Heterogenous populations: subgroup analyses for survival benefit

Paolo Bruzzi – Clinical Epidemiology
National Cancer Research Institute
Genoa - Italy

Topics

- Appropriate methods and their use
- Risks when not used
- Problems when used
- Fancy recent approaches to subgroup analyses
- Non-proportional hazards and possible developments

Should we treat differently different groups of patients?

- Different toxicity? (Genetics, Comorbidities, co-medications, etc.)
- Different efficacy? Effect Modifiers (Characteristics of Disease/Patient)
- Patient choices? (values, life-expectancy, priorities, etc.)

Subgroup analyses

Aim of subgroup analyses

To provide information on the opportunity to treat differently different groups of patients

‘Personalized’(?) Therapies

Precision (?) Medicine

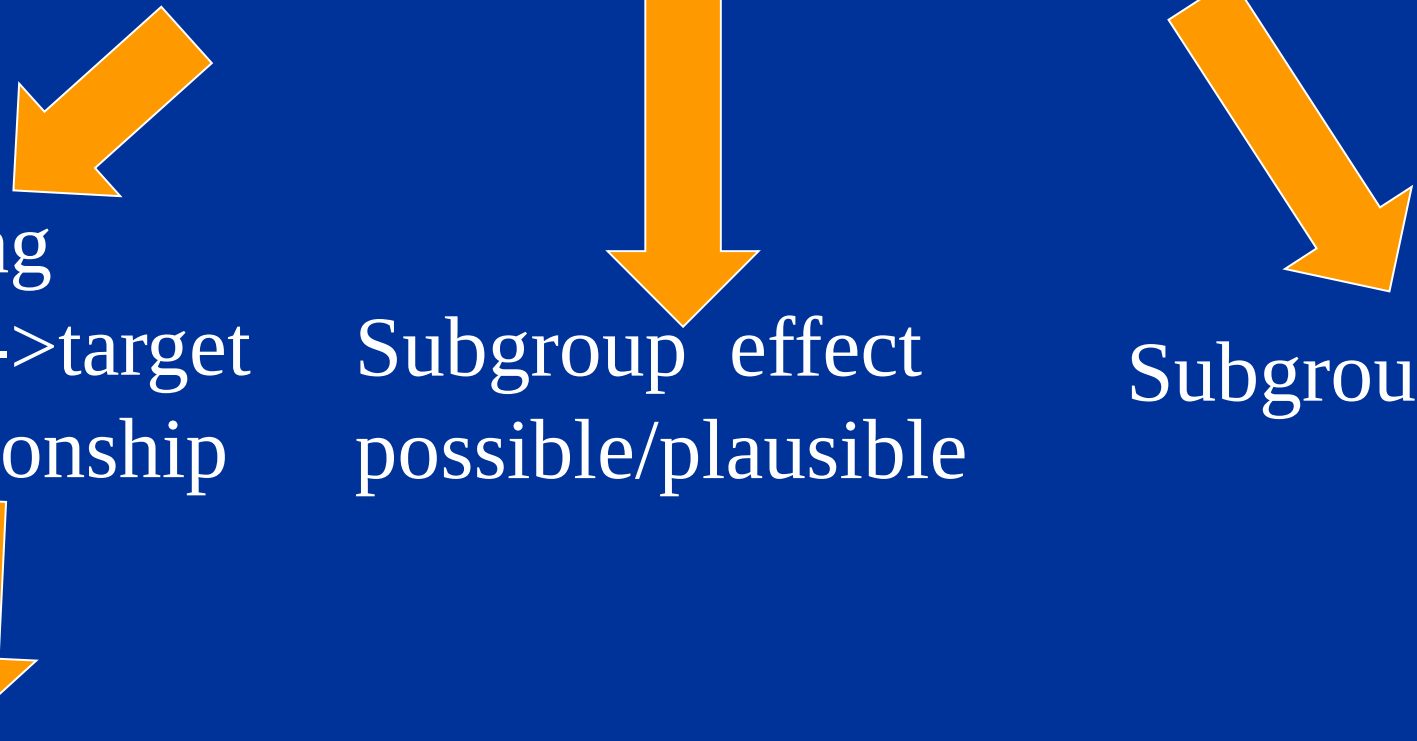
Personalised/Precision Medicine

Strong drug->target relationship

Subgroup effect possible/plausible

Subgroup effect?

Personalised/Precision Medicine



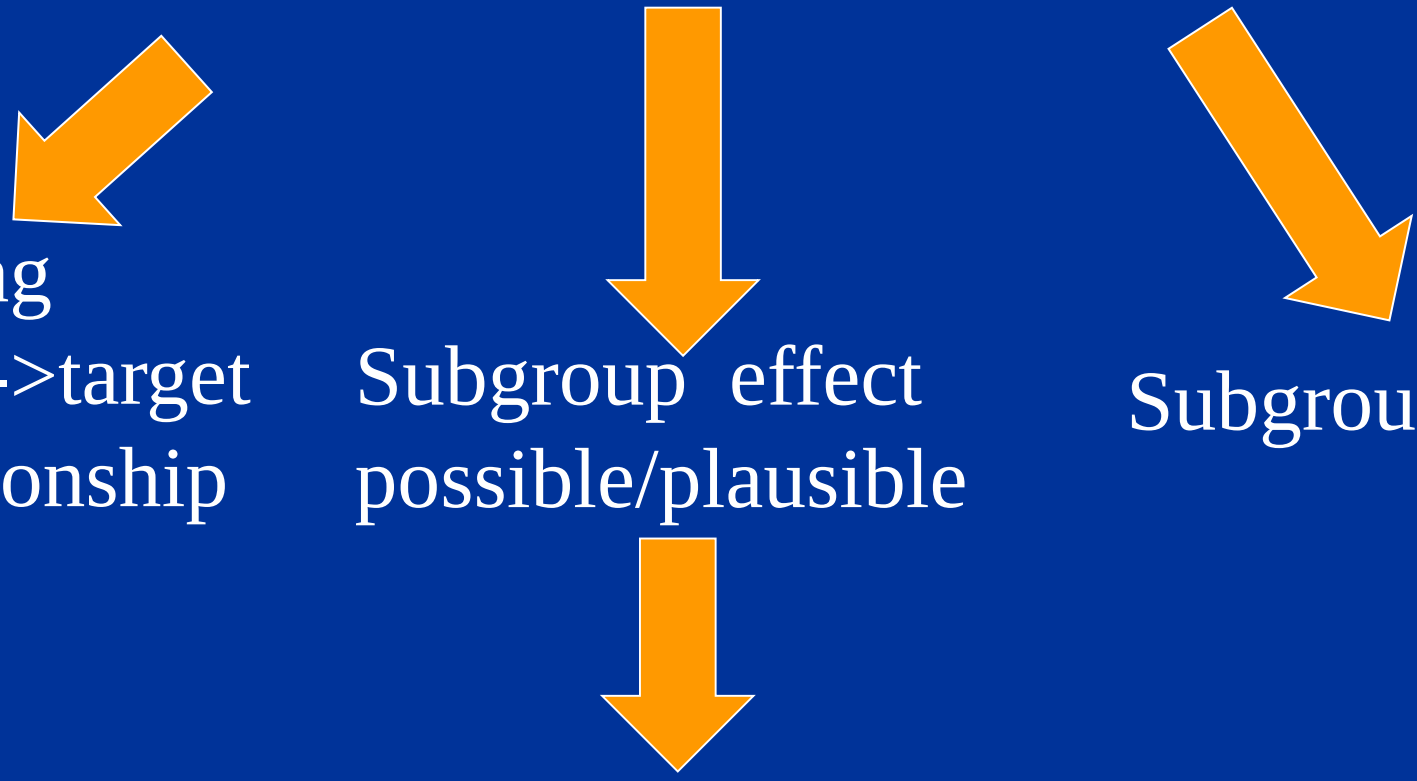
Strong drug->target relationship

Trials in subgroups (Umbrella, Basket, RCT in selected patients, etc)

Subgroup effect possible/plausible

Subgroup effect?

Personalised/Precision Medicine



Phase II in unselected pts -> Subgroup analyses ->
-> Phase III in subgroup(s)? (Adaptive designs, Seamless trials, etc) – Enriched Populations

Seamless Phase II-III trials for selection of subgroups

- Potentially one of the most powerful tools to speed up the development of cancer drugs

- Statistical issues

- **SHORT TERM ENDPOINTS** (Response, Biomarker, et.) ?

Personalised/Precision Medicine

Strong drug->target relationship

Subgroup effect possible/plausible

Subgroup effect?

Phase III in unselected patients

Positive? Subgroup analysis?

Negative? Subgroup analysis?

