLOGY 66 (2014) 711–712

Although retrospective, the results of this study are strengthened by the number of patients examined, inclusion of patients from institutions around the world, and lack of patient exclusion based on RCC histology or type of targeted agent.

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

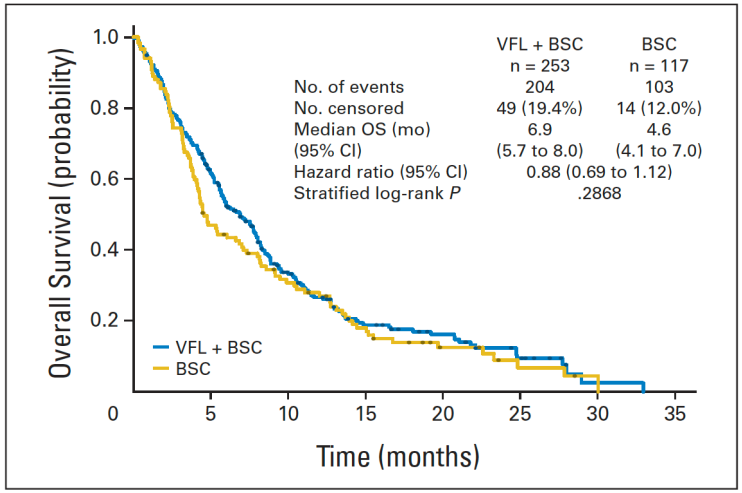


Fig 2. Overall survival (OS) in the intent-to-treat population (n = 370). VFL, vinflunine; BSC, best supportive care.

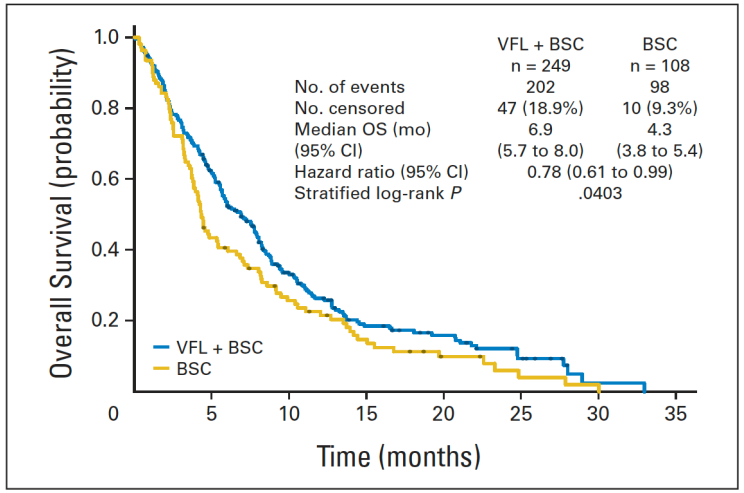


Fig 3. Overall survival (OS) in the eligible population (n = 357; 96.5% of intent-to-treat population). VFL, vinflunine; BSC, best supportive care.

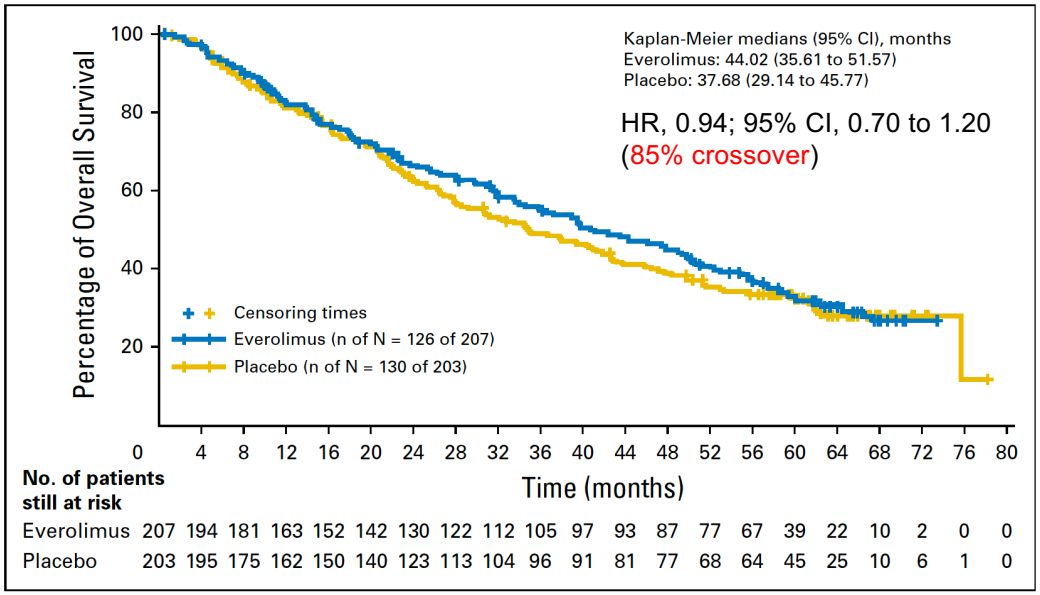
Although the objective of the median 2-month survival advantage favoring VFL BSC versus BSC was achieved (6.9 v 4.6 months, respectively), this difference was not statistically significant (P .287; Fig 2).

