April 5, 2022

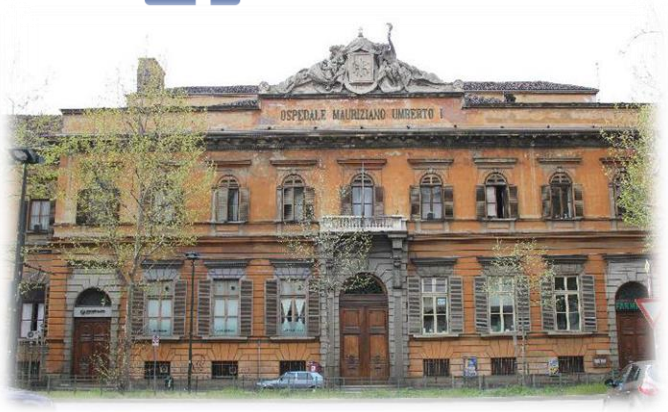
How to understand subgroup analysis in clinical studies

Massimo Di Maio

SCDU Medical Oncology
AO Ordine Mauriziano, Torino, Italy
Department of Oncology,
University of Turin
massimo.dimaio@unito.it

@MassimoDiMaio75

dimaiomax



(Above) The Prophecy, the Famous Rea Packer Cartoon of June 1, 1935. (Right) The Fulfillment, Ruth, Okla., May, 1935

(Right) Presenting an Army Pole Cup. (Lower Right) At the Veazie School at Bethesda

IN DEFENSE OF CURIOSITY

By MRS. FRANKLIN D. ROOSEVELT

A SHORT time ago a cartoon appeared depicting two miners looking up in surprise and saying with undisguised horror, "Here comes Mrs. Roosevelt!"

In strange and subtle ways, it was indicated to me that I should feel somewhat ashamed of that cartoon, and there certainly was something the matter with a woman who wanted to see so much and to know so much.

Somehow or other, most of the people who spoke to me, or wrote to me about it, seemed to feel that it was unbecoming in a woman to have a variety of interests. Perhaps that arose from the old inherent theory that woman's interests must lie only in her home. This is a kind of blindness which seems to make people feel that interest in the home stops within the four walls of the house in which you live. Few seem capable of realizing that the real reason that home is important is that it is so closely tied, by a million strings, to the rest of the world. That is what makes it an important factor in the life of only upon the buying power of people like herself but upon the buying power of the great mass of agricultural people throughout the country. The farm housewife must realize, too, that her interests are tied up with those of the wage earner and his employer throughout the nation, for her husband's products can only find a ready market when the city dweller is prosperous.

There is ever present, of course, the economic question of how to keep balanced the cost of living and the wages the man receives. The theory of low wages and low living costs has been held by many economists to be sound, for they contend what money one has will provide as much as high wages do in countries where living costs are also high.

We have gone, as a rule, on the theory, in this country, particularly in eras of prosperity, that high wages and high costs make for a higher standard of living, and that we really obtain more for our money, even though our prices are higher.

This question is argued back and forth, and the quilting pattern or recipe in the neighborhood. Isn't that better than waiting days for a letter which may never come?

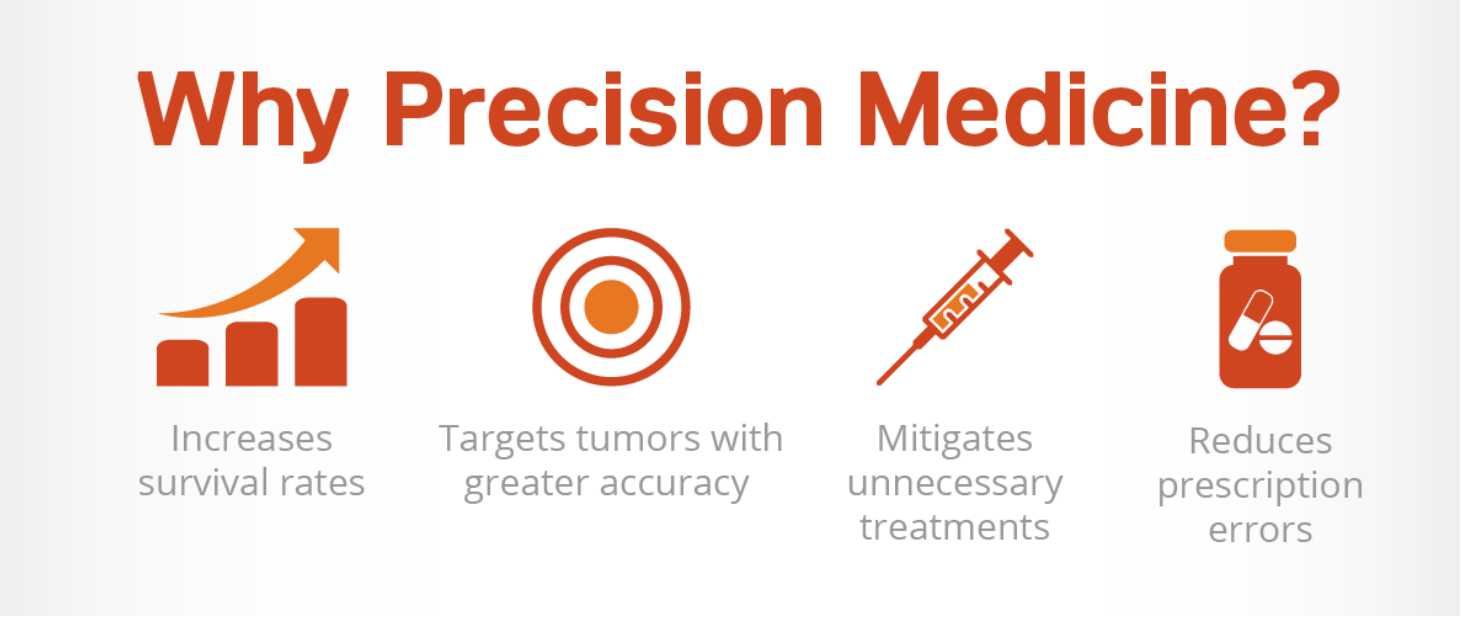
To the city or suburban dweller, the price of a subway ride is of great importance, for if it costs ten cents a day to come and go from work, he may have enough left at the end of the week to take his wife to a movie, but twenty cents a day may mean that he has nothing left for entertainment. The city dweller could also do much for the price of milk, if he realized the dairy farmer's plight and how important the consumption of milk, and its price, is to general prosperity.

This correlation of interests is something that every woman would understand if she had the curiosity to find out the reason for certain conditions instead of merely accepting them, usually with rather bad grace.

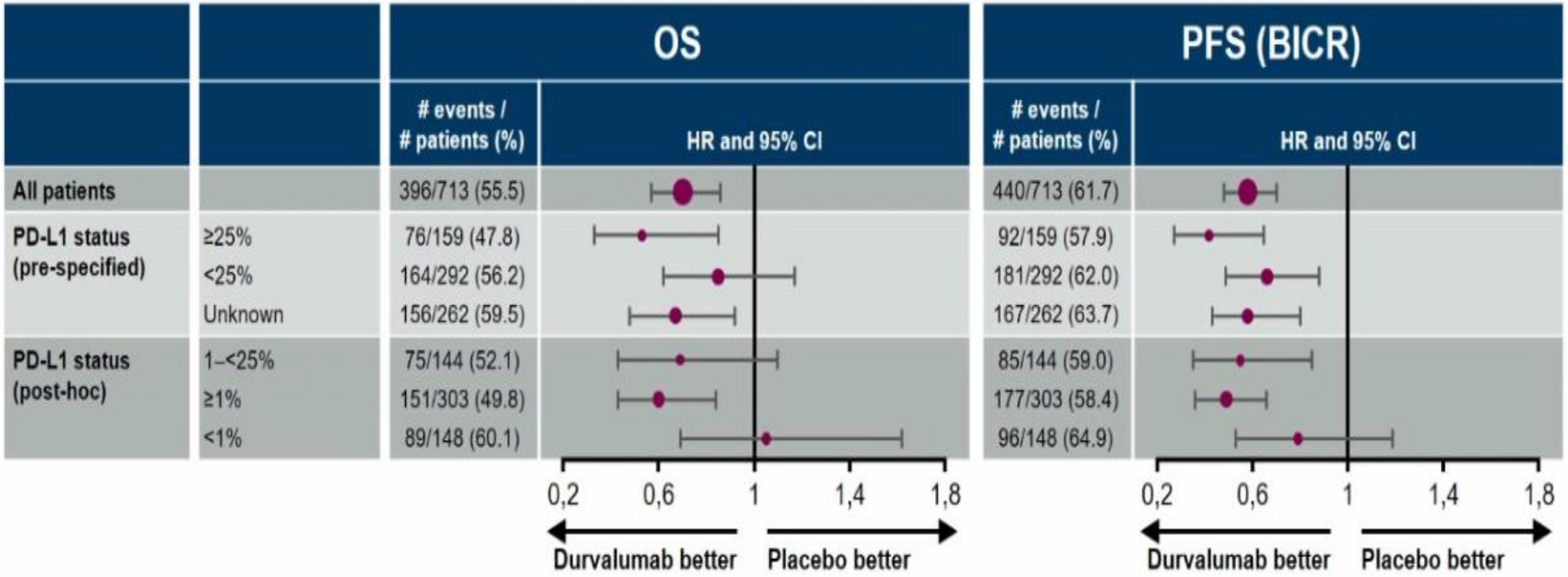
Curiosity is the Mother of Opinion

Eleanor Roosevelt, The Saturday Evening Post, August 24, 1935

…even more than before, in the era of personalized medicine and precision oncology, **subgroup analysis** seems a valuable tool for optimizing treatment choices.



Subgroup analysis may impact regulatory decisions: Durvalumab in locally advanced NSCLC

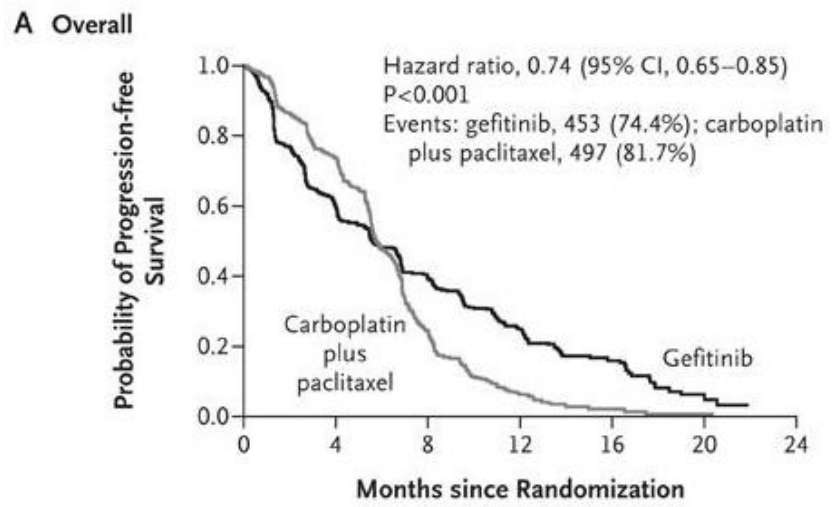


Faivre-Finn C, ESMO 2020
Annals of Oncology (2020) 31 (suppl_4): S1142-S1215

Subgroup analyses: why?

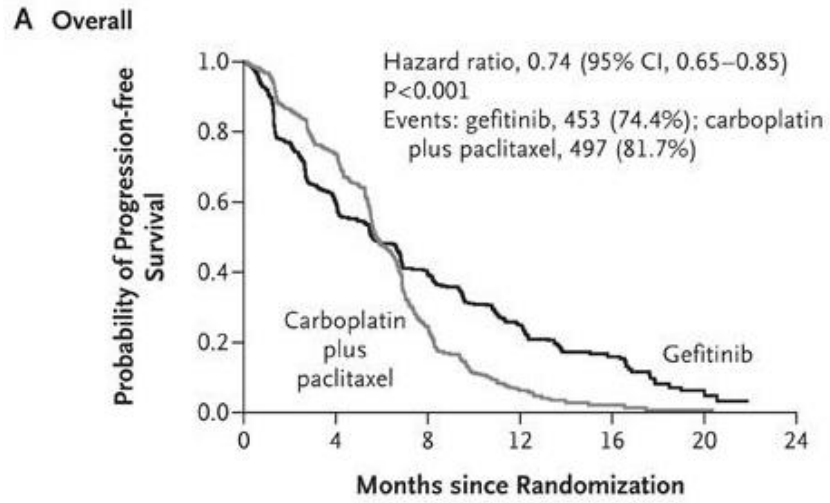
- Within a study with overall positive conclusions, subgroup analyses might help to better identify patients who benefit more, patients who benefit less or patients who don’t benefit at all.