Subgroup analyses

Traditional perspective

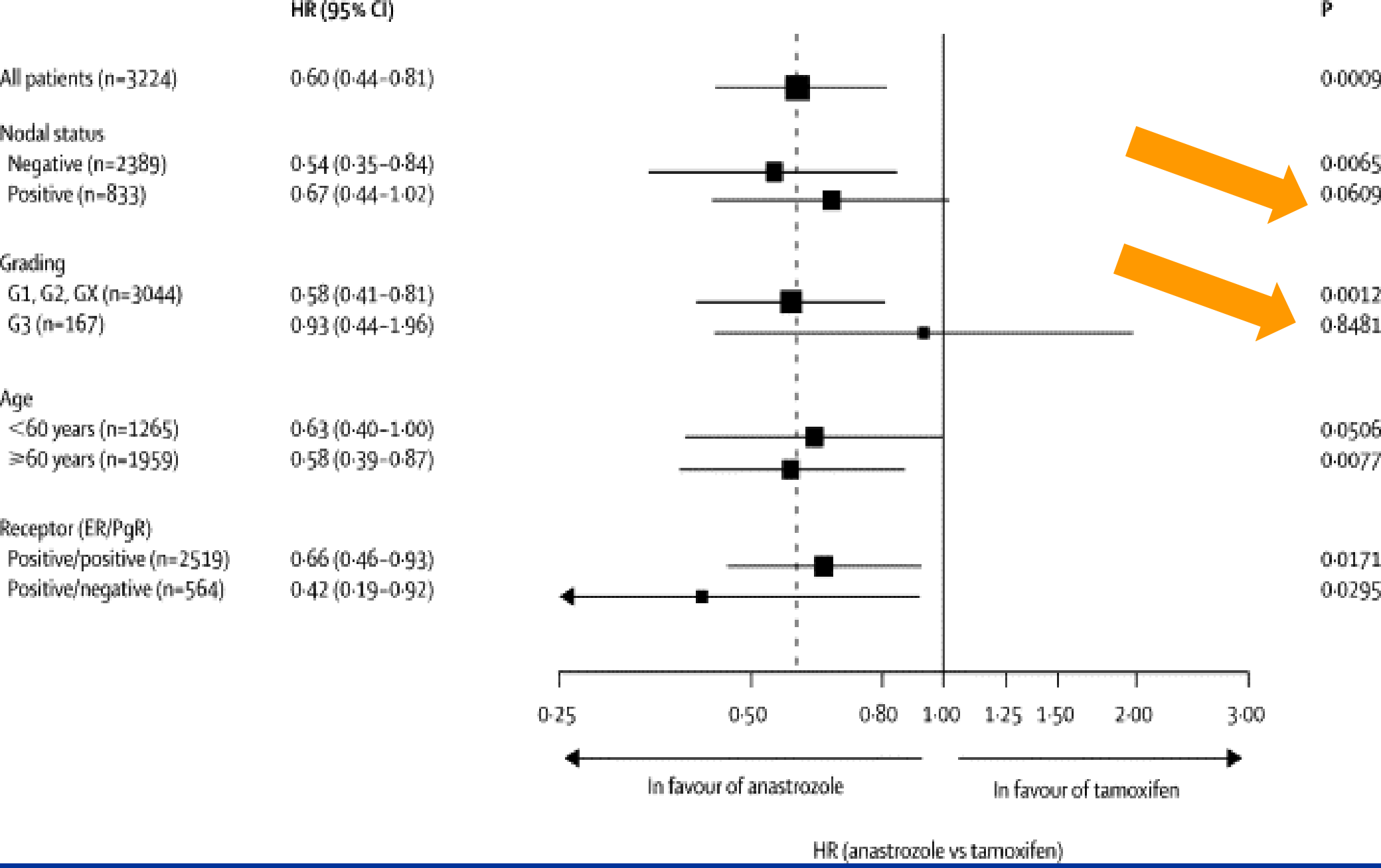
- Resuscitation of agonizing studies
- To publish irrelevant results
- Torturing data until they confess

EBM & Subgroup analyses (S.A.)

- S.A. considered with suspicion
  - Not necessary: average effect is fine for estimates of effectiveness (Cardiology, Early Breast C., Prevention)
  - **Many poorly conducted S.A.**

Provocation (R. Peto): Astrological sign

- Second International Study of Infarct Survival (ISIS 2) (*Lancet*, 1988)
  - 2 x 2 factorial design
    (aspirin vs. placebo and streptokinase vs. placebo)
  - vascular and total mortality in patients with an acute myocardial infarction (MI)
  - **Gemini or Libra astrological birth signs** did somewhat worse on aspirin while all other signs and overall results impressive and highly significant benefit from aspirin



Forrest plot (qui p x sottogruppo – sbagliate)

Subgroup analyses: Classical Problems

Metodological

- Retrospective vs prospective
- Planned vs unplanned
- Bias

Statistical

- Proper test of significance
- Multiplicity

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

1 23 January 2014
2 EMA/CHMP/539146/2013
3 Committee for Medicinal Products for Human Use (CHMP)

4 Guideline on the investigation of subgroups in 5 confirmatory clinical trials

6 DRAFT

Clinical Review & Education

Users' Guides to the Medical Literature

How to Use a Subgroup Analysis
Users' Guides to the Medical Literature

Xin Sun, PhD; John P. A. Ioannidis, MD, DSc; Thomas Agoritsas, MD; Ana C. Alba, MD; Gordon Guyatt, MD, MSc

Clinicians, when trying to apply trial results to patient care, need to individualize patient care and, potentially, manage patients based on results of subgroup analyses. Apparently compelling subgroup effects often prove spurious, and guidance is needed to differentiate credible from less credible subgroup claims. We therefore provide 5 criteria to use when assessing the validity of subgroup analyses: (1) Can chance explain the apparent subgroup effect; (2) Is the effect consistent across studies; (3) Was the subgroup hypothesis one of a small number of hypotheses developed a priori with direction specified; (4) Is there strong preexisting biological support; and (5) Is the evidence supporting the effect based on within- or between-study comparisons. The first 4 criteria are applicable to individual studies or systematic reviews, the last only to systematic reviews of multiple studies. These criteria will help clinicians deciding whether to use subgroup analyses to guide their patient care.

JAMA. 2014;311(4):405-411. doi:10.1001/jama.2013.285063

Supplemental content at jama.com

Author Affiliations: Author affiliations are listed at the end of this article.

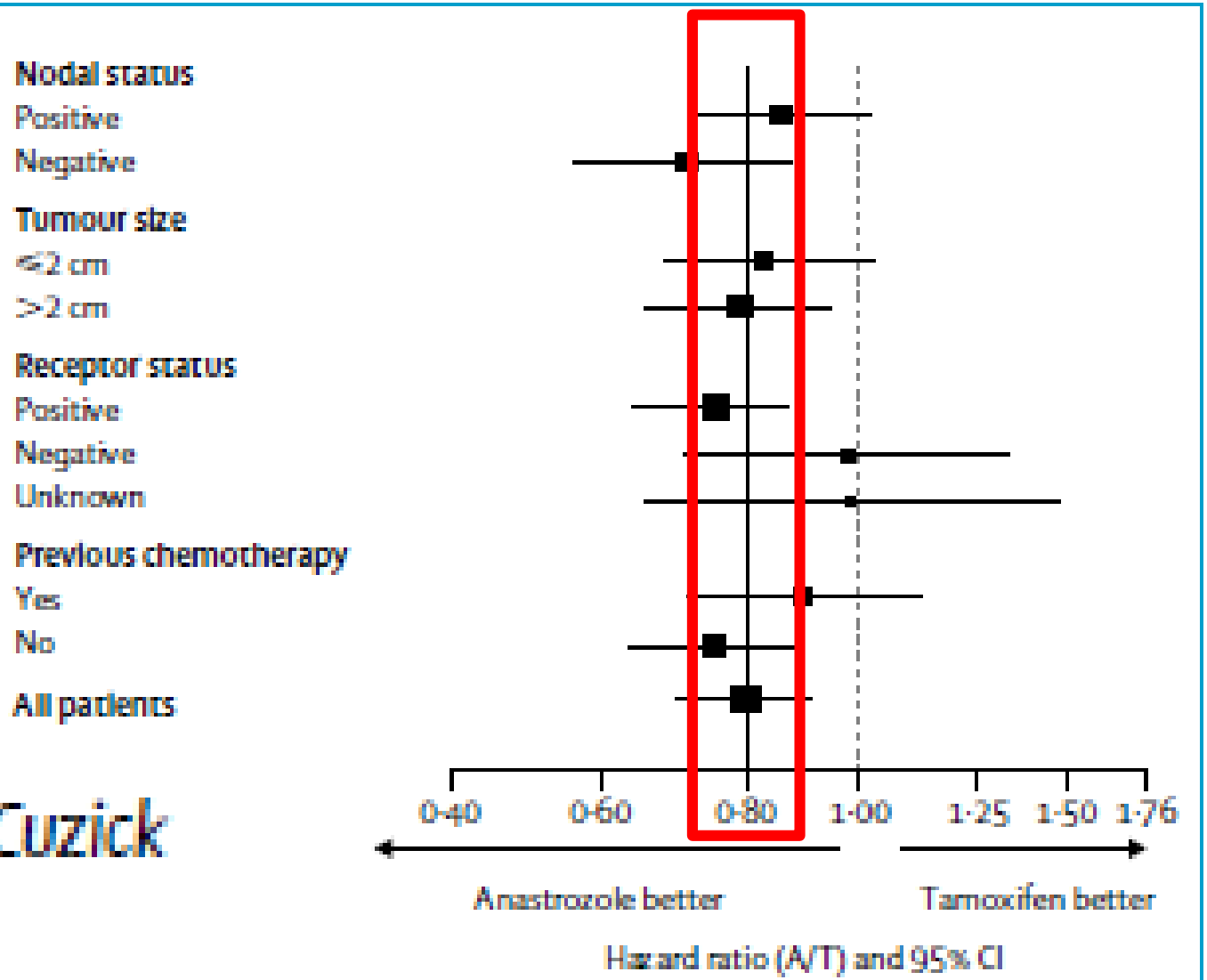
Corresponding Author: Gordon H. Guyatt, MD, MSc, Departments of Clinical Epidemiology and Biostatistics and Medicine, Faculty of Health Sciences, McMaster University, 1200 Main St W, Room 2C12, Hamilton, Ontario, L8N 3Z5, Canada (guyatt@mcmaster.ca).

Subgroup analyses

- **Methodology has become standardised**
  - Careful planning to prevent selection and assessment biases
  - Test for interaction = H0: the (lack of) effect is the same in all subgroups **(No sugroup specific p's)**
  - Multiplicity controlled (Planned vs Post-hoc, Exploratory vs confirmatory analyses, Corrections of p values)

No problems with large datasets

2003; 24: 26–66.