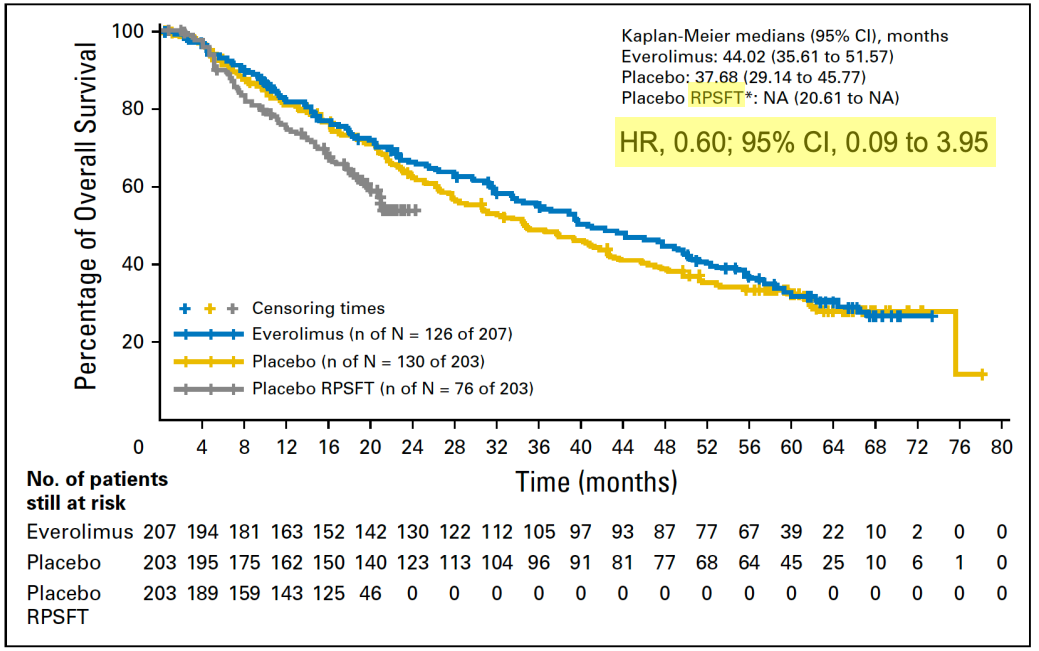
In the eligible population (Fig 3), the objective of achieving a 2-month survival difference in OS between the VFLBSC and BSC arms was met (6.9 v 4.3 months, respectively), and this difference is statistically significant (P. 040).

Everolimus for the Treatment of Advanced Pancreatic Neuroendocrine Tumors: Overall Survival and Circulating Biomarkers From the Randomized, Phase III RADIANT-3 Study

James C. Yao, Marianne Pavel, Catherine Lombard-Bohas, Eric Van Cutsem, Maurizio Voi, Ulrike Brandt, Wei He, David Chen, Jaume Capdevila, Elisabeth G.E. de Vries, Paola Tomassetti, Timothy Hobday, Rodney Pommier, and Kjell Öberg

J Clin Oncol 34:3906-3913. © 2016 by American Society of Clinical Oncology





Il Bias

Distorsione dell'effetto di un intervento diagnostico-terapeutico per effetto di una componente estranea al caso:

- selezione dei soggetti da assegnare a uno specifico trattamento, rimozione di alcuni pazienti dall'analisi, analisi non preordinata su sottogruppi di pazienti (*selection bias*)

  rimedio: randomizzazione (e stratificazione)

- **conoscenza da parte del paziente/medico/valutatore del trattamento in atto (*performance/detection bias*)**

  ***rimedio:* mascheramento**

- selezione dei risultati da comunicare a mezzo presentazione/pubblicazione (*reporting bias*)

  rimedio: database degli studi clinici in corso

Rischio di bias legato all'assenza di mascheramento

END POINTS AND ASSESSMENTS

The primary end point was progression-free survival, which was defined as the time from randomization to documented disease progression (as evaluated by independent central review by radiologists who were unaware of the treatment assignments) or death from any cause. Secondary end points included the objective response rate, overall survival (defined as the time from randomization to death from any cause), safety, and the side-effect profile.

Basso rischio di Detection Bias
(valutazione indipendente in cieco)

Alto rischio di Detection Bias

Alto rischio di Performance & Detection Bias

Basso rischio di Detection Bias
(per caratteristica intrinseca dell'outcome)

Lancet 2021; 397: 797–804

Il Bias

Distorsione dell'effetto di un intervento diagnostico-terapeutico per effetto di una componente estranea al caso:

- selezione dei soggetti da assegnare a uno specifico trattamento, rimozione di alcuni pazienti dall'analisi, analisi non preordinata su sottogruppi di pazienti (*selection bias*)

  rimedio: randomizzazione (e stratificazione)

- conoscenza da parte del paziente/medico/valutatore del trattamento in atto (*performance/detection bias*)

  rimedio: mascheramento

- **selezione dei risultati da comunicare a mezzo presentazione/pubblicazione (*reporting bias*)**

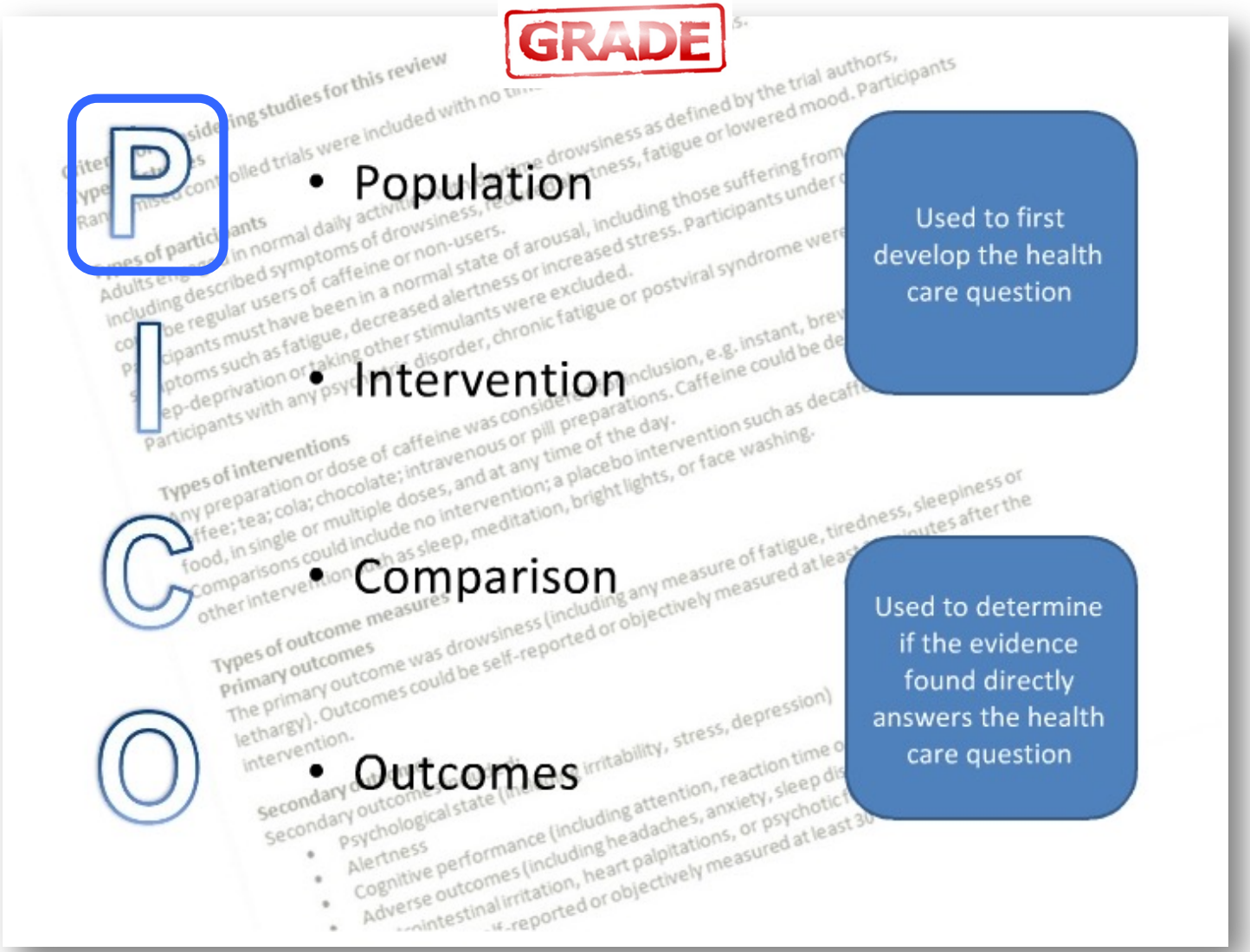
  ***rimedio:* database degli studi clinici in corso**