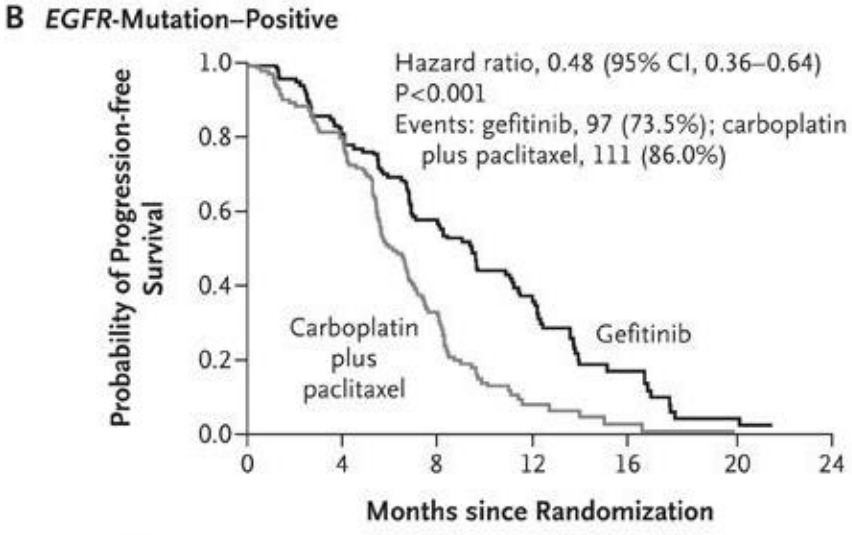
The famous example of the IPASS trial: qualitative interaction!

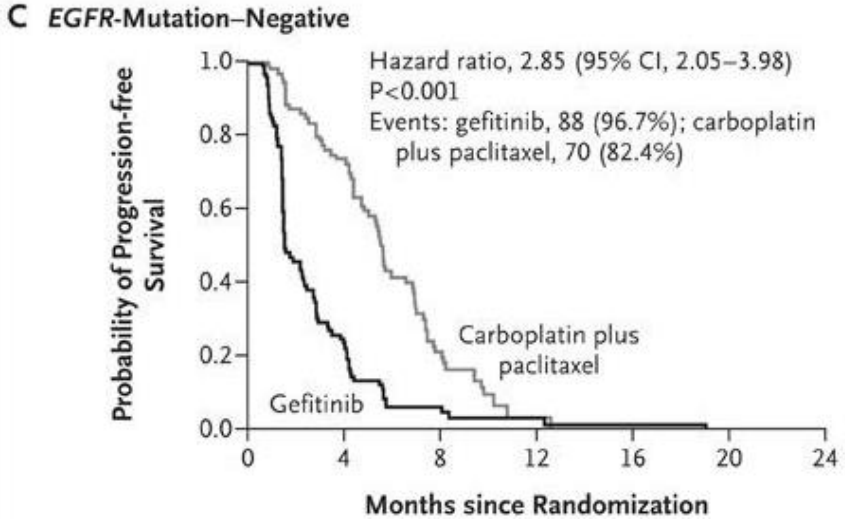


Mok TS, et al. N Engl J Med. 2009 Sep 3;361(10):947-57.

The famous example of the IPASS trial: qualitative interaction!







Mok TS, et al. N Engl J Med. 2009 Sep 3;361(10):947-57.

Subgroup analyses: why?

- Within a study with overall positive conclusions, subgroup analyses might help to better identify patients who benefit more, patients who benefit less or patients who don’t benefit at all.

- Within a study with overall negative conclusions, subgroup analyses might help to avoid **«throwing the baby out with the bath water»**, by identifying certain groups of patients in whom the experimental treatment appears to work.

…but please remember!

"Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise."

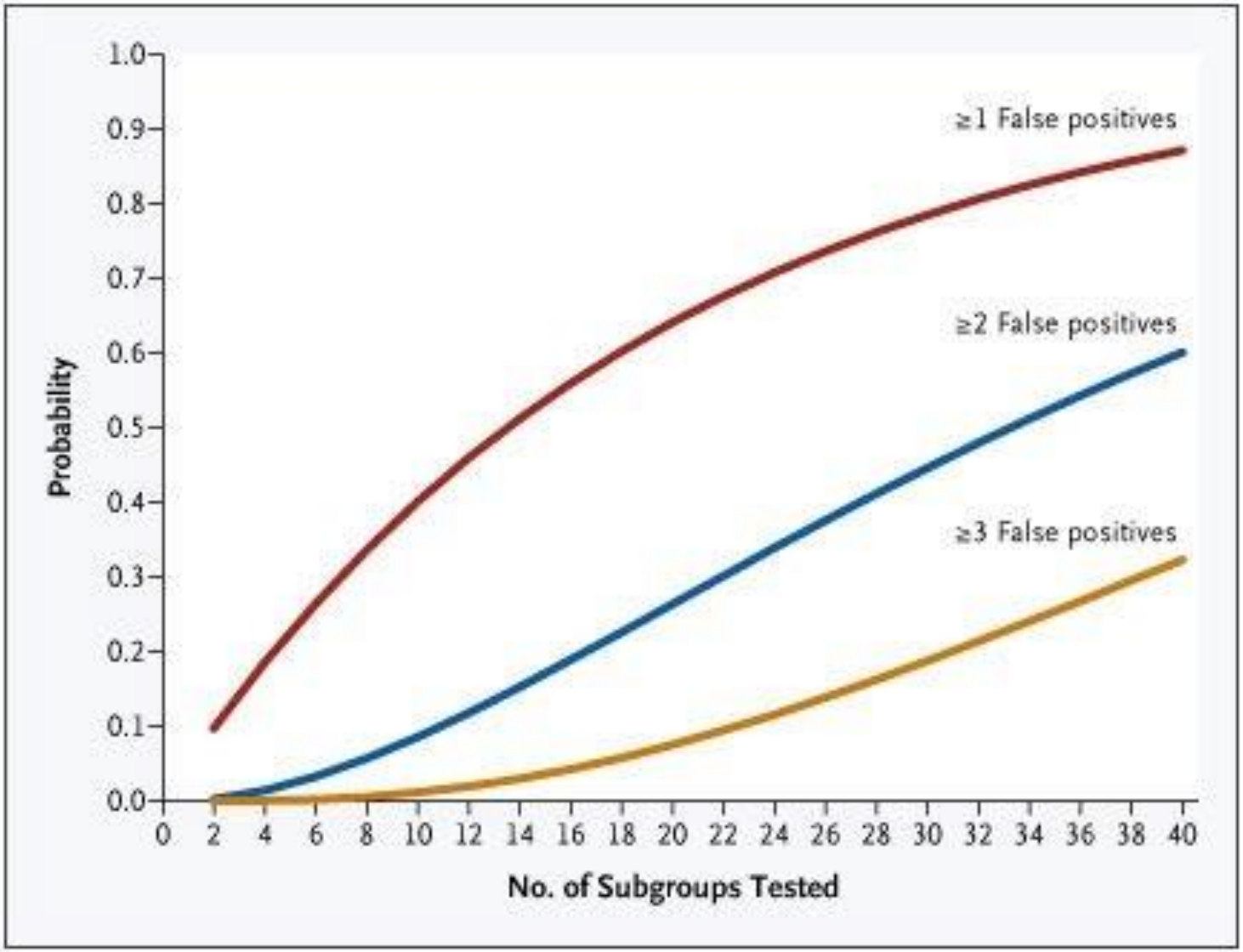
J W Tukey, 1962

Tukey JW. The future of data analysis. Ann Math Stat 1962; 33: 13–14.

Curiosity, Eugene de Blaas (1892)

In defense of curiosity…

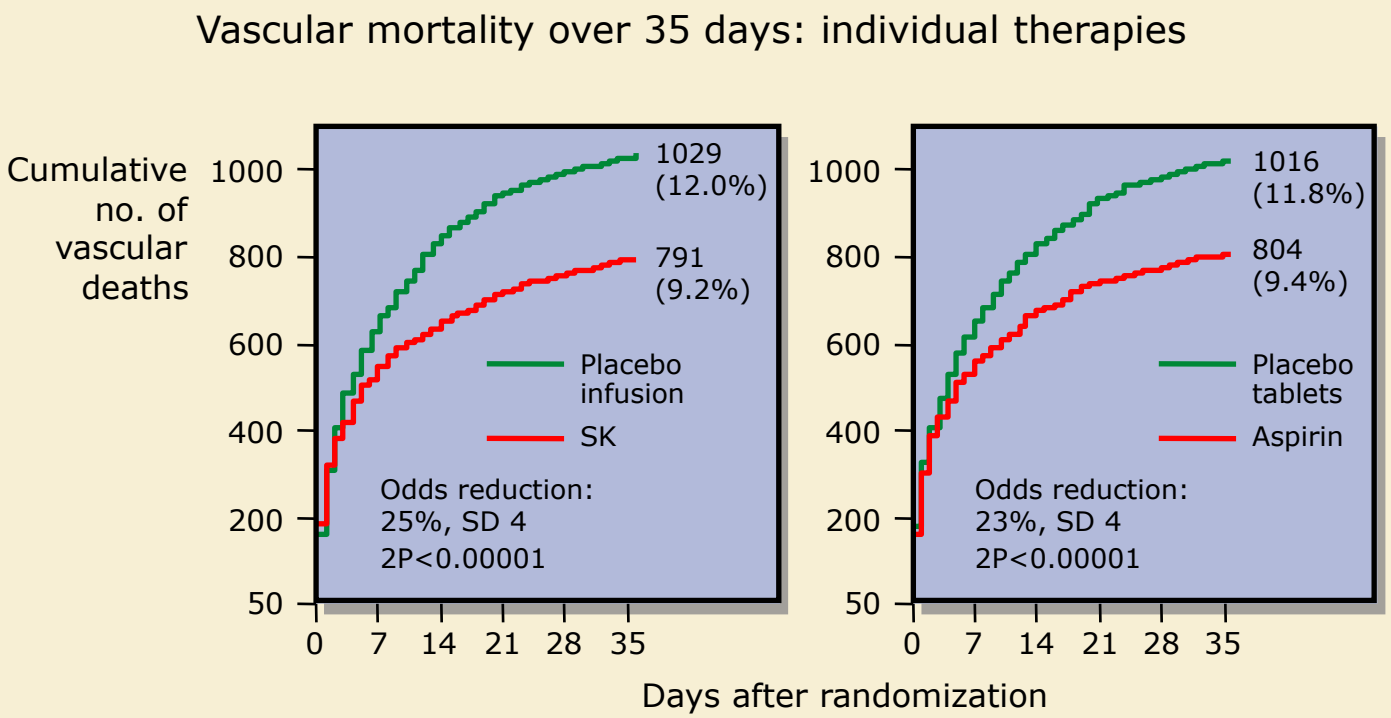
…but remember that curiosity can be dangerous!



Lagakos SW. The challenge of subgroup analyses--reporting without distorting. N Engl J Med. 2006 Apr 20;354(16):1667-9.

Let's make an example outside oncology!

ISIS-2: Second International Study of Infarct Survival



The ISIS-2 collaborative group. *Lancet* 1988; ii: 349–60.