Jack Cuzick

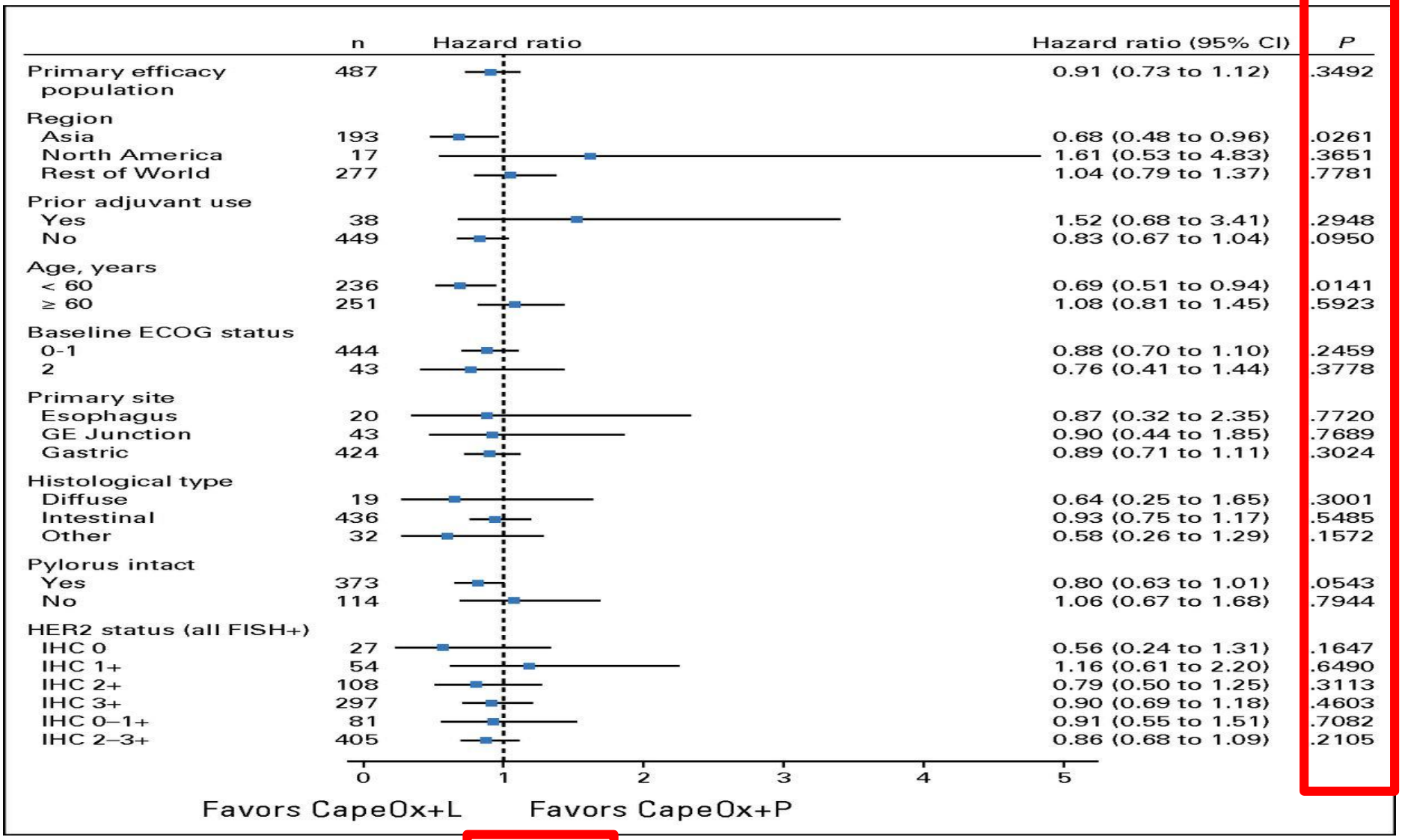
Figure: Suggested modification as applied to time to recurrence in subgroups of the ATAC trial[3]

Subgroup analyses

- **Methodology has become standardised**

This methodology is widely and uniformly used?

Forest plot for overall survival: subgroup analysis.



J. Randolph Hecht et al. JCO 2016;34:443-451

VOLUME 34 · NUMBER 10 · APRIL 1, 2016

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Final Results of the Sunbelt Melanoma Trial: A Multi-Institutional Prospective Randomized Phase III Study Evaluating the Role of Adjuvant High-Dose Interferon Alfa-2b and Completion Lymph Node Dissection for Patients Staged by Sentinel Lymph Node Biopsy

benefit from HDI in Protocol A (Fig 5). In patients with a single positive SLN, HDI was associated with an improvement in DFS only in patients with ulceration (HR, 0.43; 95% CI, 0.21 to 0.87; $P = .0183$; n = 75)[9] and with Breslow thickness more than 4 mm (HR, 0.35; 95% CI, 0.14 to 0.88; $P = .0259$; n = 42). No improvement

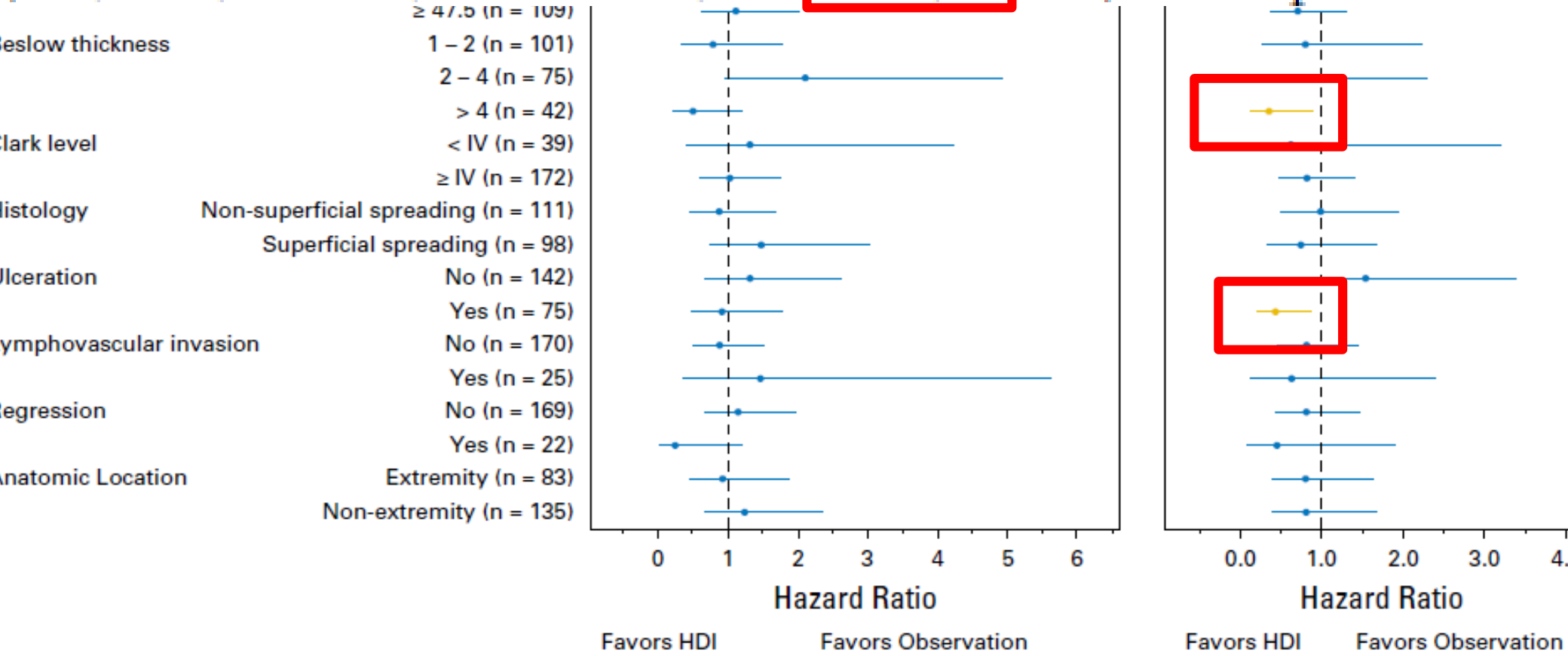


Fig 5. Subgroup analysis of Protocol A, examining differences in disease-free and overall survival in patients with a single positive sentinel lymph node randomly assi

Quality of Subgroup analyses in cancer trials
Luca Carmisciano, MD Thesis (2016)

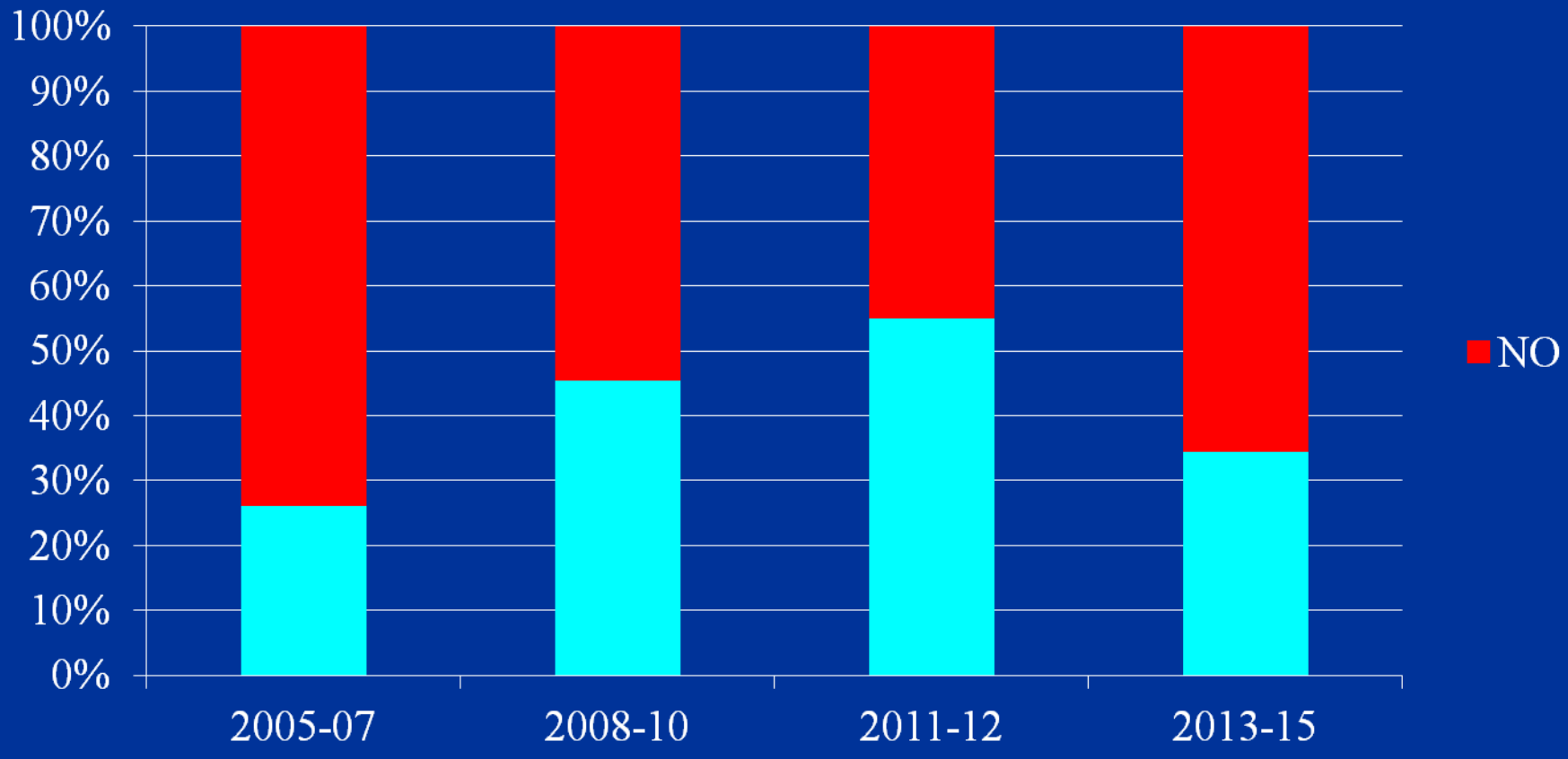
- Random sample of Phase II-III RCTs of cancer therapies from 6 major journals
- (Ann Oncol, J Clin Oncol, J Natl Cancer Inst, Lancet, Lancet Oncol, N Engl J Med)

published between 2005-2015.