Trasferibilità dei risultati di uno studio

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

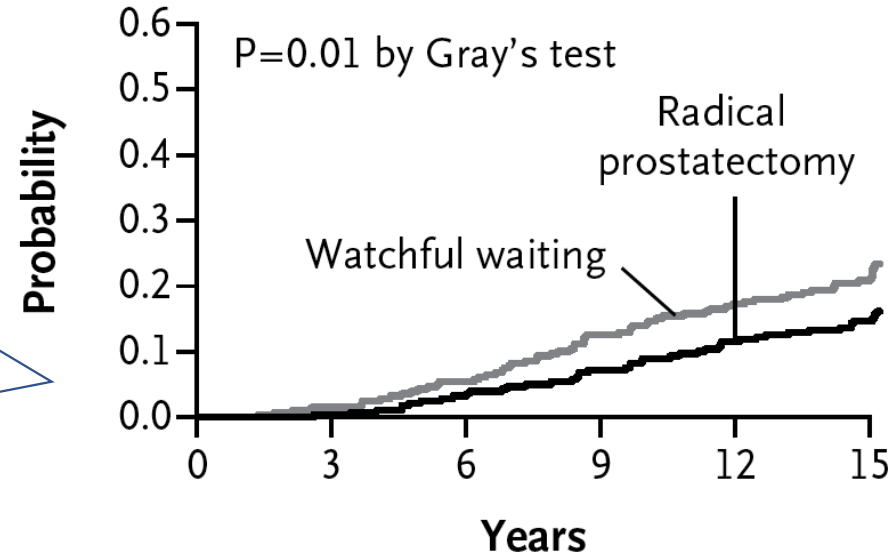

GRADE
P
• Population
I
• Intervention
C
• Comparison
O
• Outcomes
Used to first develop the health care question
Used to determine if the evidence found directly answers the health care question

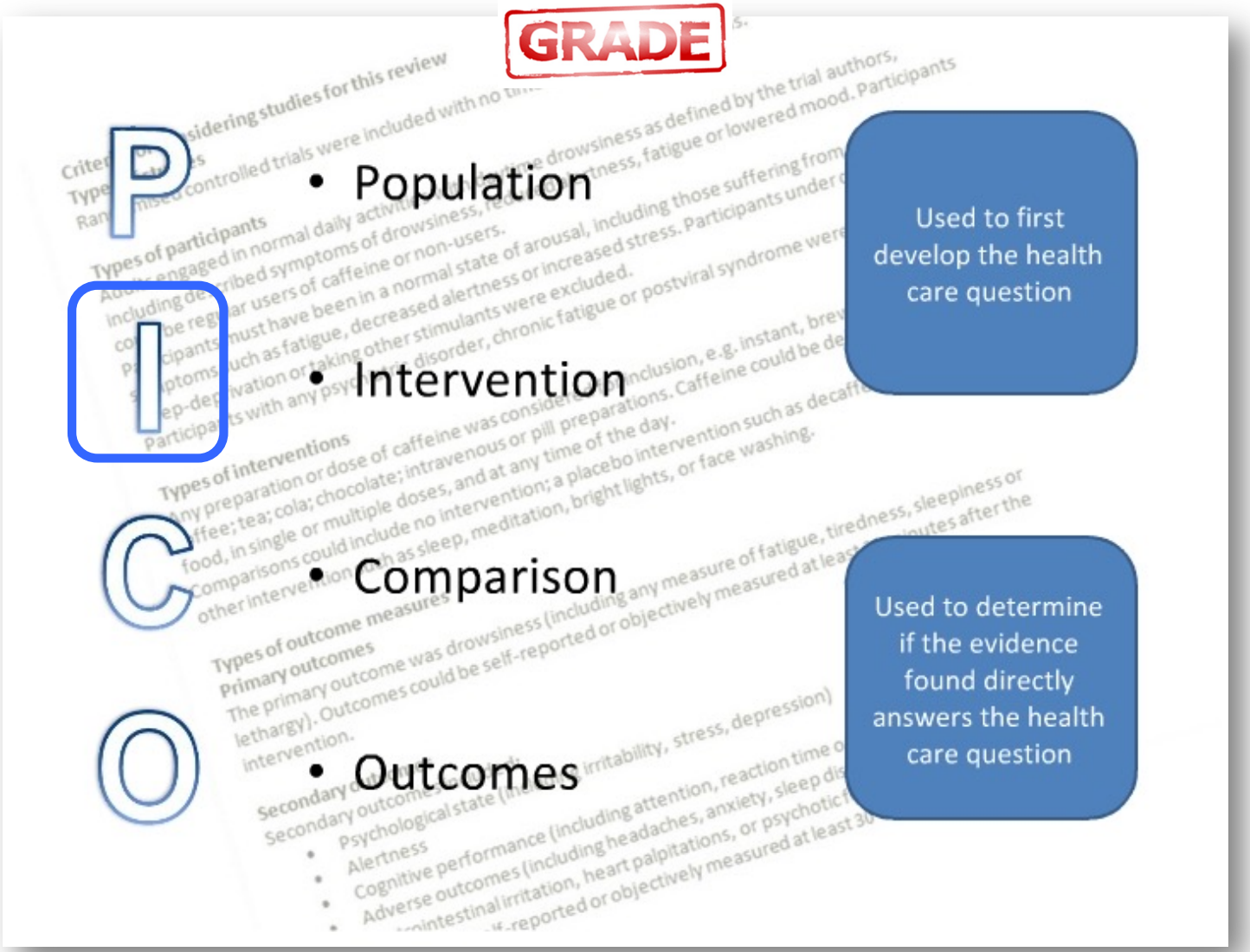
Radical Prostatectomy versus Watchful Waiting in Early Prostate Cancer