ISIS-2 trial: Aspirin vs Placebo
Mortality 1 month after myocardial infarction




	N. of deaths A vs P	P
All cases	804 vs 1016	<0.0001








ISIS-2 trial: Aspirin vs Placebo
Mortality 1 month after myocardial infarction



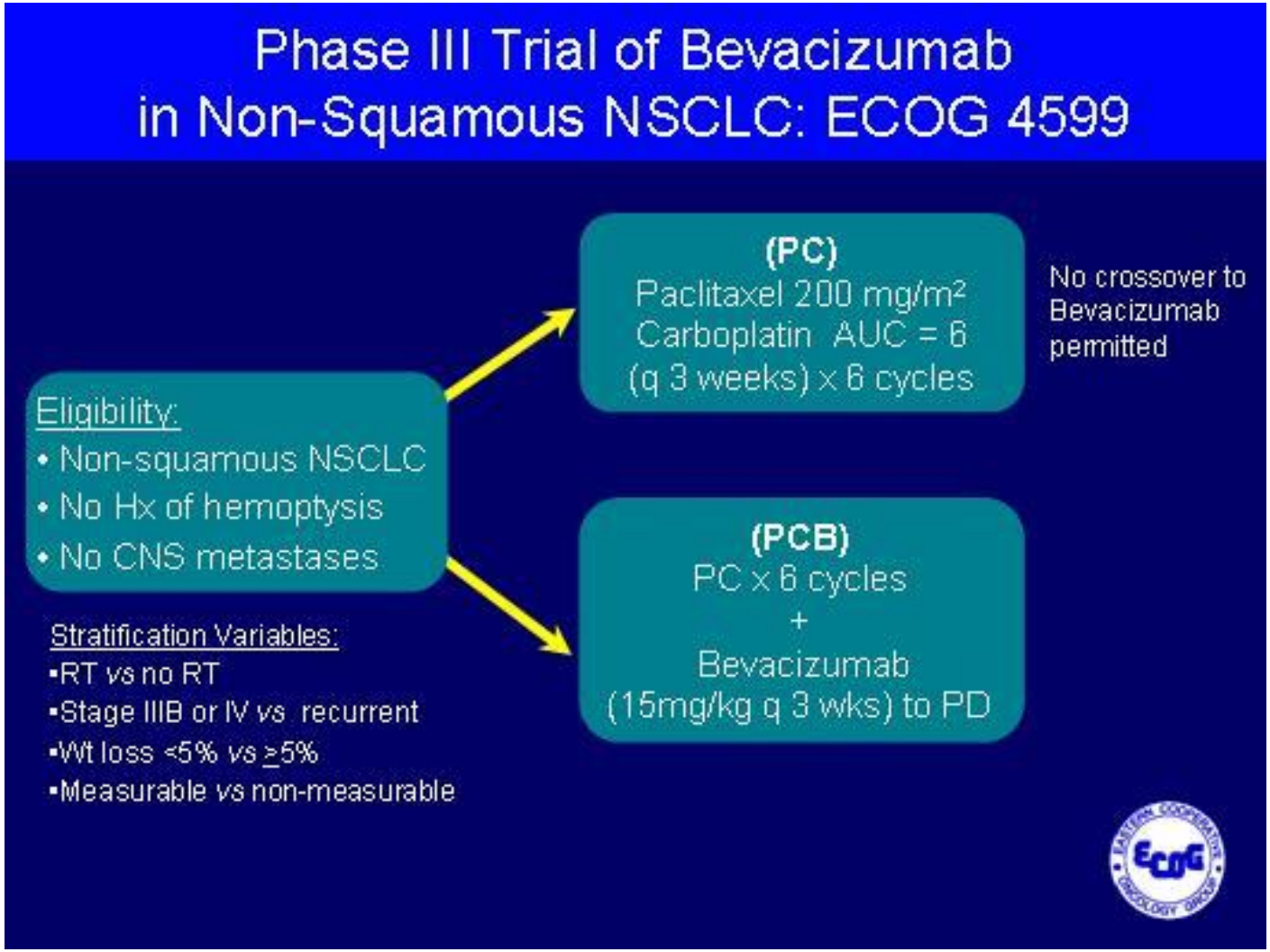

Zodiac sign	N. of deaths A vs P	P
All cases	804 vs 1016	<0.0001
Other signs	654 vs 869	<0.0001
Libra or Gemini	150 vs 147	0.5 (ns)





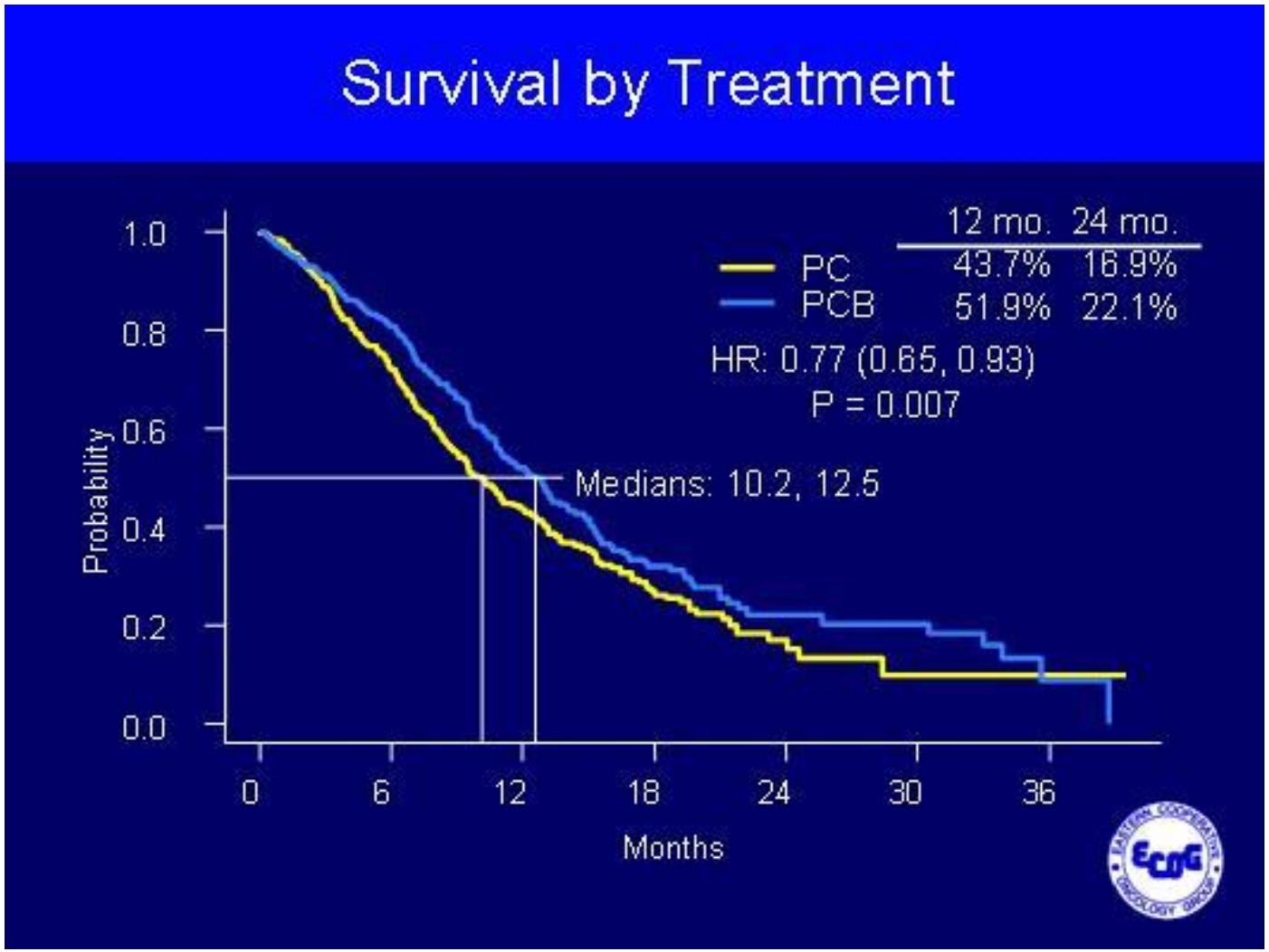



Should I suspect a risk of false negative result in a subgroup?



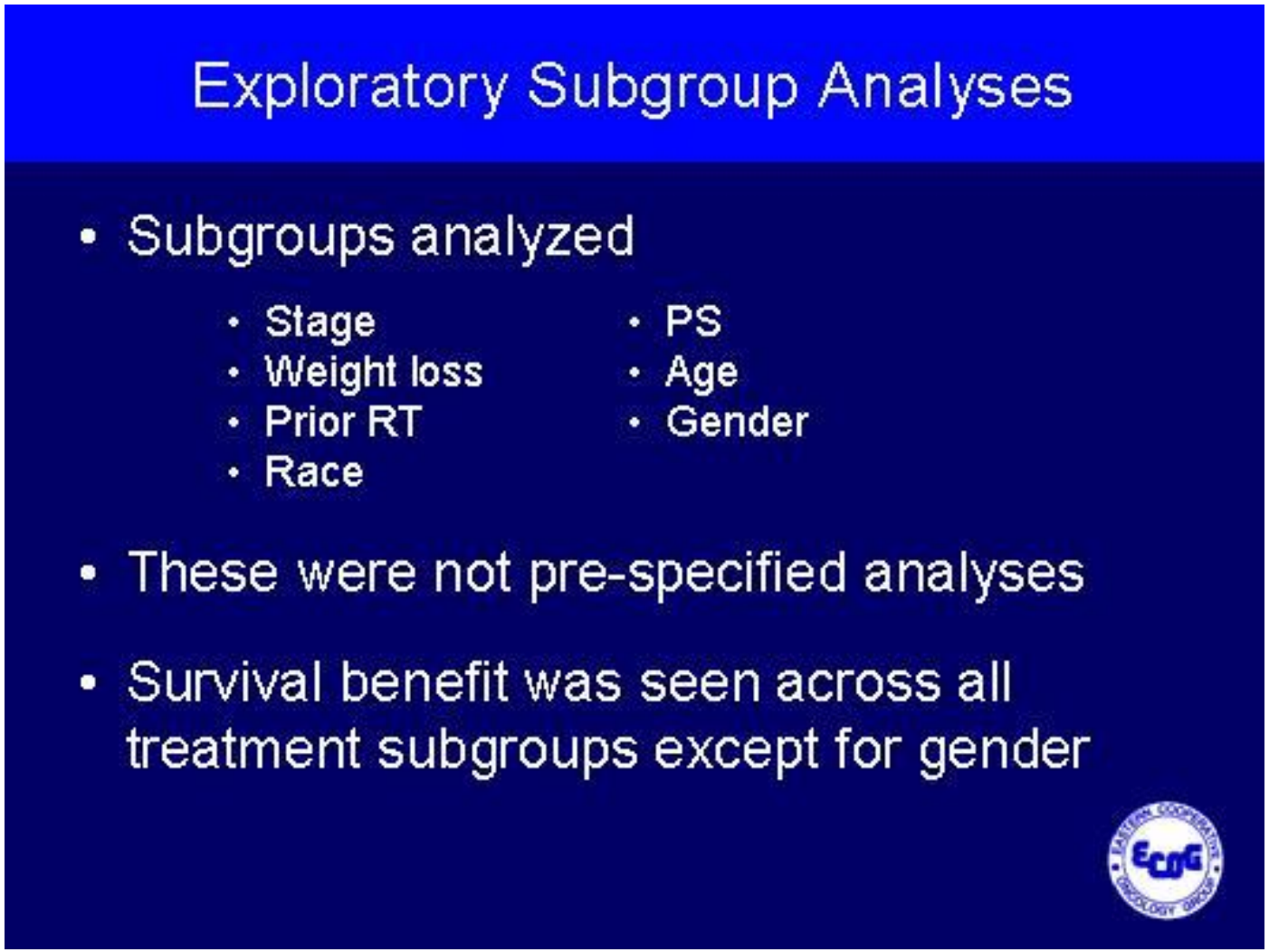
Sandler AB et al., ASCO 2005, abstract 4

Should I suspect a risk of false negative result in a subgroup?