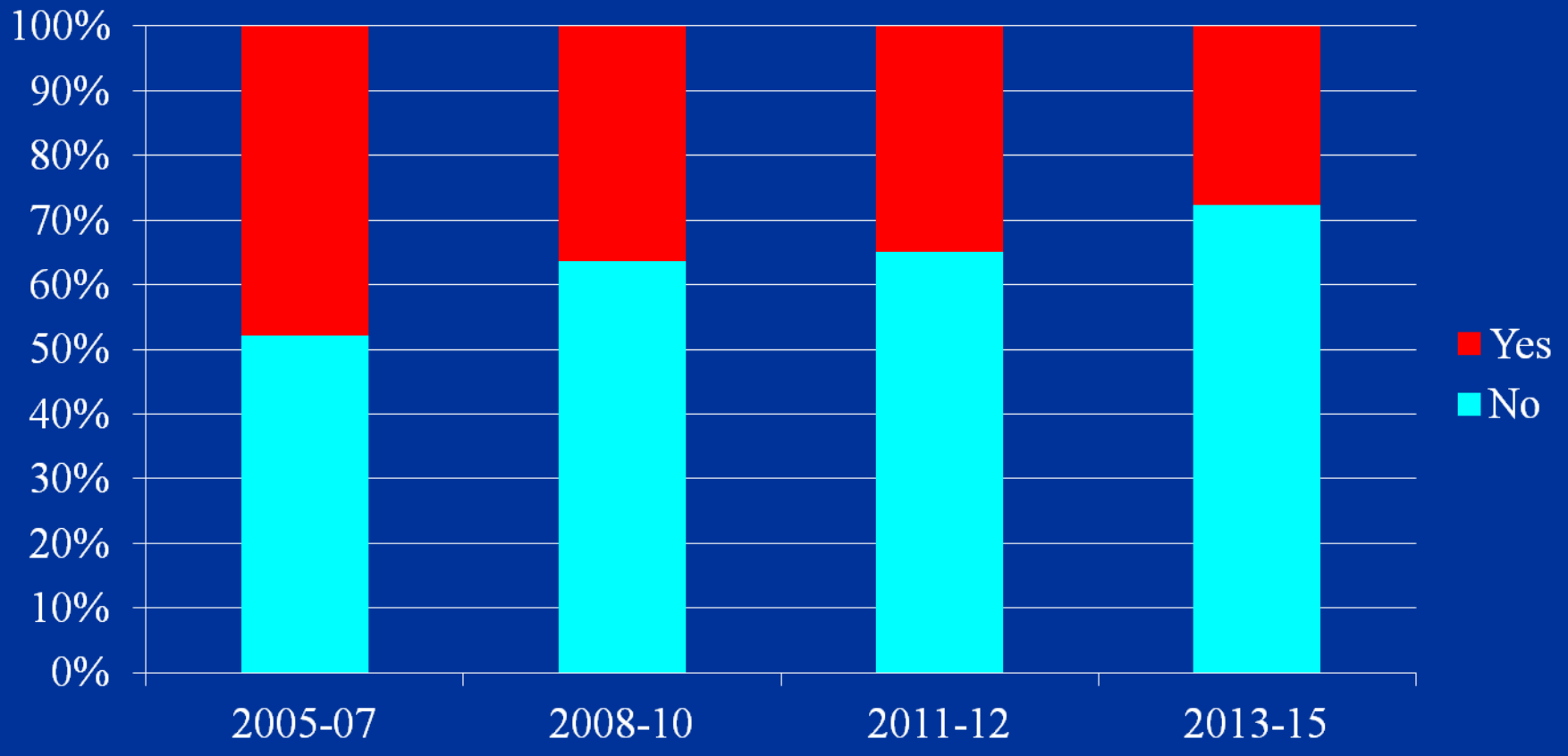
261 RCTs, 137 eligible,

94 RCTs with subgroup analyses

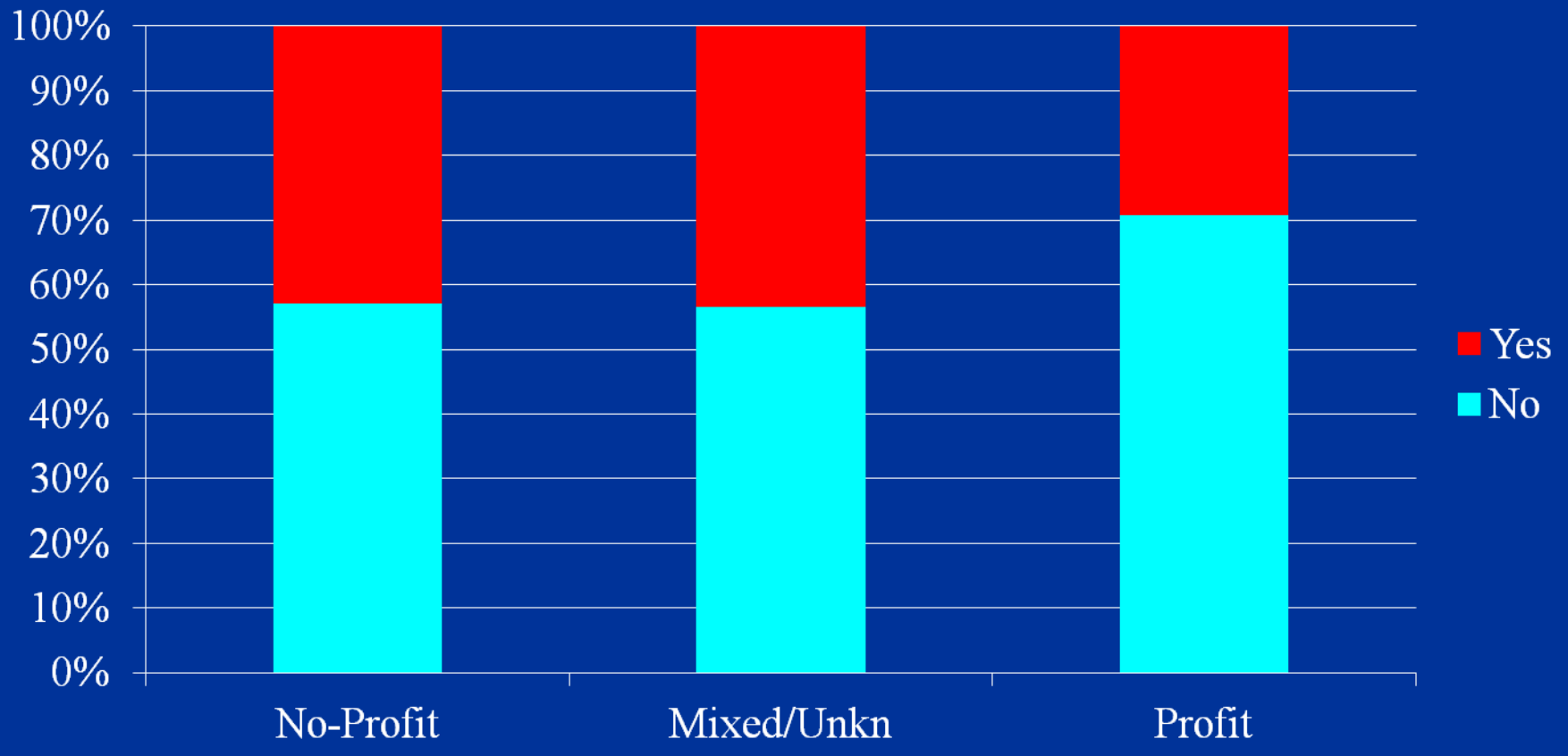
Interaction test



Subgroup specific p values



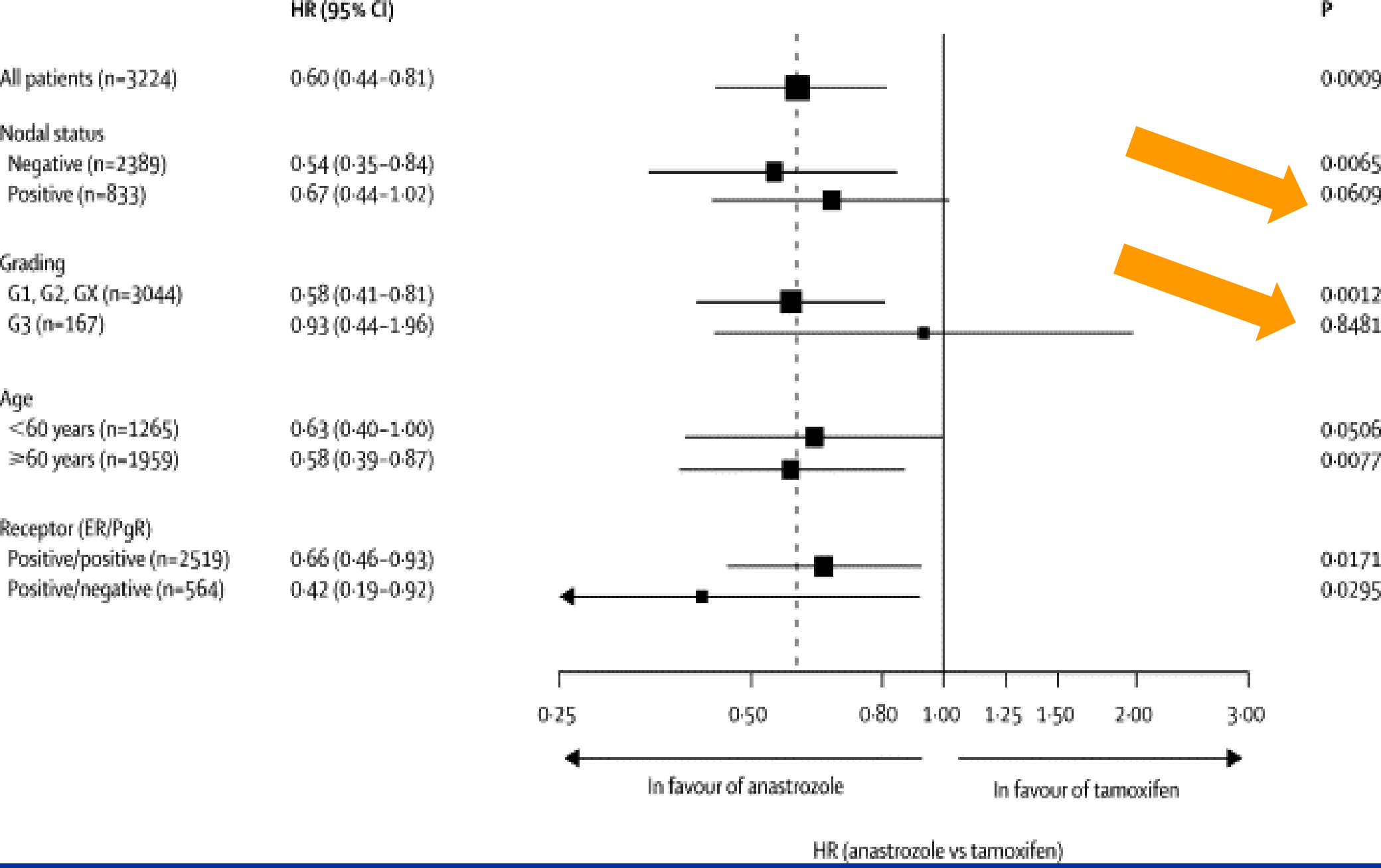
Subgroup-specific p values



Consequences of flawed S.A.

Overall Positive Study:

- False Positive S.A. (= No effect in some subgroup)
  - Small subgroups
  - Multiplicity



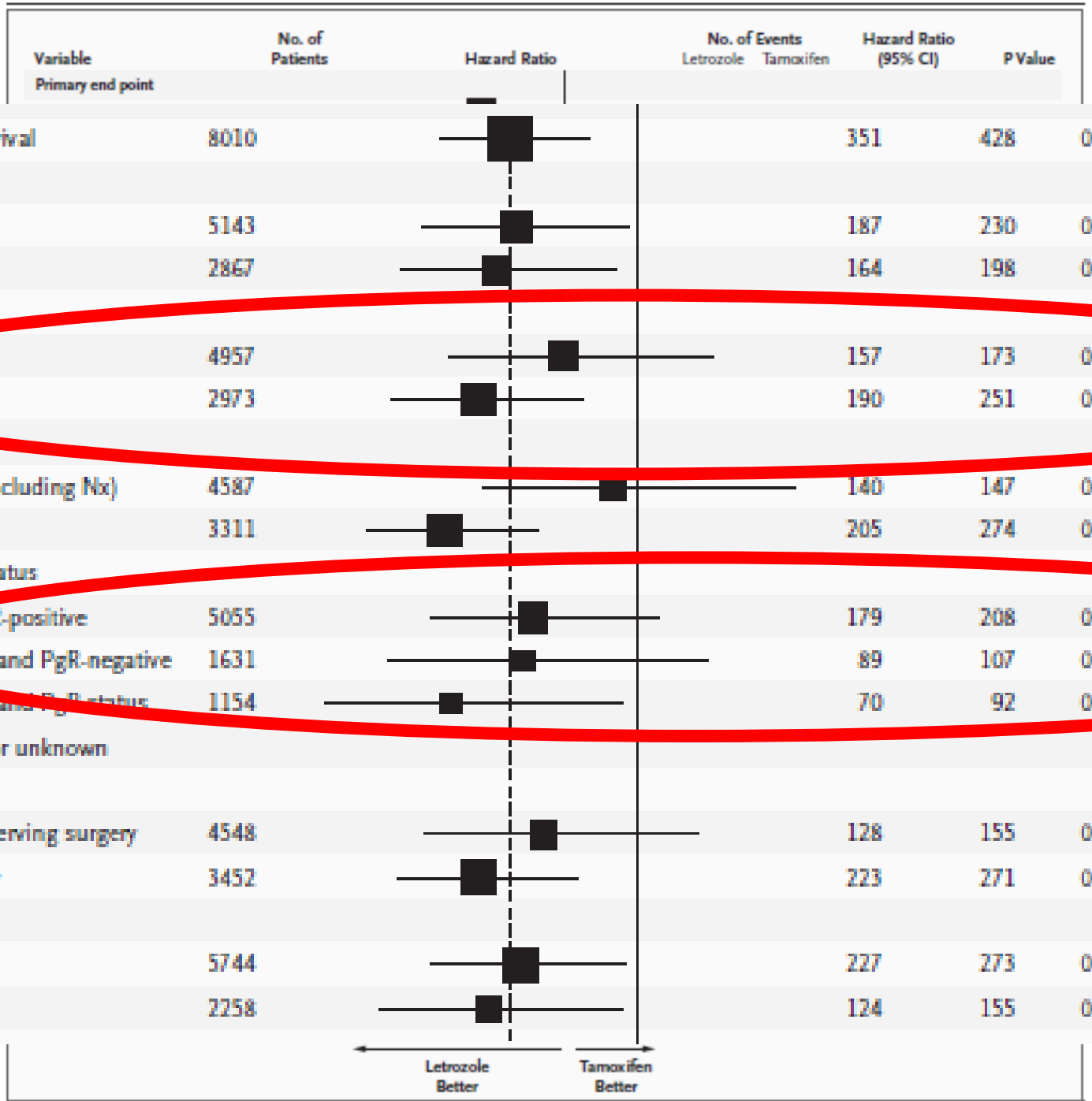
Forrest plot (qui p x sottogruppo – sbagliate)

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 DECEMBER 29, 2005 VOL. 353 NO. 26

A Comparison of Letrozole and Tamoxifen in Postmenopausal Women with Early Breast Cancer

The Breast International Group (BIG) 1-98 Collaborative Group*


Variable
No. of Patients
Hazard Ratio
No. of Events
Letrozole
Tamoxifen
Hazard Ratio (95% CI)