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.

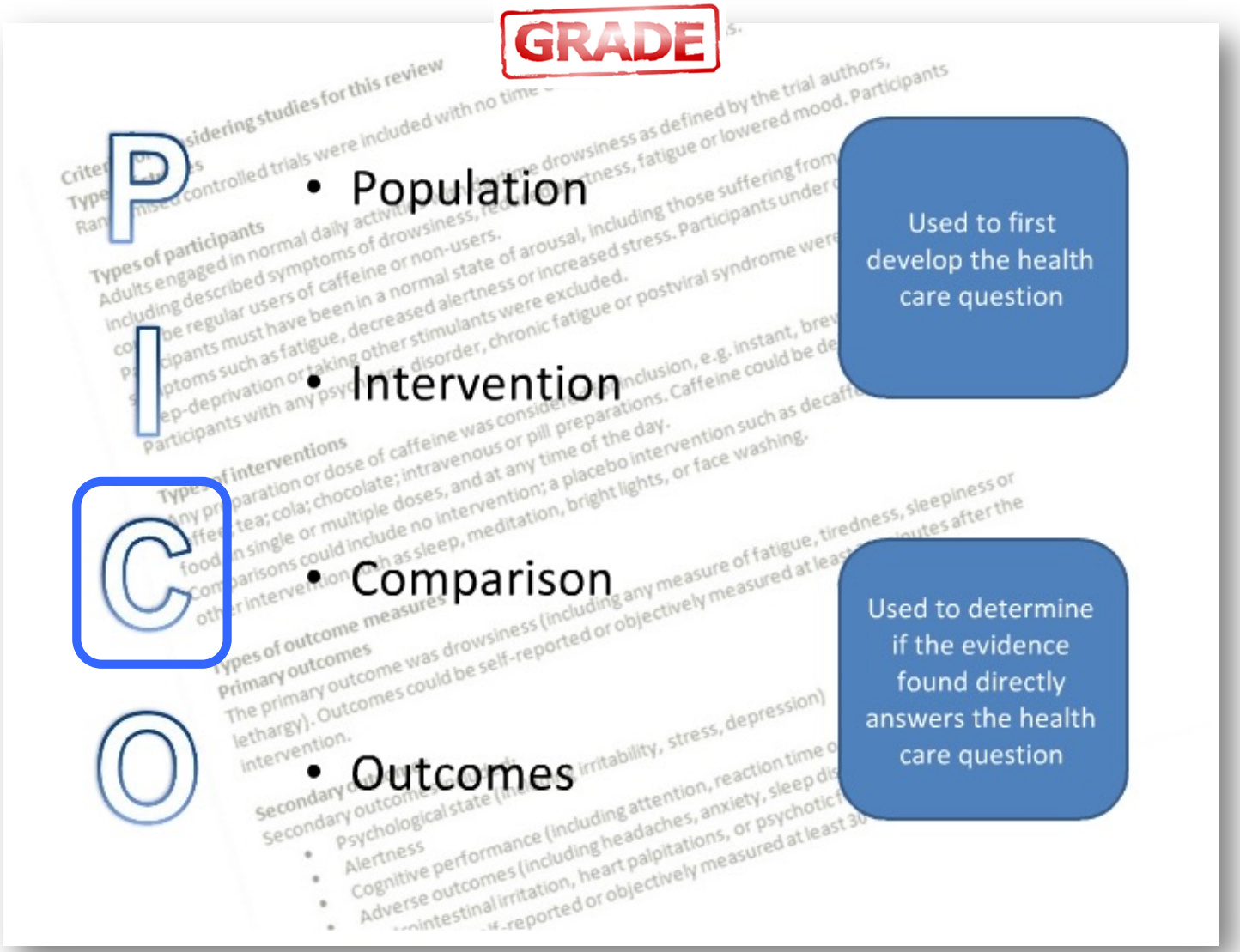
The present systemic review has the following limitations that must be taken into account.

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]