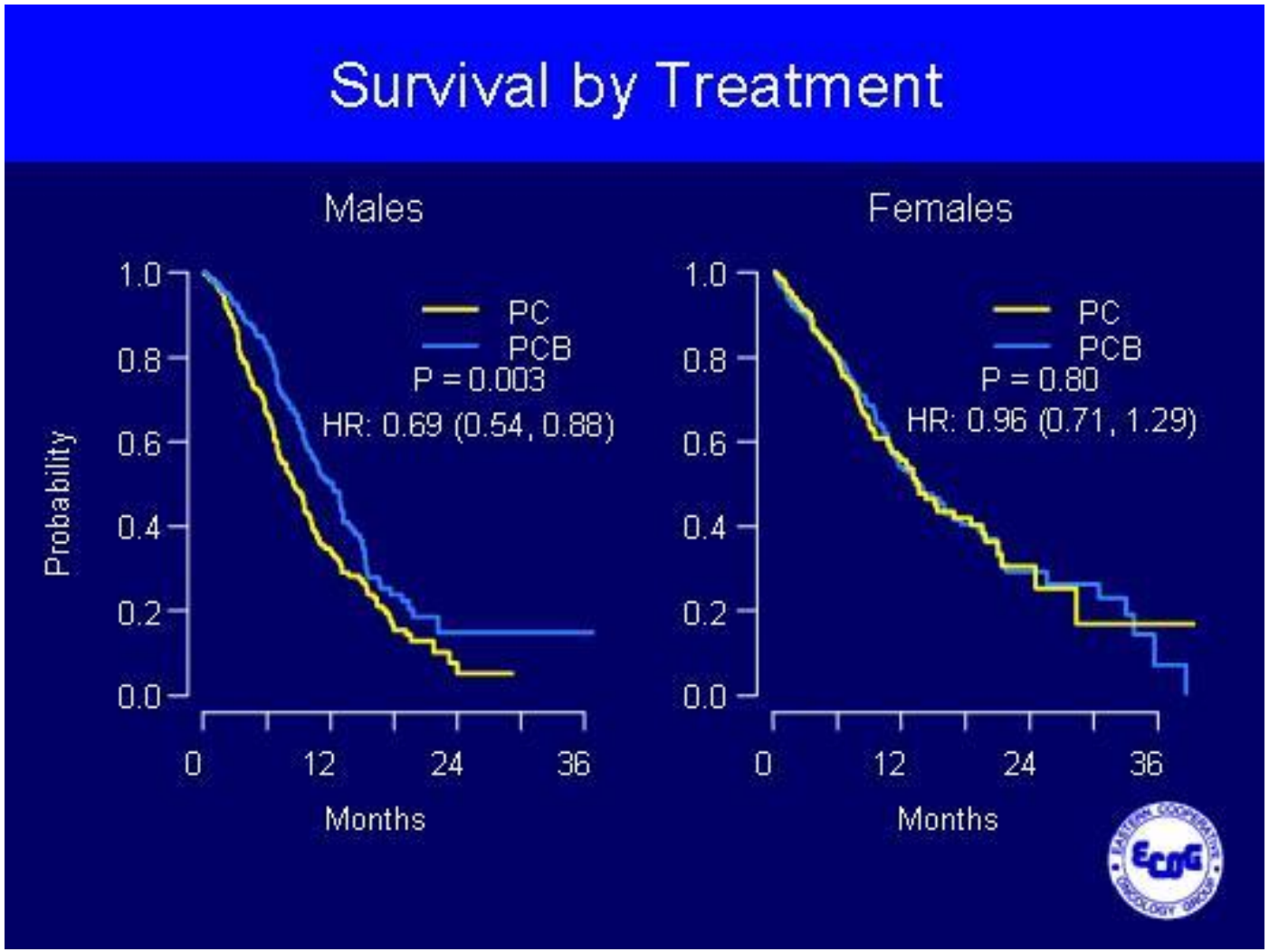
Sandler AB et al., ASCO 2005, abstract 4

Should I suspect a risk of false negative result in a subgroup?



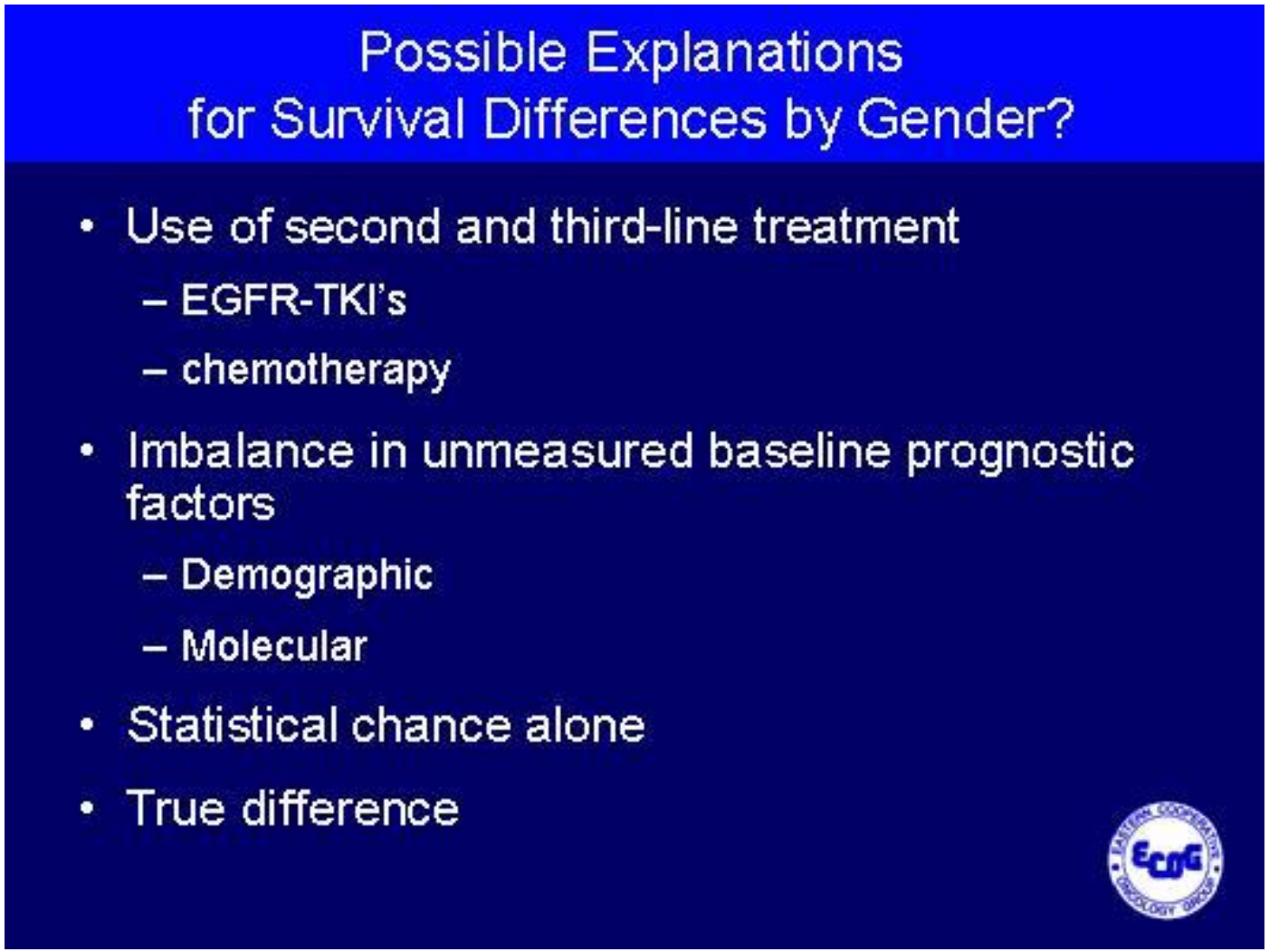
Sandler AB et al., ASCO 2005, abstract 4

Should I suspect a risk of false negative result in a subgroup?



Sandler AB et al., ASCO 2005, abstract 4

Should I suspect a risk of false negative result in a subgroup?



Sandler AB et al., ASCO 2005, abstract 4

The risk of «belief bias»…

…and the risk of HARKing

Personality and Social Psychology Review
1998, Vol. 2, No. 3, 196–217

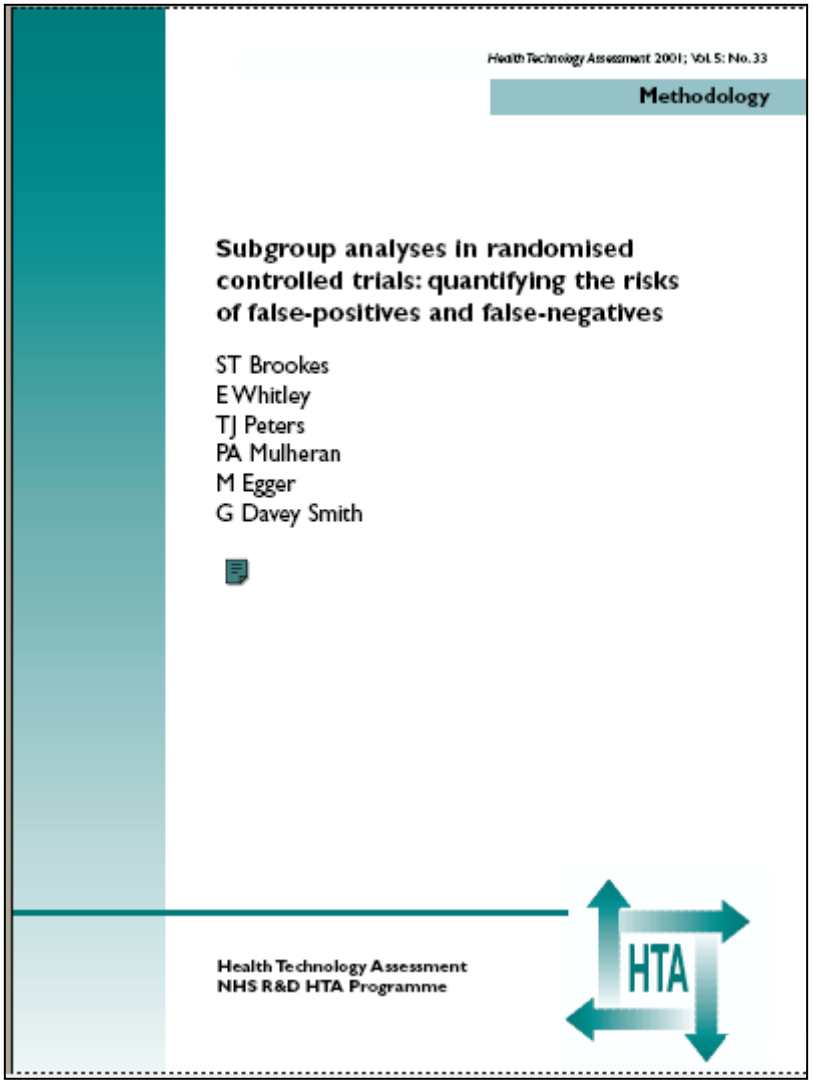
HARKing: Hypothesizing After the Results are Known

Norbert L. Kerr
Department of Psychology
Michigan State University

This article considers a practice in scientific communication termed HARKing (Hypothesizing After the Results are Known). HARKing is defined as presenting a post hoc hypothesis (i.e., one based on or informed by one's results) in one's research report as if it were, in fact, an a priori hypotheses. Several forms of HARKing are identified and survey data are presented that suggests that at least some forms of HARKing are widely practiced and widely seen as inappropriate. I identify several reasons why scientists might HARK. Then I discuss several reasons why scientists ought not to HARK. It is conceded that the question of whether HARKing's costs exceed its benefits is a complex one that ought to be addressed through research, open discussion, and debate. To help stimulate such discussion (and for those such as myself who suspect that HARKing's costs do exceed its benefits), I conclude the article with some suggestions for deterring HARKing.

Kerr NL. Pers Soc Psychol Rev. 1998;2(3):196-217.