P Value
Primary end point
Disease-free survival 8010 351 428 0.81 (0.70–0.93) 0.00
Age
<65 yr 5143 187 230 0.82 (0.67–0.99) 0.04
≥65 yr 2867 164 198 0.79 (0.64–0.97) 0.02
Tumor size
≤2 cm 4957 157 173 0.89 (0.72–1.10) 0.28
>2 cm 2973 190 251 0.76 (0.63–0.92) 0.004
Nodal
Negative (including Nx) 4587 140 147 0.96 (0.76–1.21) 0.75
Positive 3311 205 274 0.71 (0.59–0.85) <0.001
ER and PgR status
and PgR-positive 5055 179 208 0.84 (0.69–1.03) 0.09
ER-positive and PgR-negative 1631 89 107 0.83 (0.62–1.10) 0.18
ER-positive 1154 70 92 0.72 (0.53–0.98)
missing or unknown
Local therapy
Breast-conserving surgery 4548 128 155 0.86 (0.68–1.08) 0.20
Mastectomy 3452 223 271 0.76 (0.64–0.91) 0.003
Radiotherapy
Yes 5744 227 273 0.82 (0.69–0.98) 0.03
No 2258 124 155 0.77 (0.61–0.98) 0.03
Letrozole Better
Tamoxifen Better