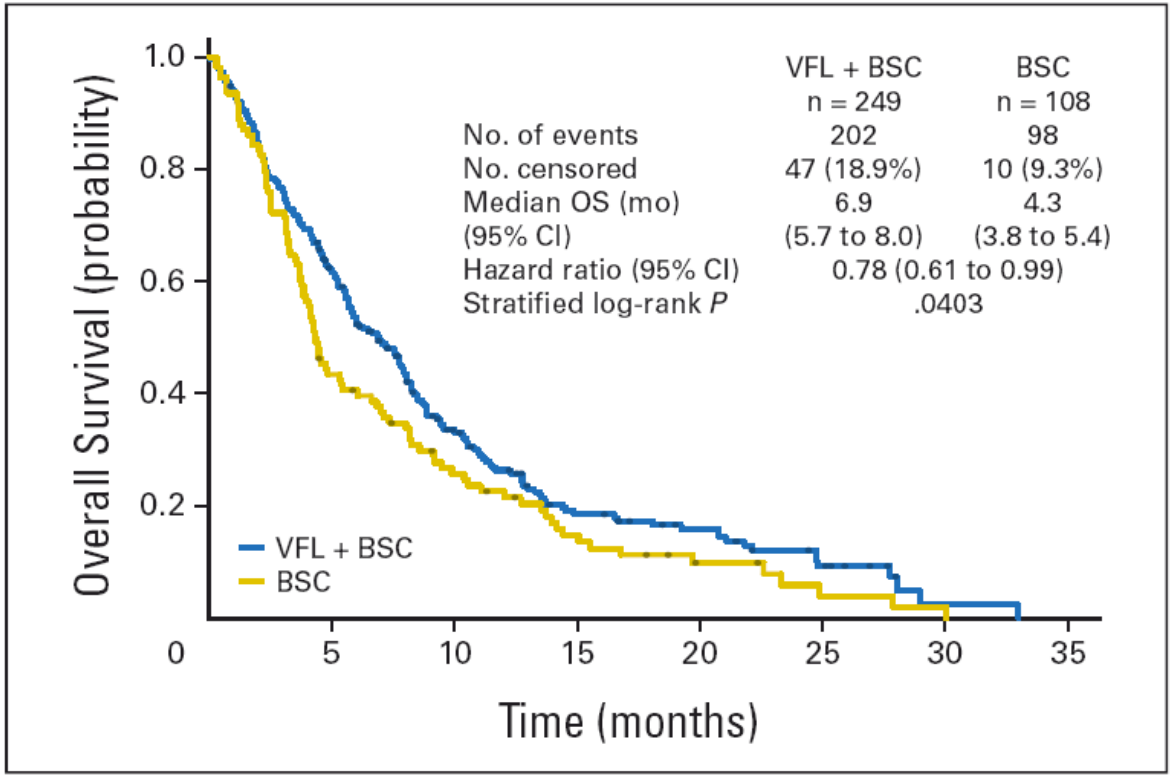
Indirectness per I. (di P.I.C.O.)

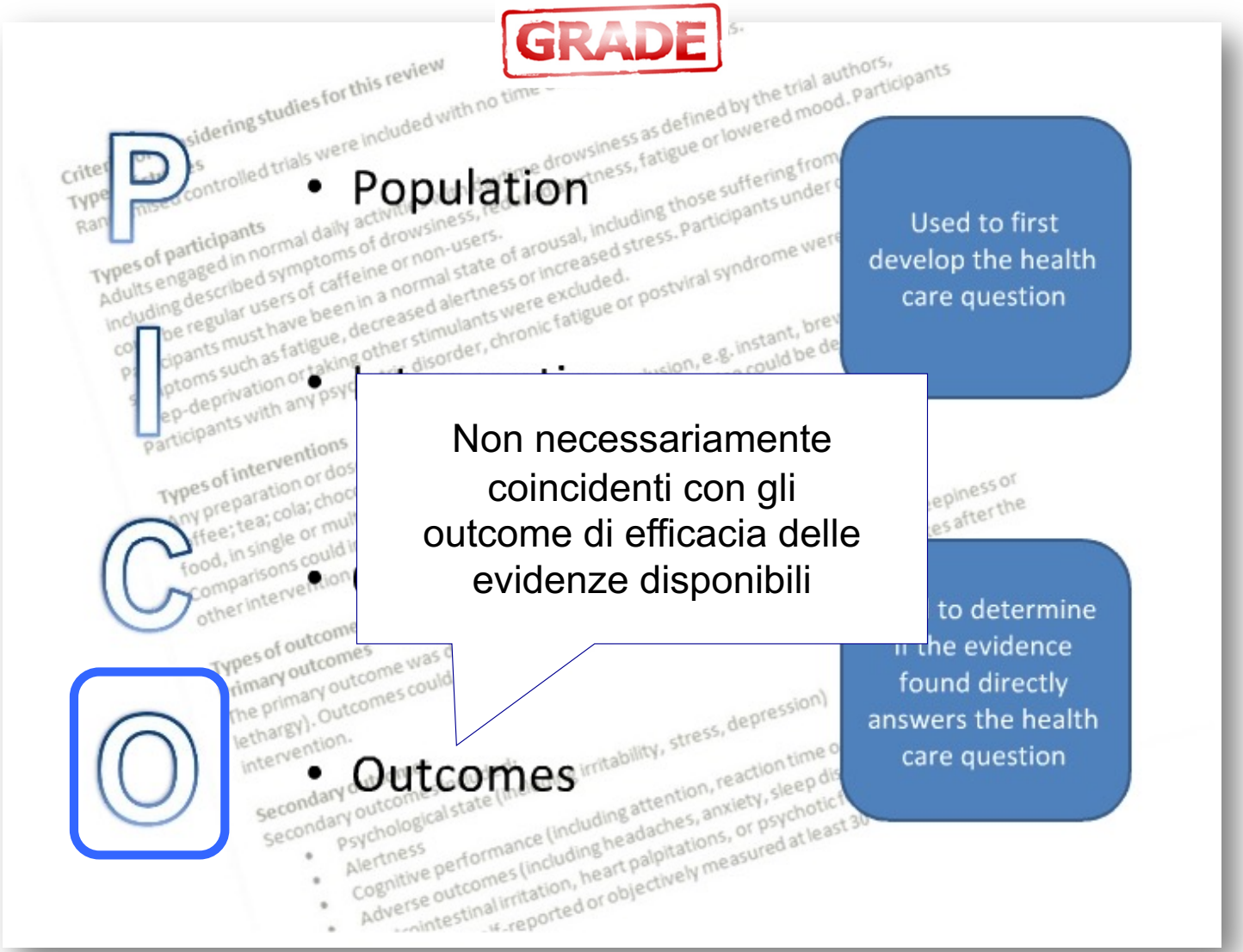

GRADE
P
• Population
I
• Intervention
C
• Comparison
O
• Outcomes
Used to first develop the health care question
Used to determine if the evidence found directly answers the health care question