An interesting lecture:



Brookes ST et al. Health Technology Assessment 2001; Vol.5: No. 33

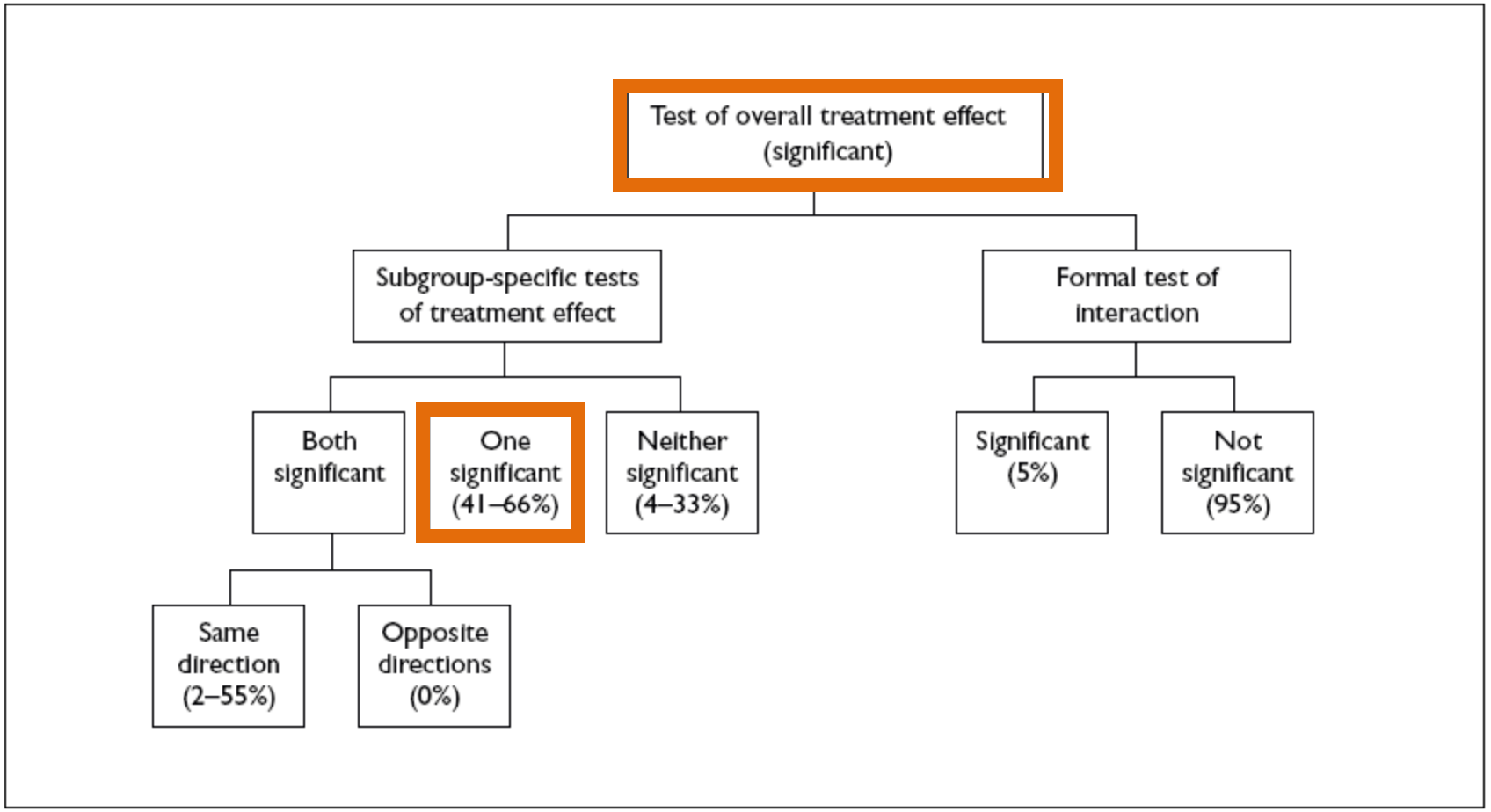
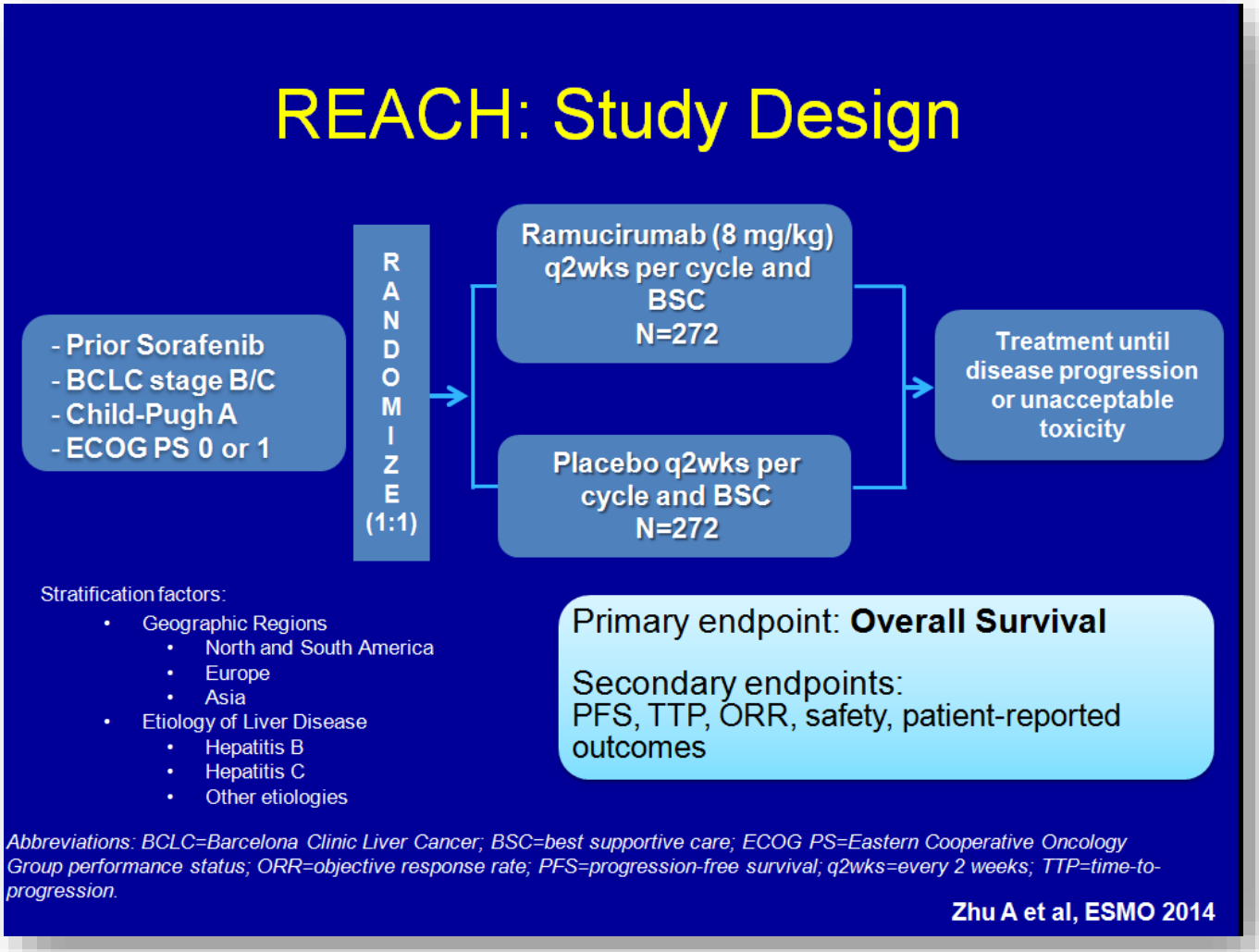


FIGURE 22 *Summary of results for the simplest case (overall test result significant). This figure combines the results from data simulated with no overall treatment effect and with a true overall treatment effect detectable at nominal powers of 50, 80, 90 and 95%*

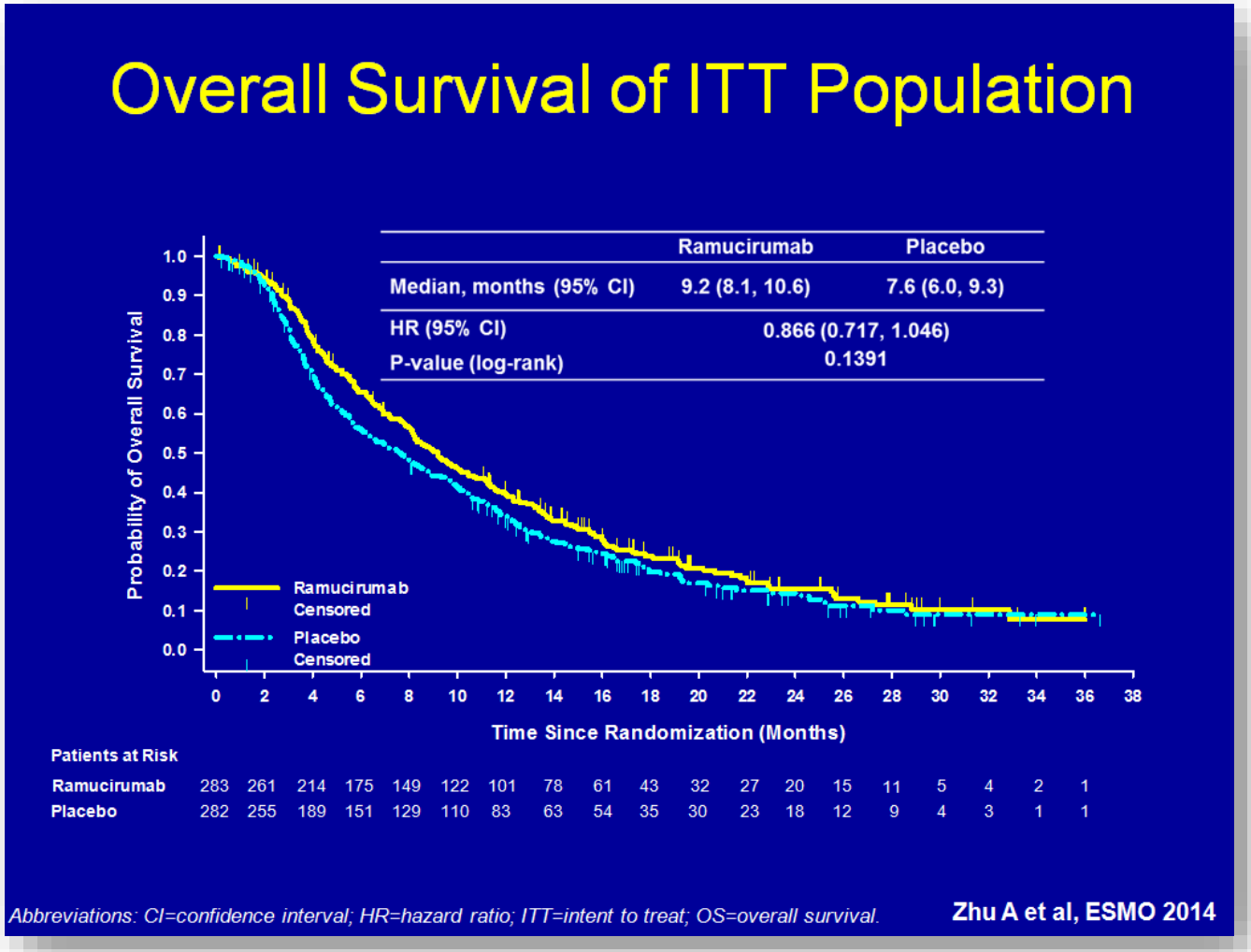
Brookes ST et al. Health Technology Assessment 2001; Vol.5: No. 33

Should I suspect a false positive result in a subgroup?



Zhu et al, ESMO 2014

Should I suspect a false positive result in a subgroup?

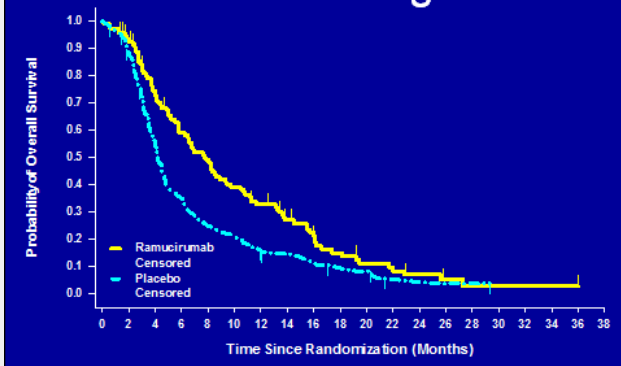


Zhu et al, ESMO 2014

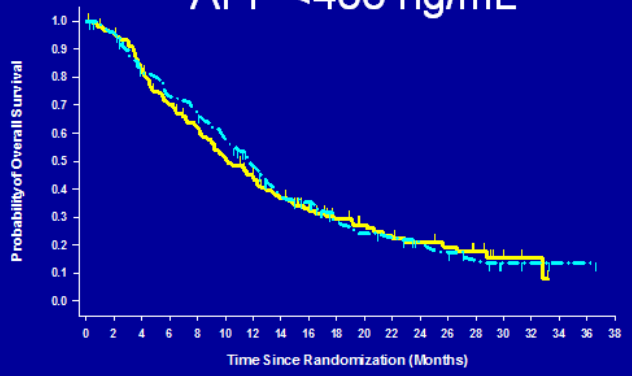
Should I suspect a false positive result in a subgroup?