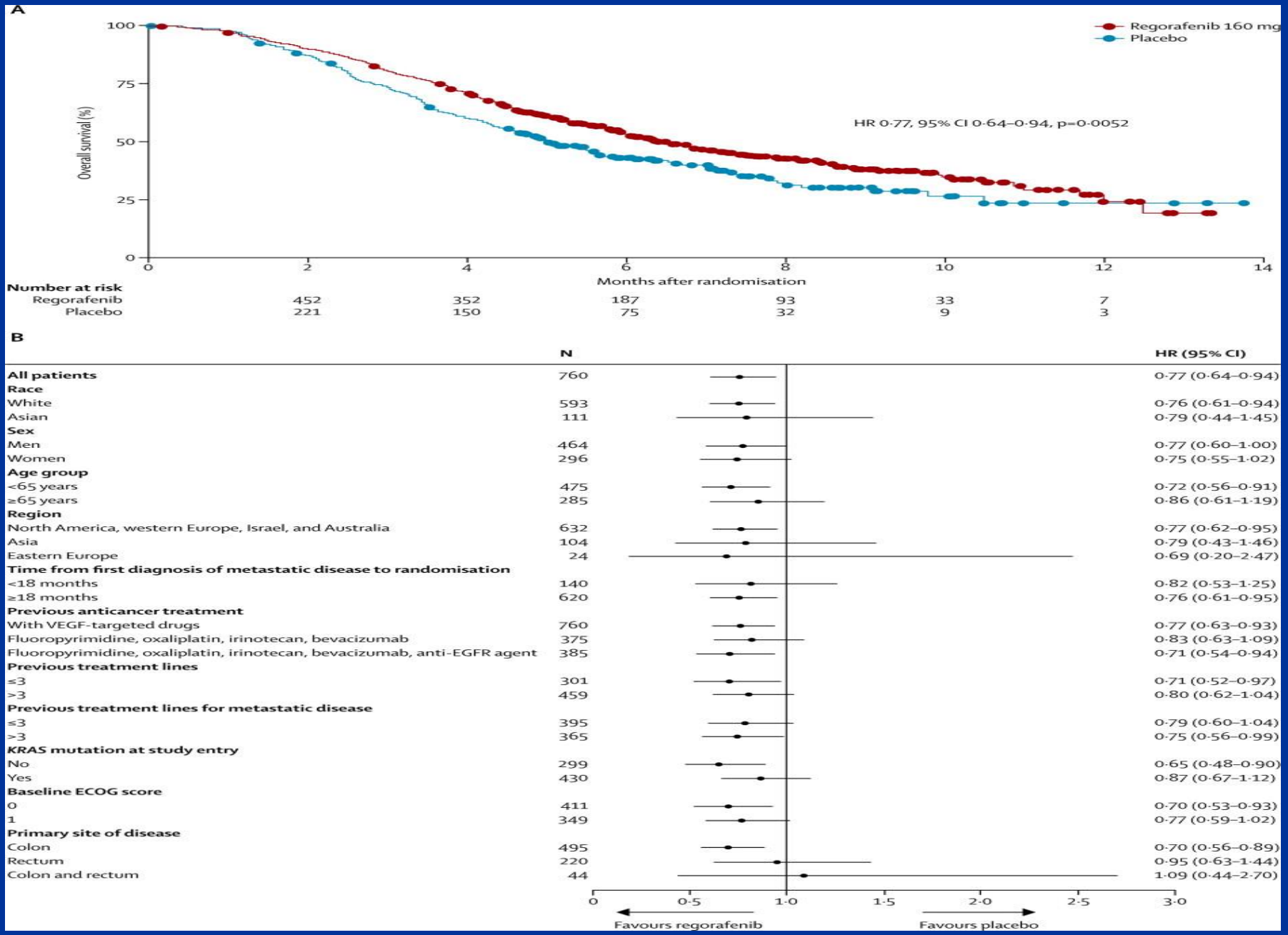
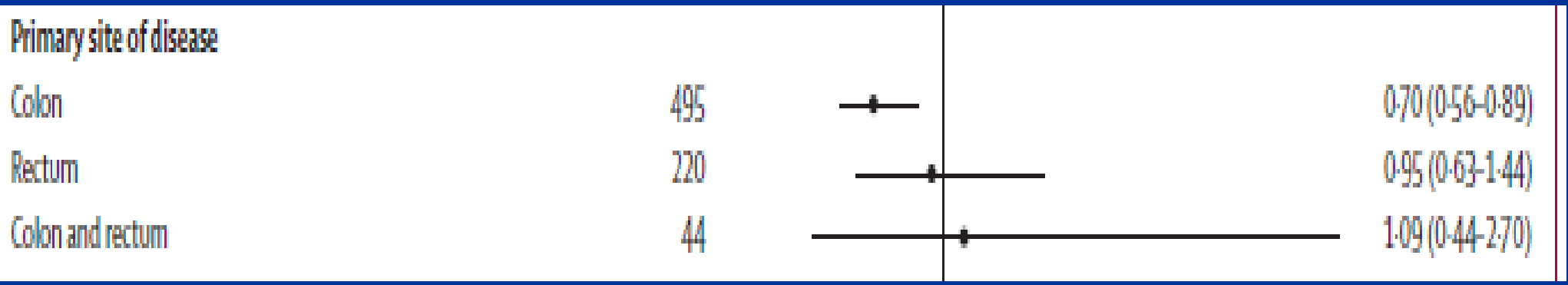


Figure 2. Overall survival(A) Kaplan-Meier analysis, intention-to-treat population. (B) Subgroup analysis. HR=hazard ratio. ECOG=Eastern Cooperative Oncology Group.

…regorafenib showed an apparent benefit in 24 of 25 subgroups, the exception being the group of patients with primary disease in colon and rectum, which was based on only a few events…???

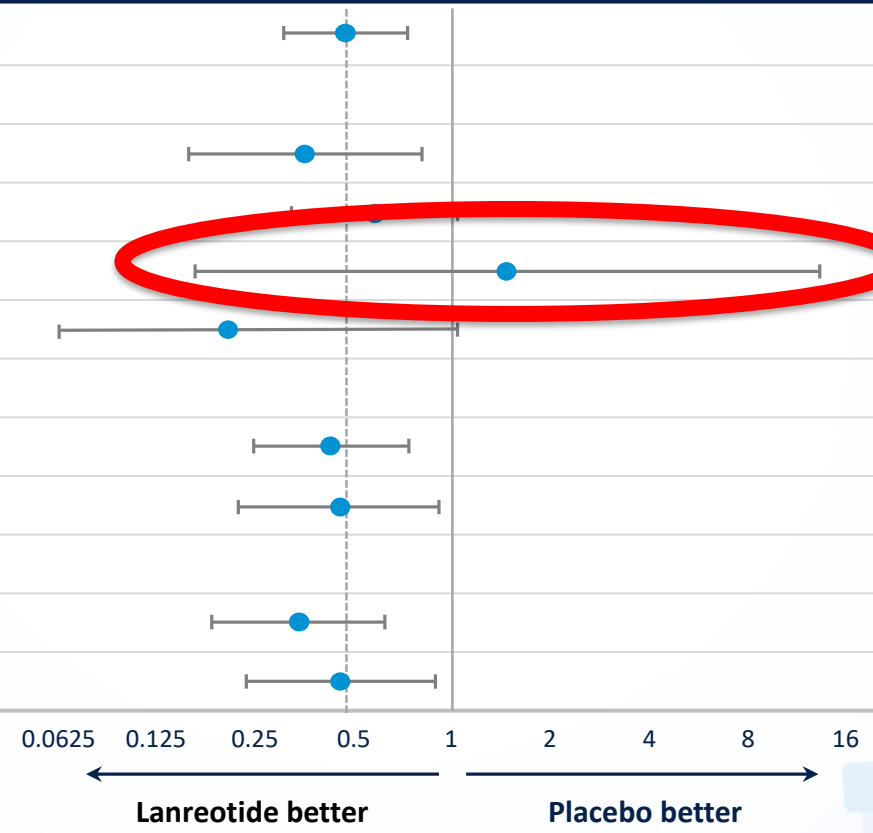


No mention anywhere of "interaction"/heterogeneity" nor of multiplcity

Lancet, 2012.

PFS: therapeutic effect in pre-defined subgroups generally consistent with overall population

Subgroup	No.	Hazard Ratio (95% Confidence Interval)
All patients	204	0.47 (0.30–0.73)
Tumour origin		
Midgut	73	0.35 (0.16–0.80)
Pancreas	91	0.58 (0.32–1.04)
Hindgut	14	1.47 (0.16–13.24)
Other/unknown	26	0.21 (0.04–1.03)
Tumour grade		
G1 tumour	141	0.43 (0.25–0.74)
G2 tumour	61	0.45 (0.22–0.91)
Hepatic tumour volume		
≤25%	137	0.34 (0.18–0.62)
>25%	67	0.45 (0.23–0.88)



Subgroups pre-defined although number of categories for hepatic tumour volume was simplified post hoc from five to two.

Table 1. Baseline Demographic and Disease Characteristics of the Patients (Intention-to-Treat Population).*

Variable	Lanreotide (N=101)	Placebo (N=103)
Male sex — no. (%)	53 (52)	54 (52)
Age — yr	63.3±9.8	62.2±11.1
Time since diagnosis — mo		
Mean	32.6±46.1	34.4±41.4
Median	13.2	16.5
Prior treatment for neuroendocrine tumor — no. (%)	16 (16)	16 (16)
Primary tumor resected — no. (%)	40 (40)	39 (38)
Origin of neuroendocrine tumor — no. (%)†		
Pancreas	42 (42)	49 (48)
Midgut	33 (33)	40 (39)
Hindgut	11 (11)	3 (3)
Unknown or other	15 (15)	11 (11)
Tumor progression — no. (%)	4 (4)	5 (5)
Tumor grade — no. (%)‡		
1: Ki-67 0–2%	69 (68)	72 (70)
2: Ki-67 3–10%	32 (32)	29 (28)
Data missing	0	2 (2)

CLARINET

PFS: therapeutic effect in pre-defined subgroups generally consistent with overall population

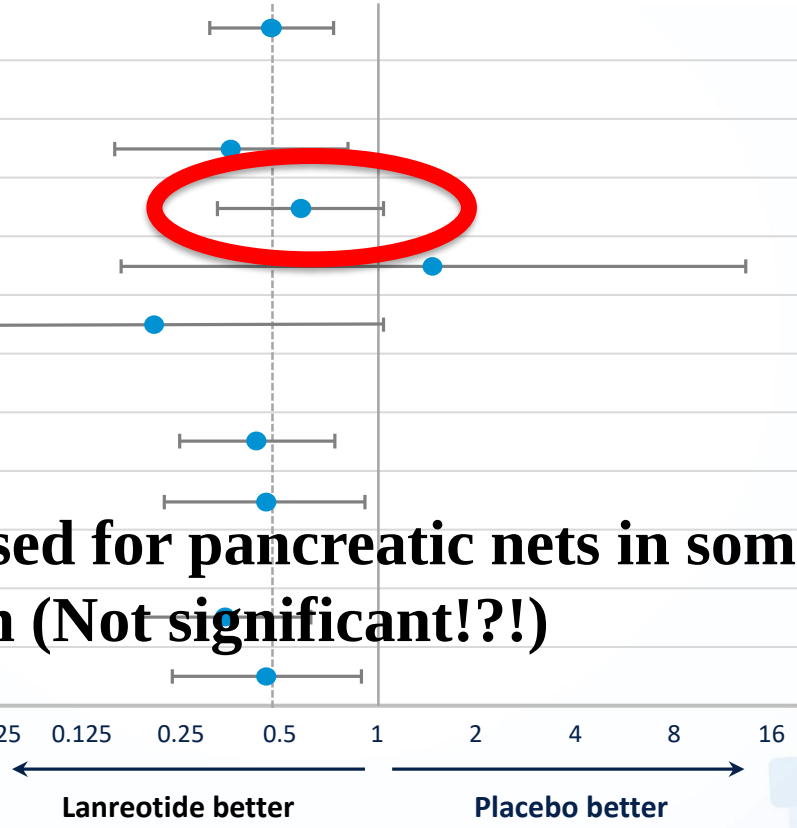
Subgroup	No.	Hazard Ratio (95% Confidence Interval)	
All patients	204		0.47 (0.30–0.73)
Tumour origin			
Midgut	73		0.35 (0.16–0.80)
Pancreas	91		0.58 (0.32–1.04)
Hindgut	14		1.47 (0.16–13.24)
Other/unknown	26		0.21 (0.04–1.03)
Tumour grade			
G1 tumour	141		0.43 (0.25–0.74)

Hazard ratios for disease progression or death generally favored lanreotide over placebo in the predefined subgroups. ***The exceptions were the smaller subgroups (e.g., the subgroup of patients with tumors originating in the hindgut*** *[Fig. 2]), for which the hazard ratios had wide confidence intervals and the findings were imprecise.*

Wrong—Test of interaction!