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

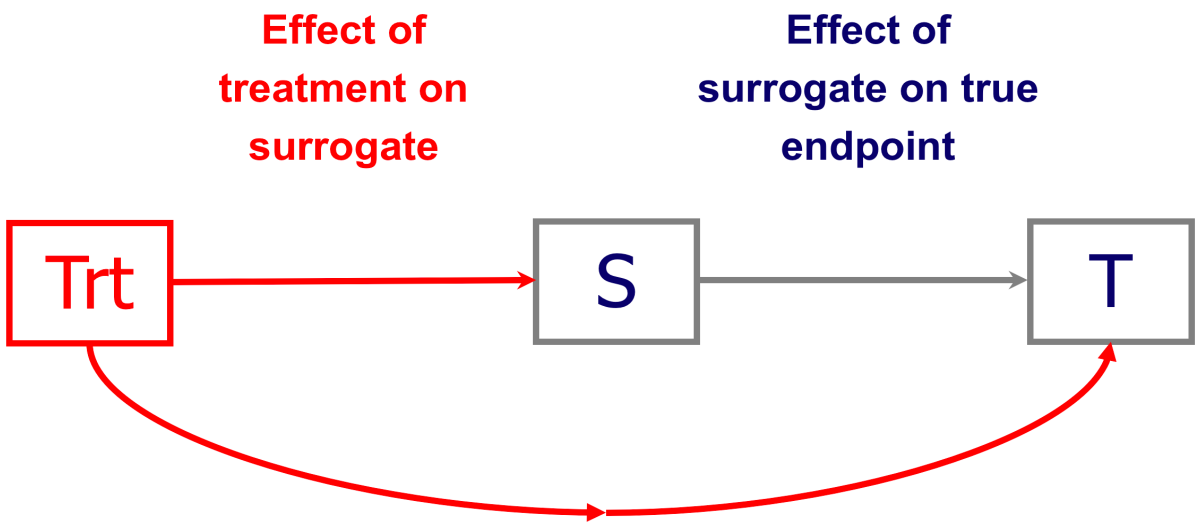
Joaquim Bellmunt, Christine Théodore, Tomasz Demkov, Boris Komyakov, Lisa Sengelov, Gedske Daugaard, Armelle Caty, Joan Carles, Agnieszka Jagiello-Gruszfeld, Oleg Karyakin, François-Michel Delgado, Patrick Hurteloup, Eric Winquist, Nassim Morsli, Yacine Salhi, Stéphane Culine, and Hans von der Maase

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




GRADE
P
• Population
I
C
O
• 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"

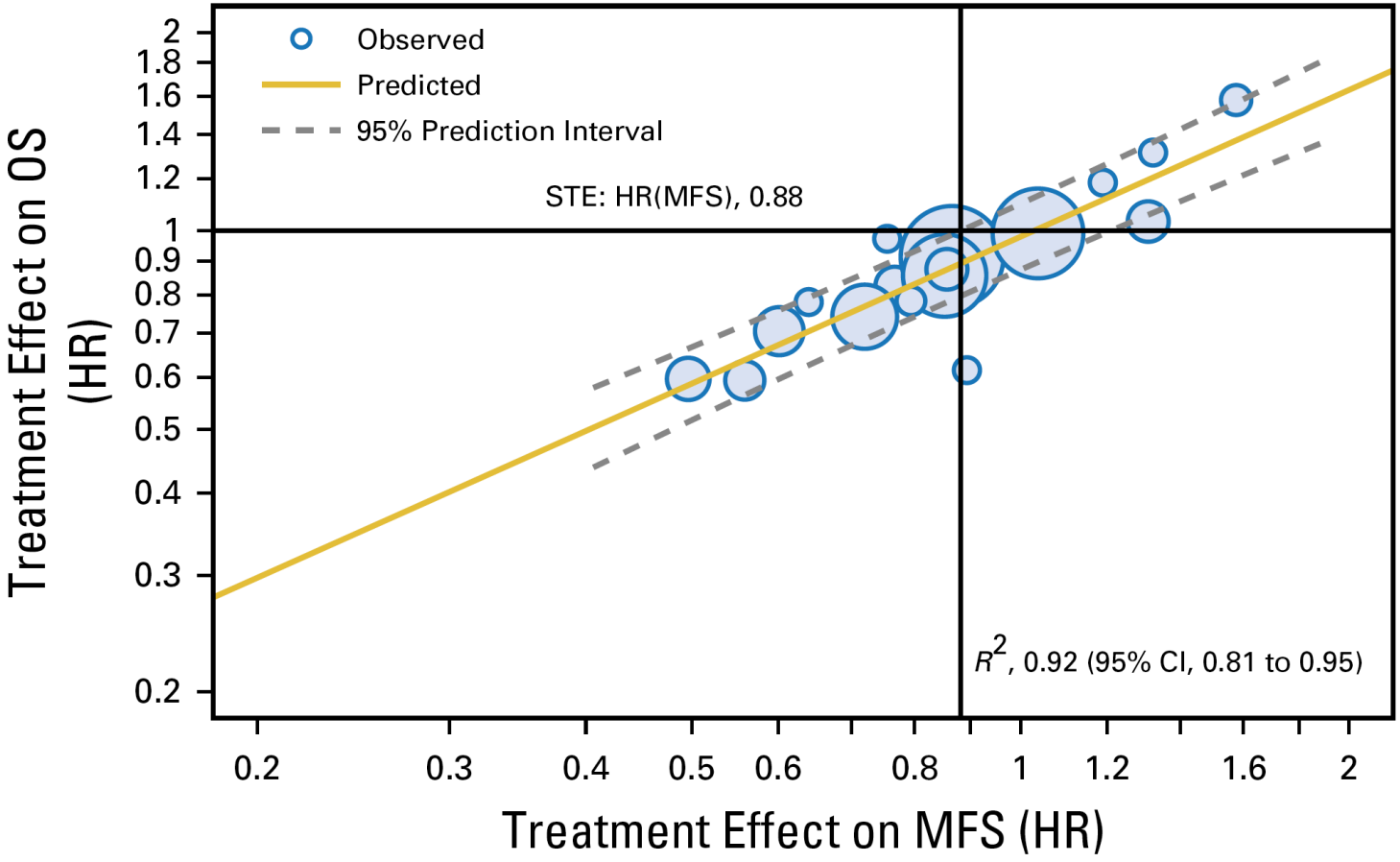


*Prentice, Statist Med 1989;**8**:431.*

Metastasis-Free Survival Is a Strong Surrogate of Overall Survival in Localized Prostate Cancer