Overall Survival in Patients With Baseline Alpha-fetoprotein ≥ or <400 ng/mL

AFP ≥400 ng/mL



AFP <400 ng/mL



	Ramucirumab (N=119)	Placebo (N=131)
Median, months	7.8	4.2
(95% CI)	(5.8, 9.3)	(3.7, 4.8)
HR (95% CI)	0.674 (0.508, 0.895)	
P-value (log-rank)	0.0059	

	Ramucirumab (N=160)	Placebo (N=150)
Median, months	10.1	11.8
(95% CI)	(8.7, 12.3)	(9.9, 13.1)
HR (95% CI)	1.093 (0.836, 1.428)	
P-value (log-rank)	0.5059	

Zhu A et al, ESMO 2014

Zhu et al, ESMO 2014

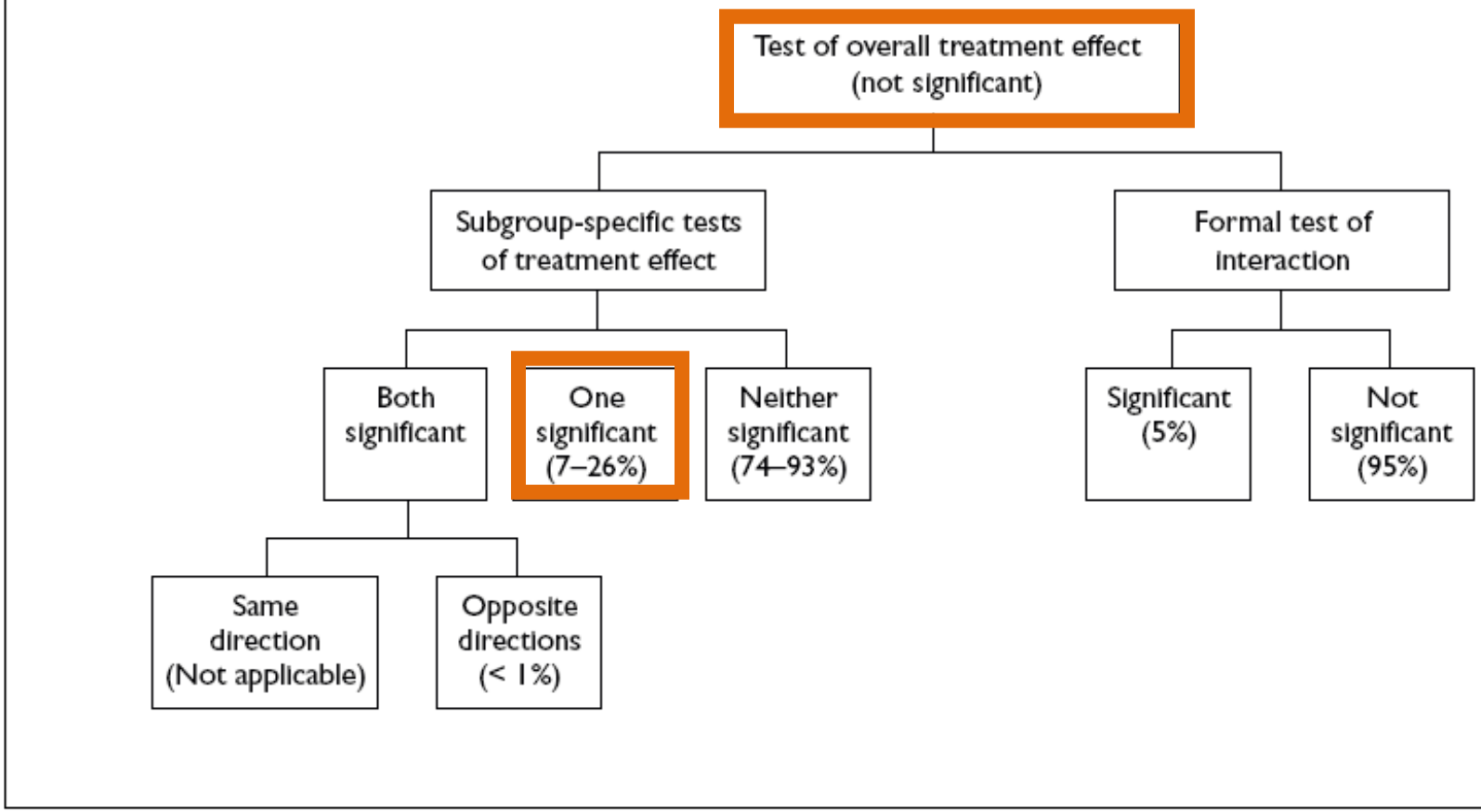
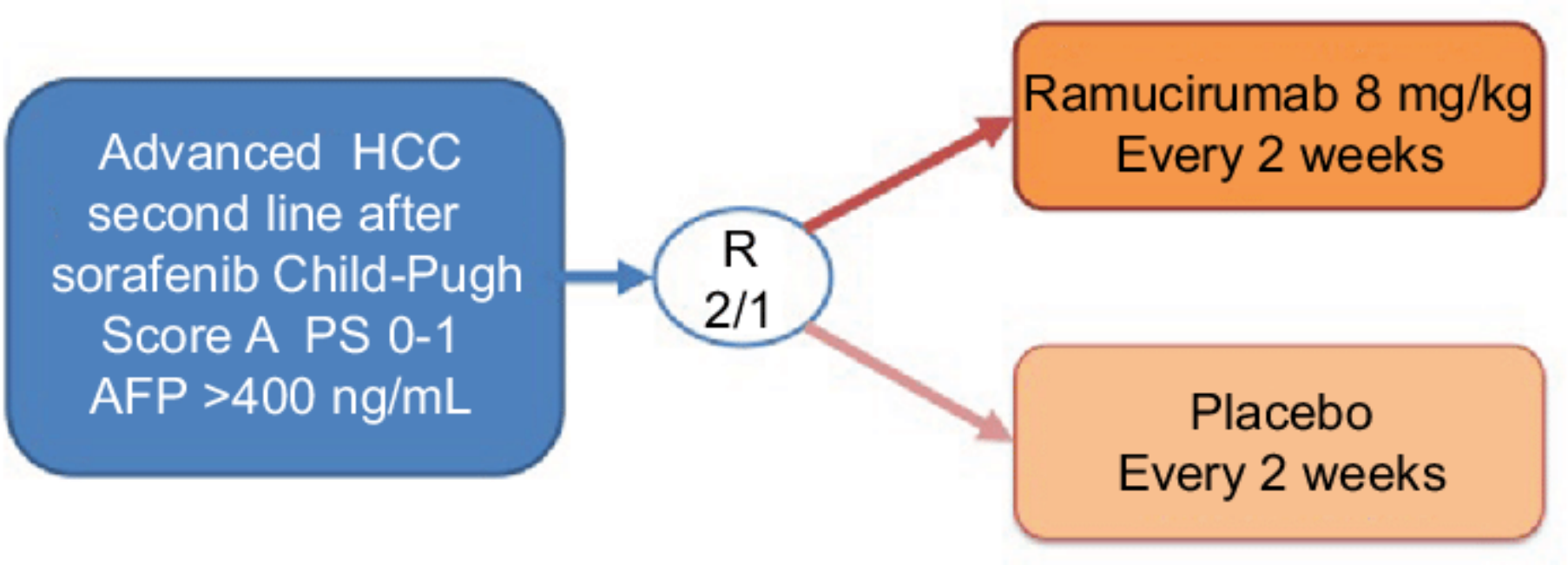


FIGURE 21 *Summary of results for the simplest case (overall test result not significant). This figure combines the results from data simulated with no overall treatment effect and with a true overall treatment effect detectable at nominal powers of 50, 80, 90 and 95%*

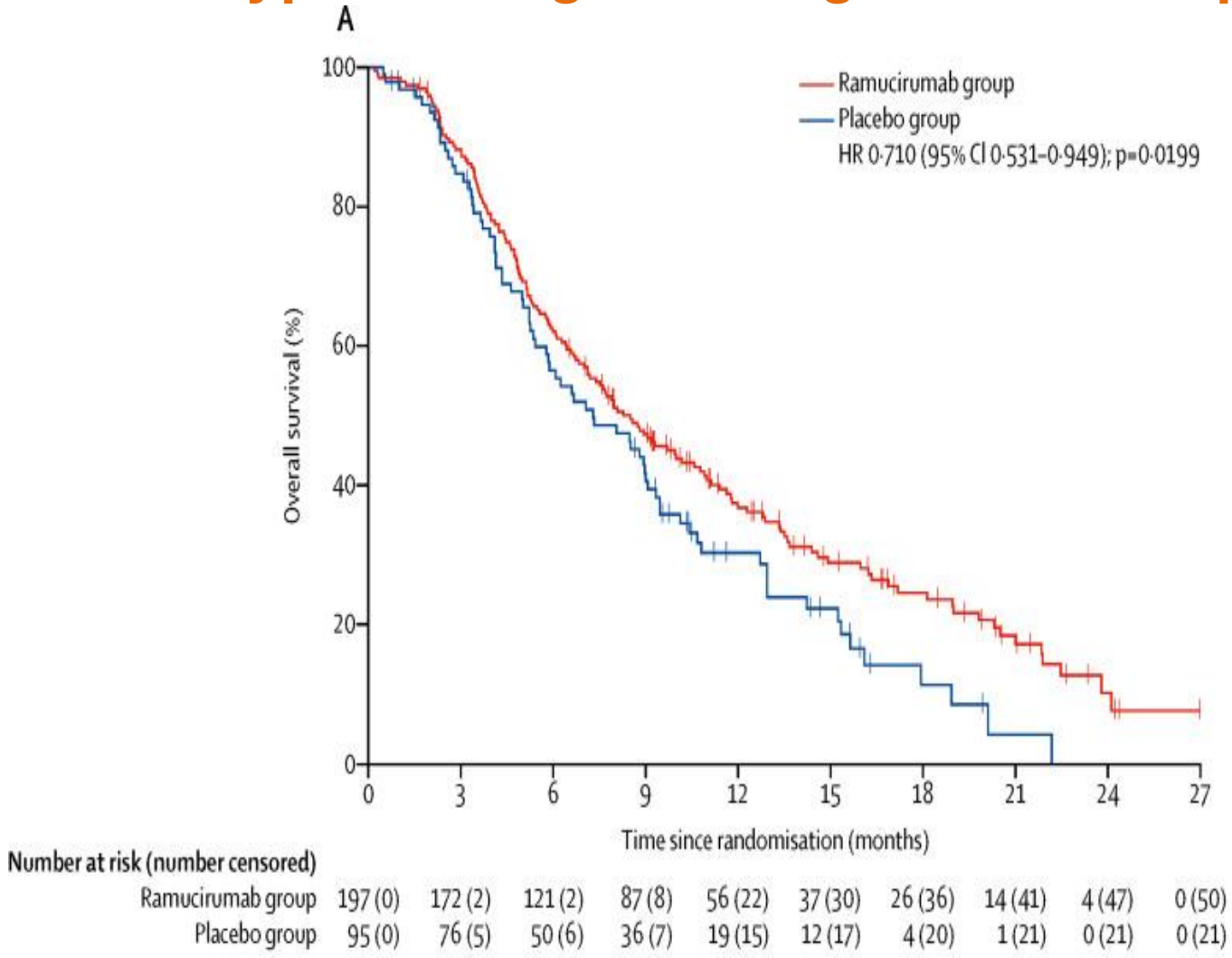
Brookes ST et al. Health Technology Assessment 2001; Vol.5: No. 33

Subgroup analysis
can be hypothesis-generating for a subsequent trial!