PFS: therapeutic effect in pre-defined subgroups generally consistent with overall population

Subgroup	No.	Hazard Ratio (95% Confidence Interval)
All patients	204	0.47 (0.30–0.73)
Tumour origin		
Midgut	73	0.35 (0.16–0.80)
Pancreas	91	0.58 (0.32–1.04)
Hindgut	14	1.47 (0.16–13.24)
Other/unknown	26	0.21 (0.04–1.03)
Tumour grade		
G1 tumour	141	0.43 (0.25–0.74)
G2 tumour	61	0.45 (0.22–0.91)
Hepatic tumour volume		
≤25%	137	0.34 (0.18–0.62)
>25%	67	0.45 (0.23–0.88)



Not reimbursed for pancreatic nets in some Italian region (Not significant!?!)

Subgroups pre-defined although number of categories for hepatic tumour volume was simplified post hoc from five to two.

Consequences of flawed S.A.

Overall Positive Study:

- False Positive S.A.
- False Negative S.A. ?

Overall Negative Study:

- **FALSE POSITIVE S.A.!**

(Multiplicity - subgroup specific p's)

Consequences of flawed S.A.

Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives

ST Brookes
E Whitley
TJ Peters
PA Mulheran
M Egger
G Davey Smith

NHS HTA report

Health Technology Assessment 2001; Vol. 5: No. 33

Methodology

Consequences of flawed S.A.

Simulation study: False positive rate (assuming no interaction)

Interaction test: 5% (as expected)

Subgroup specific analyses: 7- 66% (depending on presence and size of the overall treatement effect)

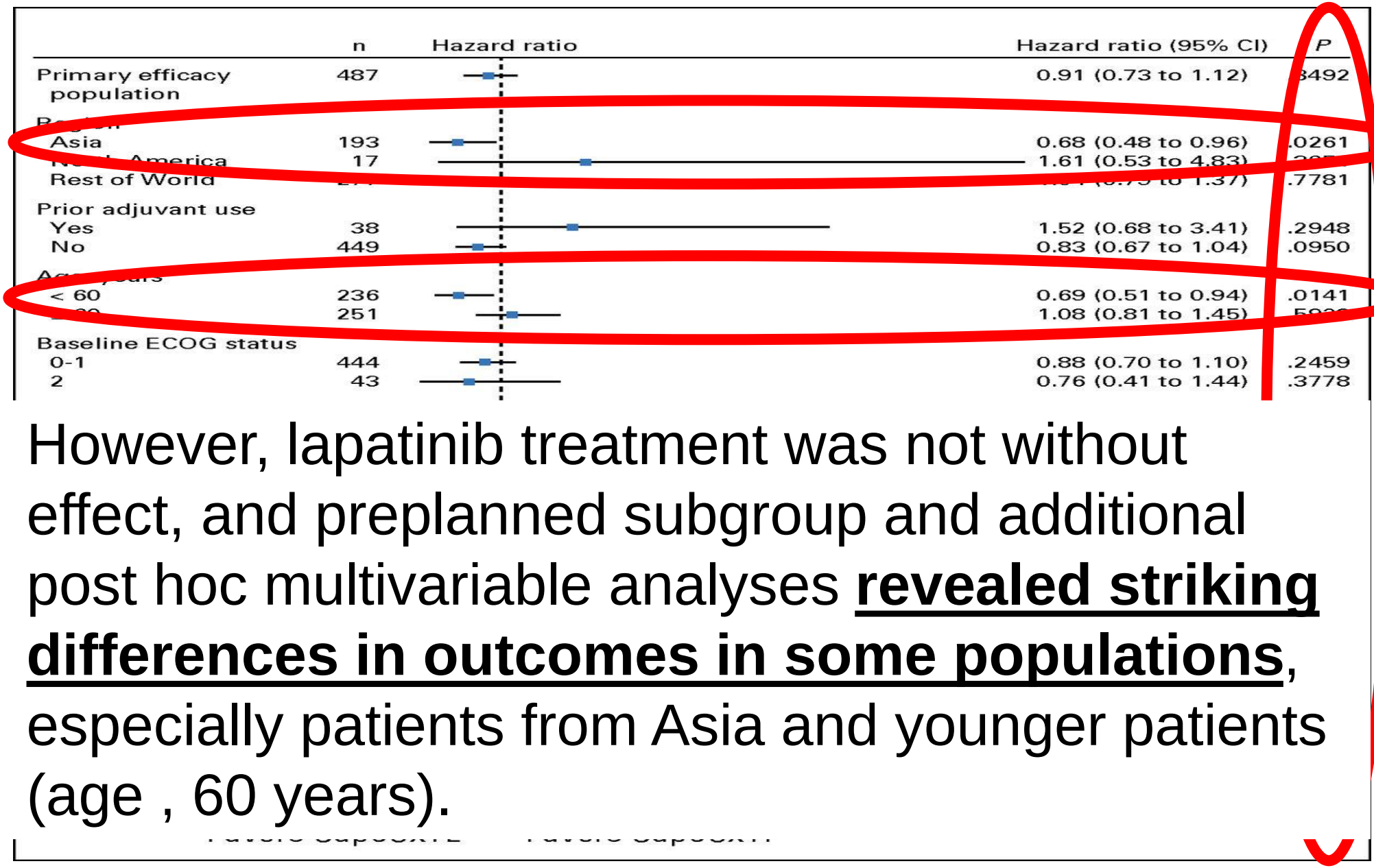
Subgroup specific analyses are particularly unreliable and are affected by many factors. Subgroup analyses should be always based on formal tests of interaction

NHS HTA report

Health Technology Assessment 2001; Vol. 5: No. 33

Methodology

Forest plot for overall survival: subgroup analysis.



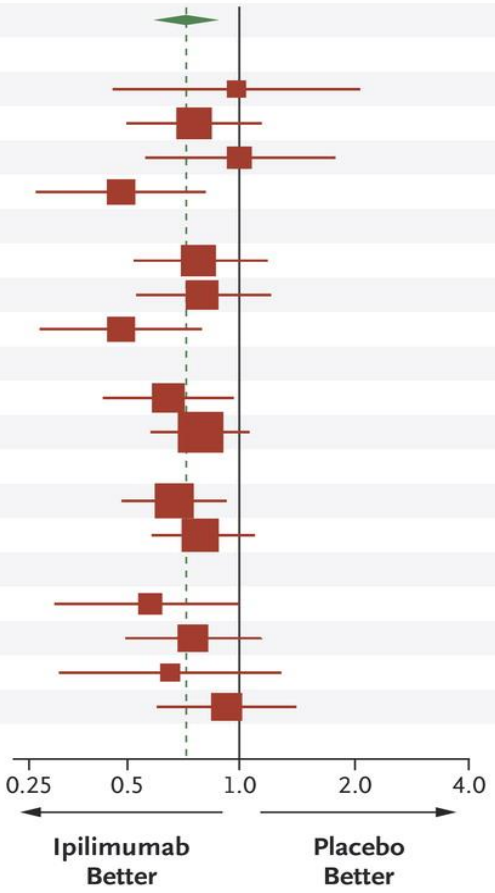
However, lapatinib treatment was not without effect, and preplanned subgroup and additional post hoc multivariable analyses **revealed striking differences in outcomes in some populations**, especially patients from Asia and younger patients (age , 60 years).

J. Randolph Hecht et al. JCO 2016;34:443-451

JOURNAL OF CLINICAL ONCOLOGY ASCO

Correct subgroup analysis

Subgroup	Ipilimumab	Placebo	Hazard Ratio (95% or 99% CI)	P Value
	no. of deaths/total no.			
All patients	162/475	214/476	0.72 (0.59–0.88)	0.001
Disease stage				0.07
IIIA	24/98	22/88	0.98 (0.46–2.09)	
IIIB	68/213	85/207	0.75 (0.50–1.14)	
IIIC with 1–3 positive lymph nodes	34/69	45/83	1.00 (0.56–1.80)	
IIIC with ≥4 positive lymph nodes	36/95	62/98	0.48 (0.28–0.81)	
No. of positive lymph nodes				0.09
1	65/217	82/220	0.79 (0.52–1.21)	
2 or 3	61/163	70/158	0.83 (0.53–1.30)	
≥4	36/95	62/98	0.48 (0.28–0.81)	
Type of positive lymph node				0.21
Microscopic	54/210	76/193	0.61 (0.39–0.96)	
Macroscopic	108/265	138/283	0.80 (0.58–1.11)	
Ulceration				0.29
Yes	73/197	110/203	0.64 (0.44–0.94)	
No	79/257	88/244	0.80 (0.54–1.20)	
Lymph-node and ulceration status				0.35
Microscopic and ulceration	28/99	43/88	0.54 (0.29–0.99)	
Macroscopic and ulceration	45/98	67/115	0.76 (0.46–1.23)	
Microscopic and no ulceration	21/104	29/97	0.62 (0.30–1.29)	
Macroscopic and no ulceration	58/153	59/147	0.90 (0.56–1.45)	