Wanling Xie, Meredith M. Regan, Marc Buyse, Susan Halabi, Philip W. Kantoff, Oliver Sartor, Howard Soule, Noel W. Clarke, Laurence Collette, James J. Dignam, Karim Fizazi, Wendy R. Paruleker, Howard M. Sandler, Matthew R. Sydes, Bertrand Tombal, Scott G. Williams, and Christopher J. Sweeney, on behalf of the ICECaP Working Group

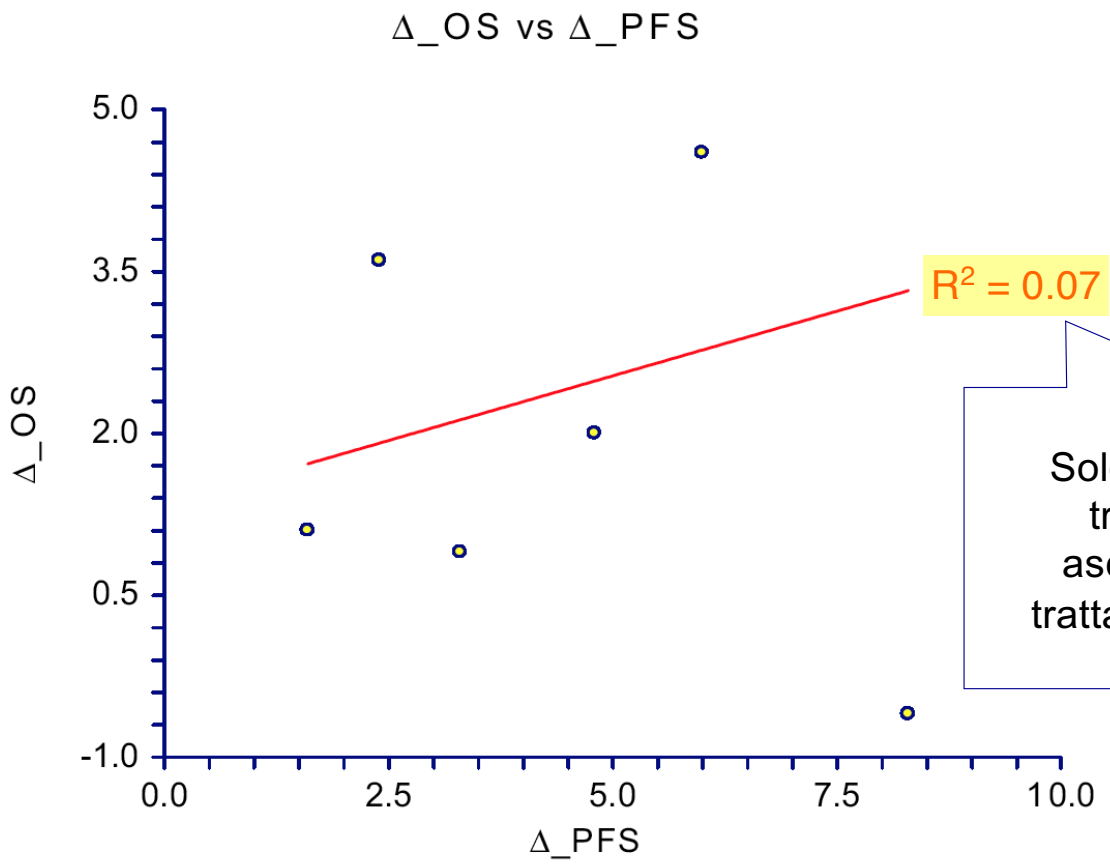
J Clin Oncol 35:3097-3104. © 2017 by American Society of Clinical Oncology



Surrogate End Points in Renal Cell Carcinoma: An Analysis of First-Line Trials With Targeted Therapies

Fausto Petrelli, Sandro Barni

Clinical Genitourinary Cancer, Vol. 11, No. 4, 385-9 © 2013 Elsevier Inc.



Solo il 7% dell'effetto del trattamento su OS è ascrivibile all'effetto del trattamento stesso su PFS