Zhu AX, Lancet Oncol. 2019 Feb;20(2):282-296.

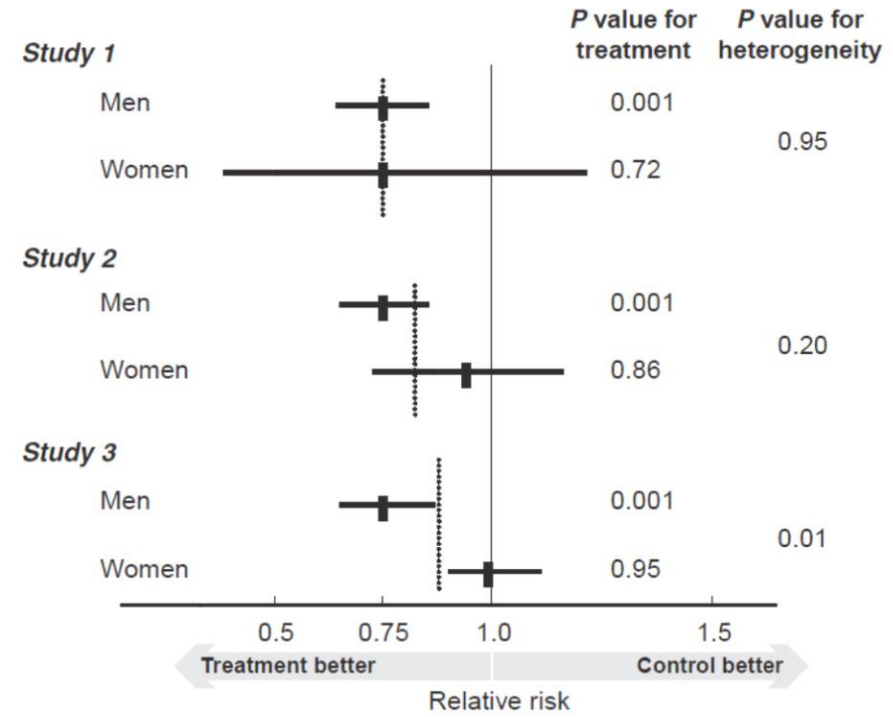
Subgroup analysis
can be hypothesis-generating for a subsequent trial!



Zhu AX, Lancet Oncol. 2019 Feb;20(2):282-296.

How to correctly interpret subgroup analyses?

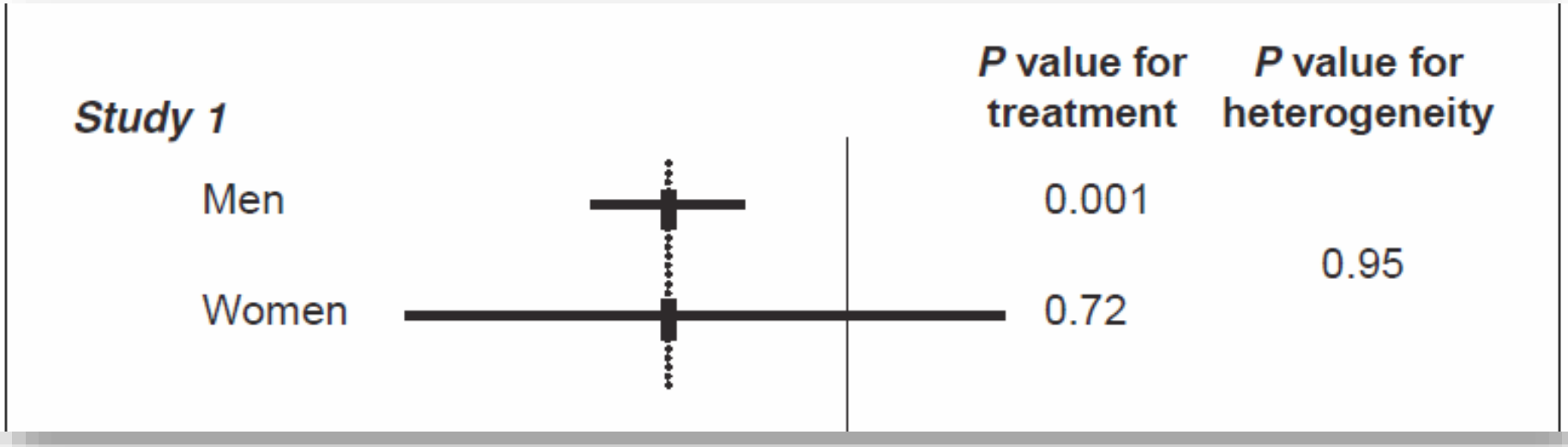
1: Treatment effects in subgroups of men and women in three hypothetical trials



Overall relative risk: 0.75 for Study 1; 0.80 for Study 2; 0.87 for Study 3; represented by the vertical dashed line in each case.

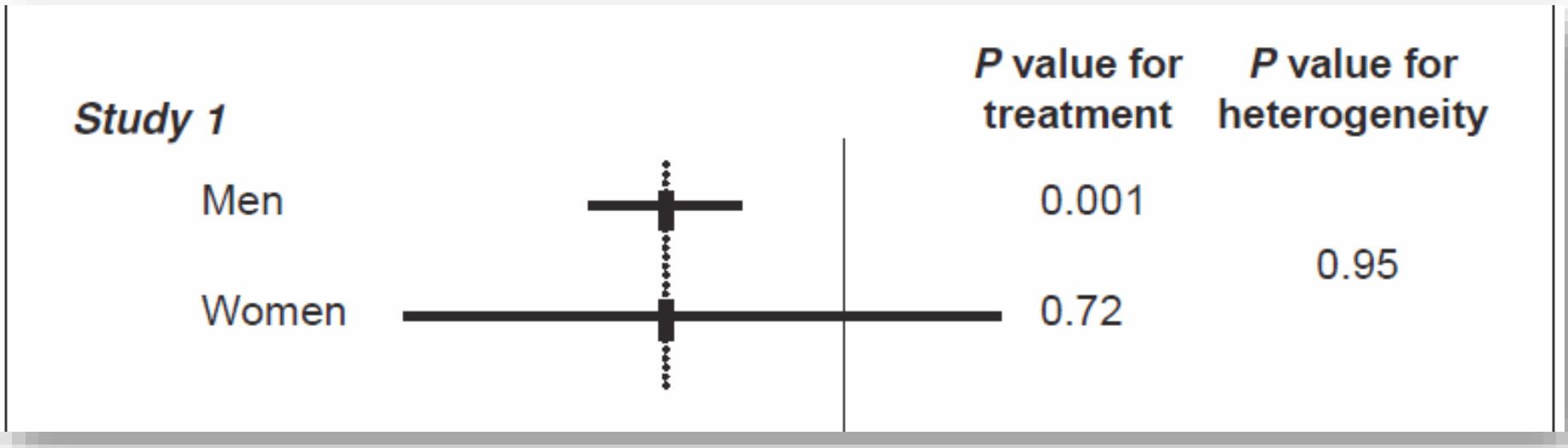
Simes RJ et al, MJA 2004

How to correctly interpret subgroup analyses?



Simes RJ et al, MJA 2004

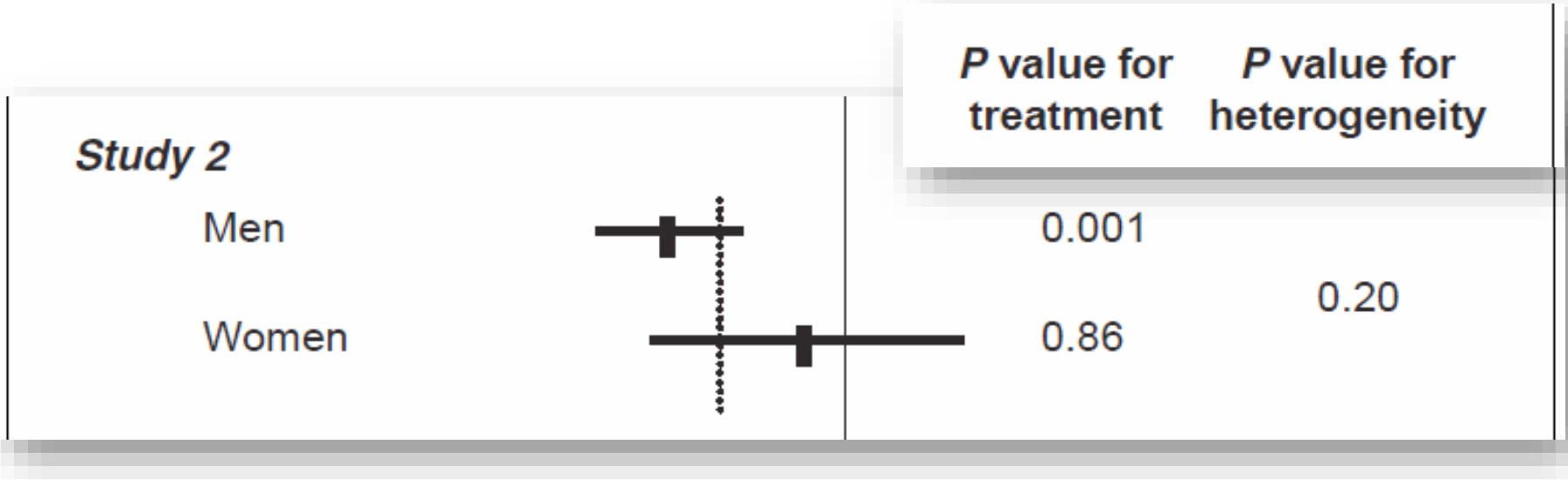
How to correctly interpret subgroup analyses?



In cases like this, please **DO NOT CLAIM** that experimental treatment is significantly effective in men but not in women!

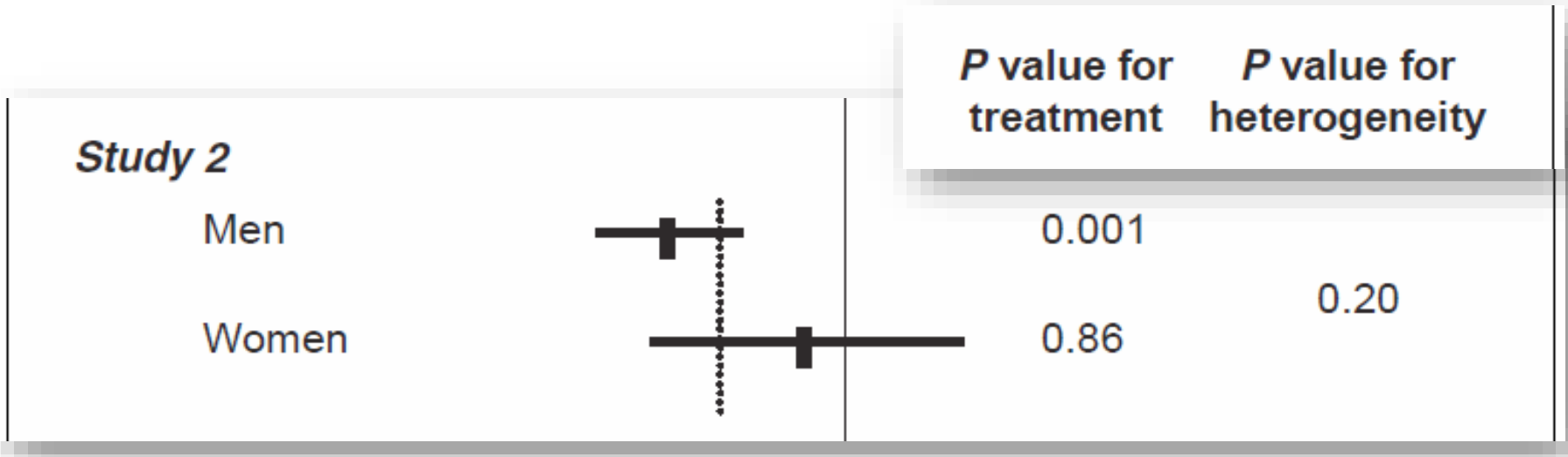
Simes RJ et al, MJA 2004

How to correctly interpret subgroup analyses?



Simes RJ et al, MJA 2004

How to correctly interpret subgroup analyses?