Analisi per sottogruppi

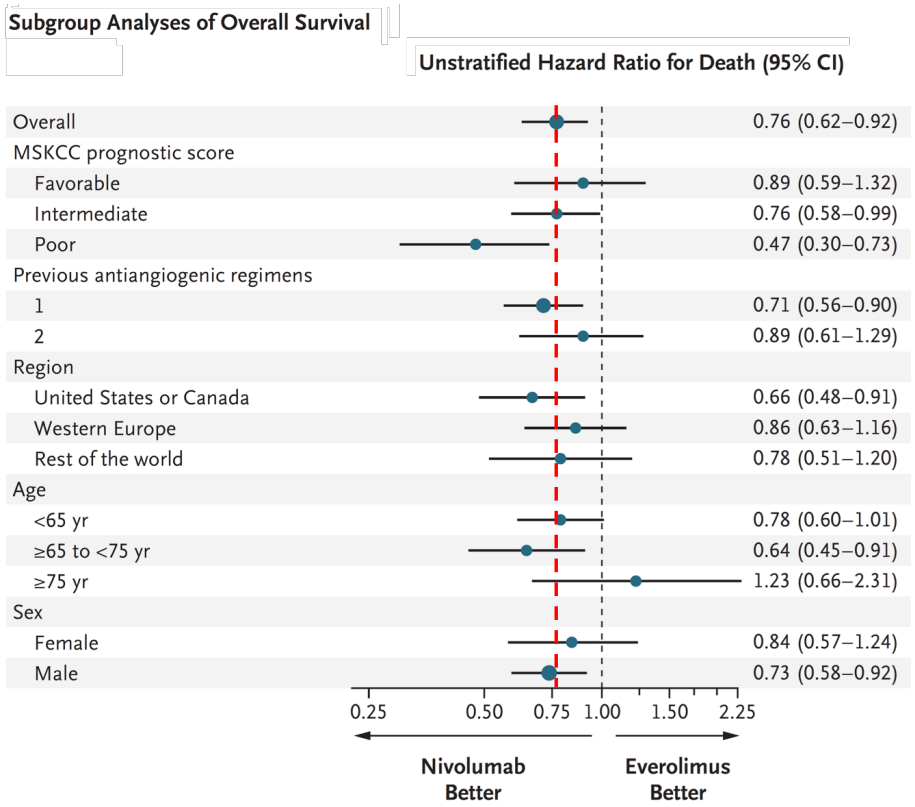
ANALISI PER SOTTOGRUPPI

- Il campione originale viene suddiviso in vari strati (**sottogruppi**) contraddistinti da caratteristiche peculiari.
- Tipologia di analisi:
  - ***Post hoc*** (analisi retrospettiva)
  - ***Pre-specified (preannunciata)*** (prevista dal protocollo di studio; criterio di stratificazione?)
  - ***Pre-planned (prepianificata)*** (prevista dal protocollo di studio con piano di analisi specifico)

Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma

R.J. Motzer, B. Escudier, D.F. McDermott, S. George, H.J. Hammers, S. Srinivas, S.S. Tykodi, J.A. Sosman, G. Procopio, E.R. Plimack, D. Castellano, T.K. Choueiri, H. Gurney, F. Donskov, P. Bono, J. Wagstaff, T.C. Gauler, T. Ueda, Y. Tomita, F.A. Schutz, C. Kollmannsberger, J. Larkin, A. Ravaud, J.S. Simon, L.-A. Xu, I.M. Waxman, and P. Sharma, for the CheckMate 025 Investigators*

N Engl J Med 2015;373:1803-13.



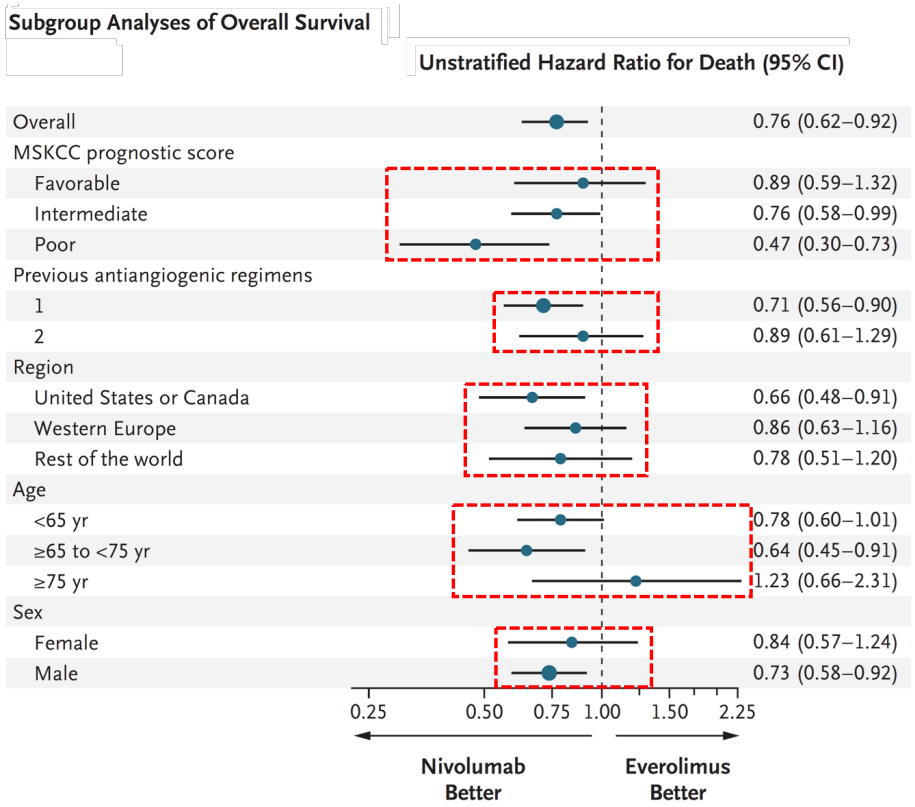
Il riferimento non è (più) la linea di non-effetto, bensì la linea tracciata in corrispondenza dell'effetto osservato sull'intero campione…

… un trattamento non dovrebbe essere limitato / escluso in una specifica sottopopolazione (solo) sulla base di un LC95% al di là (al di qua) della linea di non-effetto!

Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma

R.J. Motzer, B. Escudier, D.F. McDermott, S. George, H.J. Hammers, S. Srinivas, S.S. Tykodi, J.A. Sosman, G. Procopio, E.R. Plimack, D. Castellano, T.K. Choueiri, H. Gurney, F. Donskov, P. Bono, J. Wagstaff, T.C. Gauler, T. Ueda, Y. Tomita, F.A. Schutz, C. Kollmannsberger, J. Larkin, A. Ravaud, J.S. Simon, L.-A. Xu, I.M. Waxman, and P. Sharma, for the CheckMate 025 Investigators*

N Engl J Med 2015;373:1803-13.



The heterogeneity of the treatment effect within each subgroup shown in Figure 2A was tested with the use of an interaction test in a Cox proportional-hazards model with treatment, subgroup, and treatment-by-subgroup interaction as covariates. None of the interaction terms were significant at the 0.05 level.

Potenziali problemi…

- **High Risk of Bias** (sottogruppo non oggetto di stratificazione)
- **Multiplicity** (*type I error inflation* dovuta a confronti ripetuti)
- **Imprecision** (LC95% compatibili con interpretazioni cliniche di segno opposto)

Only one thing is worse than doing subgroup analyses…

believing the results!

R. Peto

ANALISI PER SOTTOGRUPPI

- Il campione originale viene suddiviso in vari strati (**sottogruppi**) contraddistinti da caratteristiche peculiari.
- Tipologia di analisi:
  - ***Post hoc*** (analisi retrospettiva)
  - ***Pre-specified (preannunciata)*** (prevista dal protocollo di studio; criterio di stratificazione?)
  - ***Pre-planned (prepianificata)*** (prevista dal protocollo di studio con piano di analisi specifico)