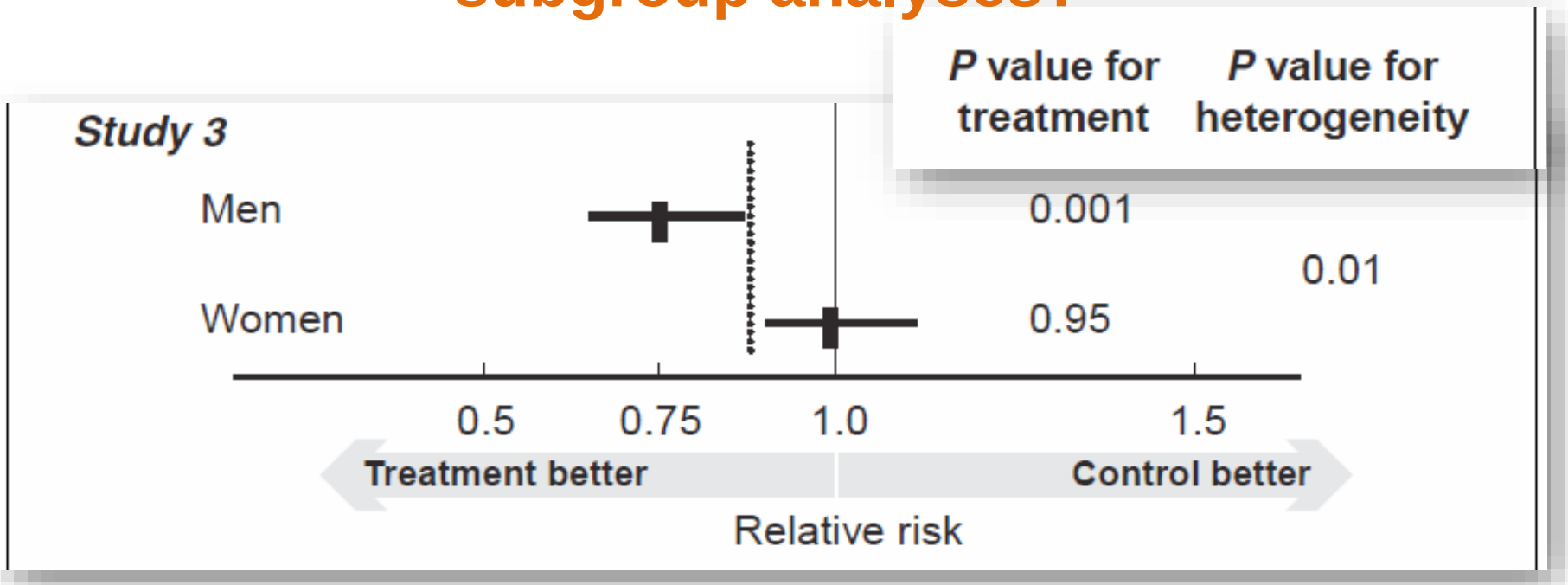


In cases like this, it is legitimate to suspect that treatment effiacy could be different…

…unfortunately, we cannot exclude that the difference we are observing is due to chance!

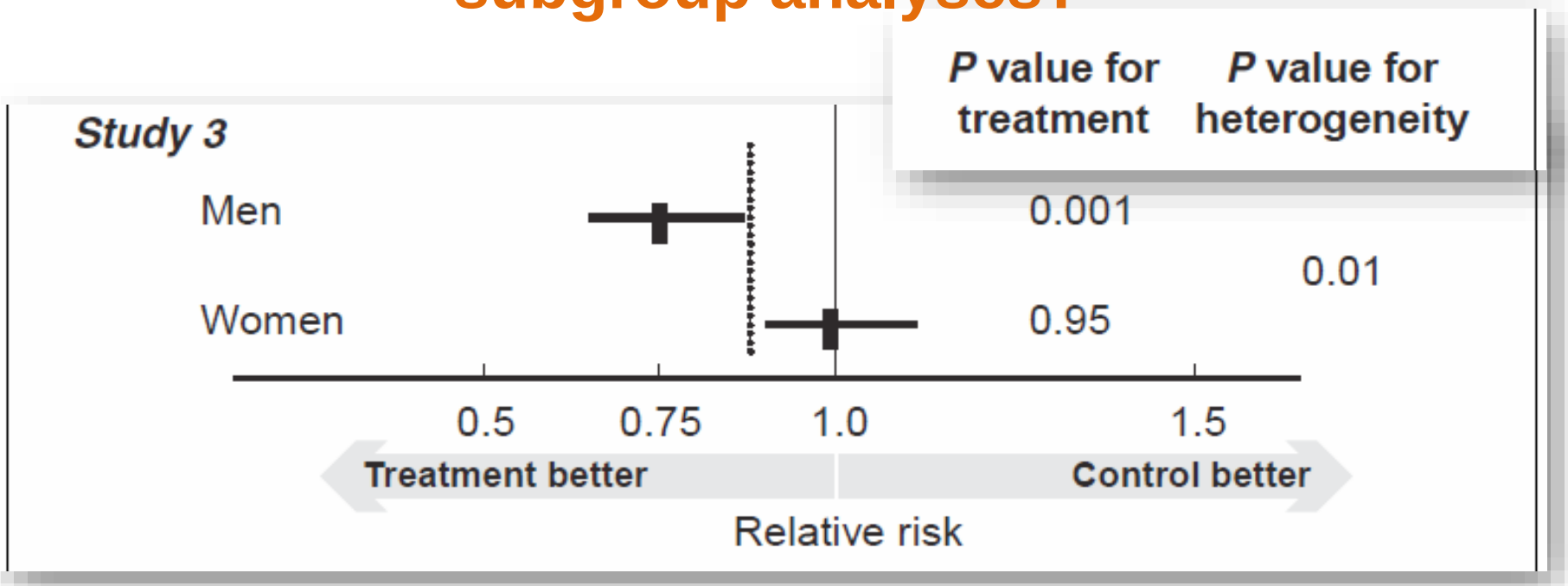
Simes RJ et al, MJA 2004

How to correctly interpret subgroup analyses?



Simes RJ et al, MJA 2004

How to correctly interpret subgroup analyses?



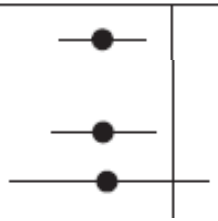
In cases like this, it is legitimate to discuss the heterogeneity of treatment effect between men and women.

Interaction test tells us that this difference is unlikely to be due to chance.

Simes RJ et al, MJA 2004

Scenario n.1

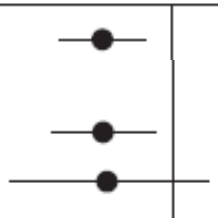
	Number of patients	Median overall survival (months)		Unstratified hazard ratio for death (95% CI)
		Nivolumab plus ipilimumab group (n=303)	Chemotherapy group (n=302)	
All randomly assigned	605	18·1	14·1	0·75 (0·62–0·91)
Sex				
Male	467	17·5	13·7	0·74 (0·60–0·92)
Female	138	21·4	18·0	0·76 (0·50–1·16)



Baas et al, Lancet 2021

Scenario n.1

	Number of patients	Median overall survival (months)		Unstratified hazard ratio for death (95% CI)
		Nivolumab plus ipilimumab group (n=303)	Chemotherapy group (n=302)	
All randomly assigned	605	18·1	14·1	0·75 (0·62–0·91)
Sex				
Male	467	17·5	13·7	0·74 (0·60–0·92)
Female	138	21·4	18·0	0·76 (0·50–1·16)

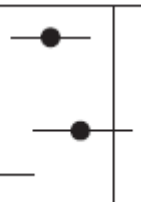


- **Interaction test (p=0.91) is NOT significant: heterogeneity of efficacy between men and women is NOT demonstrated**

Di Maio M, Tagliamento M.
Heterogeneity of treatment effects in malignant pleural mesothelioma.
Lancet. 2021 Jul 24;398(10297):301-302..

Scenario n.2

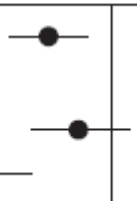
	Number of patients	Median overall survival (months)		Unstratified hazard ratio for death (95% CI)
		Nivolumab plus ipilimumab group (n=303)	Chemotherapy group (n=302)	
All randomly assigned	605	18·1	14·1	0·75 (0·62–0·91)
Tumour histology				
Epithelioid	456	18·7	16·5	0·86 (0·69–1·08)
Non-epithelioid	149	18·1	8·8	0·46 (0·31–0·68)



Baas et al, Lancet 2021

Scenario n.2

	Number of patients	Median overall survival (months)		Unstratified hazard ratio for death (95% CI)
		Nivolumab plus ipilimumab group (n=303)	Chemotherapy group (n=302)	
All randomly assigned	605	18·1	14·1	0·75 (0·62–0·91)
Tumour histology				
Epithelioid	456	18·7	16·5	0·86 (0·69–1·08)
Non-epithelioid	149	18·1	8·8	0·46 (0·31–0·68)



Interaction test is significant (p=0.007)
Heterogeneity of efficacy between epithelioid and non epithelioid tumors is demonstrated

Di Maio M, Tagliamento M.
Heterogeneity of treatment effects in malignant pleural mesothelioma.
Lancet. 2021 Jul 24;398(10297):301-302..

Subgroup analyses: take home messages

Subgroup analyses: take home messages

- **Caution!**

Subgroup analyses: take home messages

- **Caution!**
- **Hypothesis generation**

Subgroup analyses: take home messages

- **Caution!**
- **Hypothesis generation**
- **Multiplicity: risks of false positive and false negative**

Subgroup analyses: take home messages

- **Caution!**
- **Hypothesis generation**
- **Multiplicity: risks of false positive and false negative**
- **Look at consistency among studies**

Subgroup analyses: take home messages

- **Caution!**
- **Hypothesis generation**
- **Multiplicity: risks of false positive and false negative**
- **Look at consistency among studies**
- **Plausibility (but beware of belief bias!)**

Subgroup analyses: take home messages

- **Caution!**
- **Hypothesis generation**
- **Multiplicity: risks of false positive and false negative**
- **Look at consistency among studies**
- **Plausibility (but beware of belief bias!)**
- **Look at the interaction test!**

Subgroup analyses in randomized phase III trials of systemic treatments in advanced solid tumours: a systematic review of trials published between 2017 and 2020

DEPARTMENT OF ONCOLOGY
UNIVERSITY OF TURIN

Department of Oncology, University of Turin, Italy
Chiara Paratore
Clizia Zichi
Maria Lucia Reale
Annapaola Mariniello
Marco Audisio
Maristella Bungaro
Teresa Gamba
Andrea Caglio
Massimo Di Maio








ISTITUTO NAZIONALE TUMORI
IRCCS - Fondazione Pascale

Clinical Trials Unit, INT G.Pascale, Napoli, Italy
Piera Gargiulo
Raimondo Di Liello
Francesco Perrone


massimo.dimaio@unito.it
@MassimoDiMaio75
dimaiomax

Selected references

- Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. The Lancet, 2005; 365 (9454), 8–14: 176-186
- https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-subgroups-confirmatory-clinical-trials_en.pdf
- Peto R. Current misconception 3: that subgroup-specific trial mortality results often provide a good basis for individualising patient care. Br J Cancer. 2011 Mar 29; 104(7): 1057–1058.
- Nugent C, et al. Bayesian Approaches to Subgroup Analysis and Related Adaptive Clinical Trial Designs. JCO Precis Oncol. 2019 Oct 24;3:PO.19.00003.
- Zhang S, et al. Subgroup Analyses in Reporting of Phase III Clinical Trials in Solid Tumors. *J Clin Oncol*. 2015;33(15):1697-1702.
- Sun X, et al. How to Use a Subgroup Analysis: Users' Guide to the Medical Literature. *JAMA.* 2014;311(4):405–411.
- Wang R et al. Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials N Engl J Med 2007; 357:2189-2194

April 5, 2022

How to understand subgroup analysis in clinical studies

Massimo Di Maio

SCDU Medical Oncology
AO Ordine Mauriziano, Torino, Italy
Department of Oncology,
University of Turin
massimo.dimaio@unito.it

@MassimoDiMaio75

dimaiomax