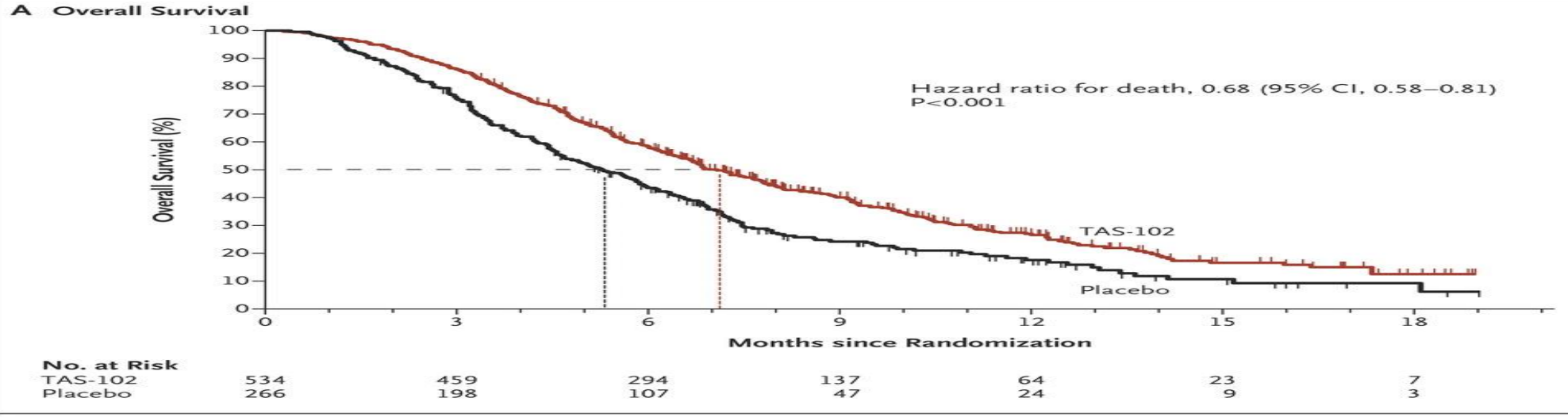
Eggermont AM et al. N Engl J Med 2016. DOI: 10.1056/NEJMoa1611299

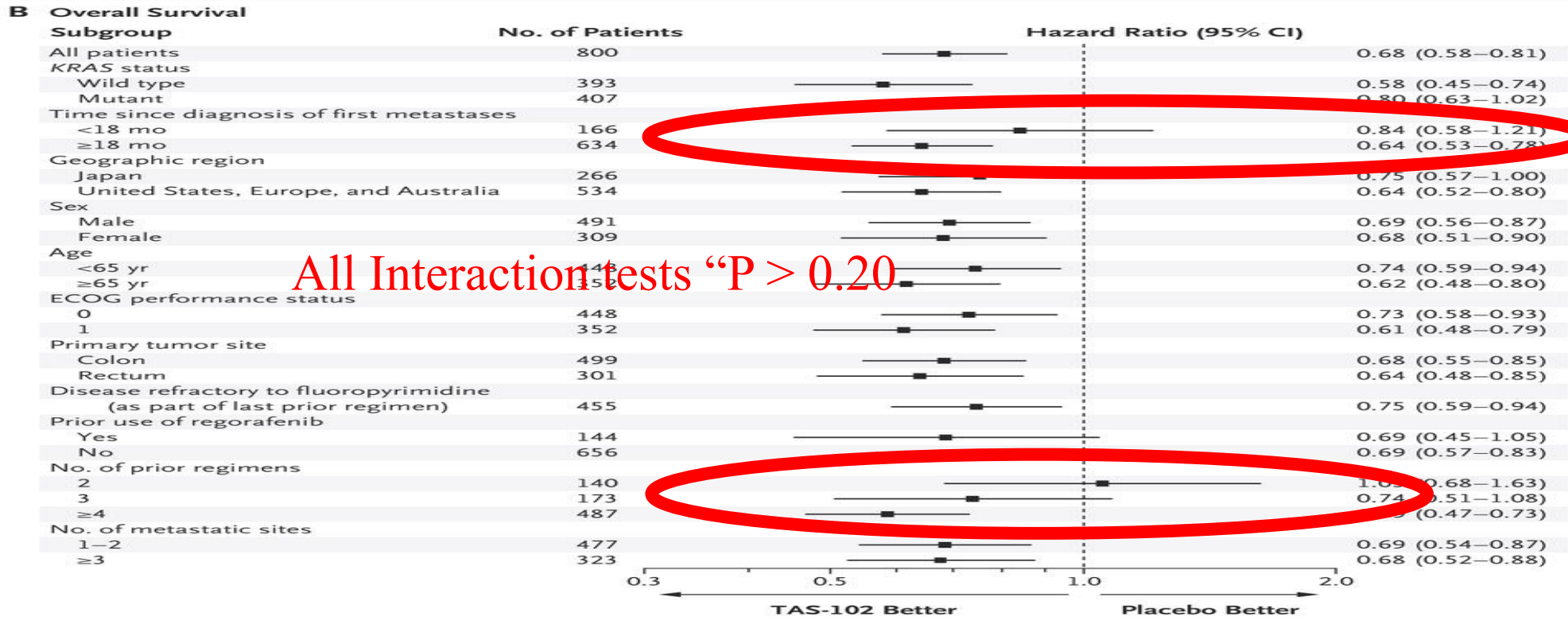
The NEW ENGLAND JOURNAL of MEDICINE

EBM & Subgroup analyses

- *S.A. considered with suspicion*
  - *Many poorly conducted S.A.*
  - *Not necessary: average effect is fine for estimates of effectiveness (Cardiology, BC, Prevention)*
- **S.A. may provide important information**
  - **Properly conducted** S.A. are unbiased …
  - But require many patients (Interaction tests -> Low power)

  Risk of false negative results!





Mayer RJ et al. N Engl J Med 2015;372:1909-1919.

The NEW ENGLAND JOURNAL of MEDICINE

Low power of subgroup analyses:

Solutions

1) Inflation (Increase in sample size needed to preserve the desired power in the subgroup analysis – No correction for multiplicity)

Inflation factor :

1 if HR(i) = twice HR(overall)

4 if HR(i) = HR (overall)

increases exponentialy as HR(i)/HR(o) decreases

Low power of subgroup analyses:

Solutions

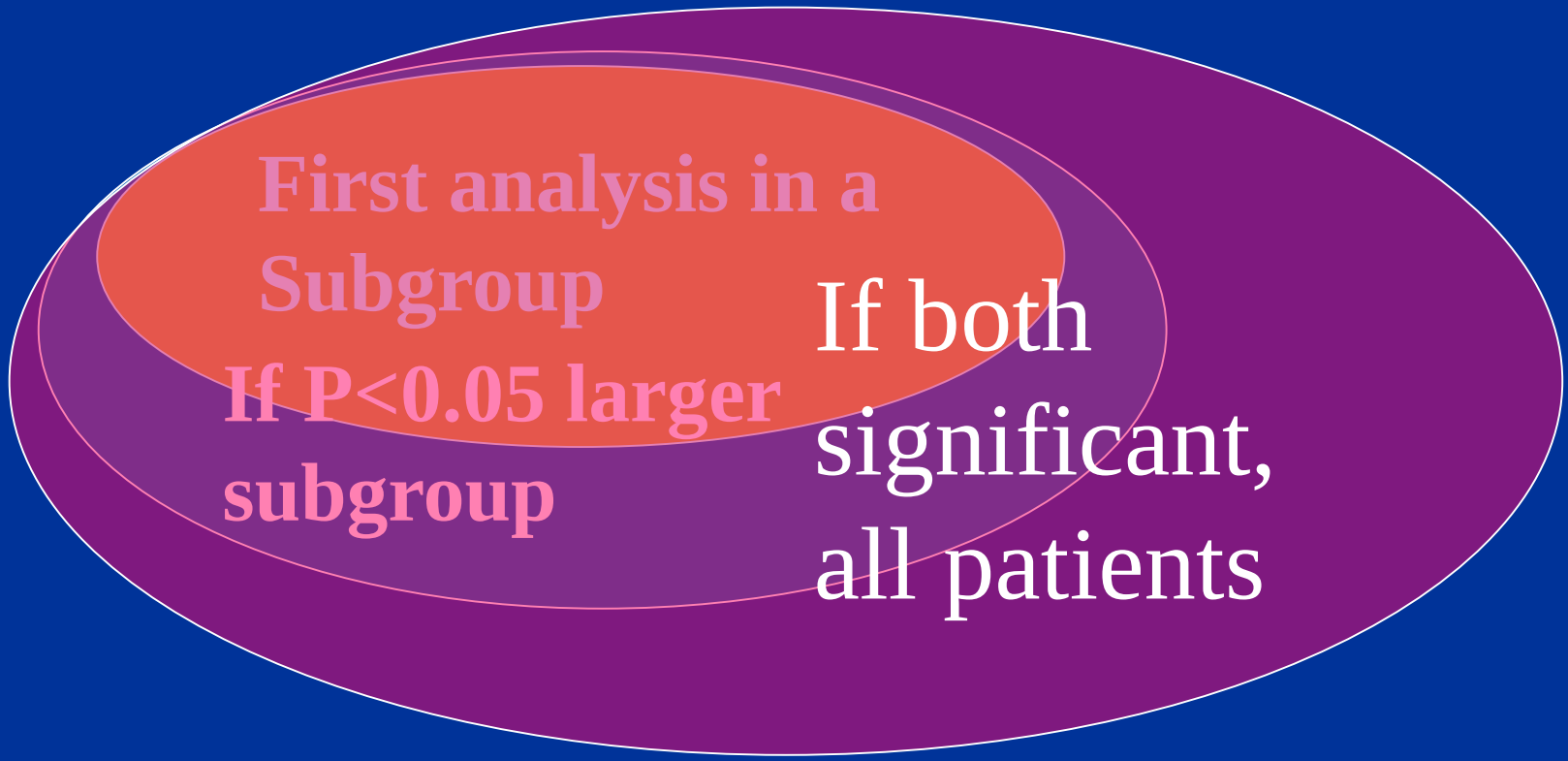
1) Inflation
2) Meta-analyses
   - Difficult with modern drugs (1 condition 1 trial)
   - Across similar drugs (e.g. anti-PD1)?????

Low power of subgroup analyses:

Solutions

1) Inflation
2) Meta-analyses
3) Enriched populations (Oversampling of candidate subgroup)
   - Statistical Problems
     - Primary H0?
     - Hierarchy of tests? (Closed test procedure?)

New Approach: “Reversed” subgroup analysis



Example of reversed S.A.

(*Shortened from Lancet Oncol 2014; 15: 143–55*)

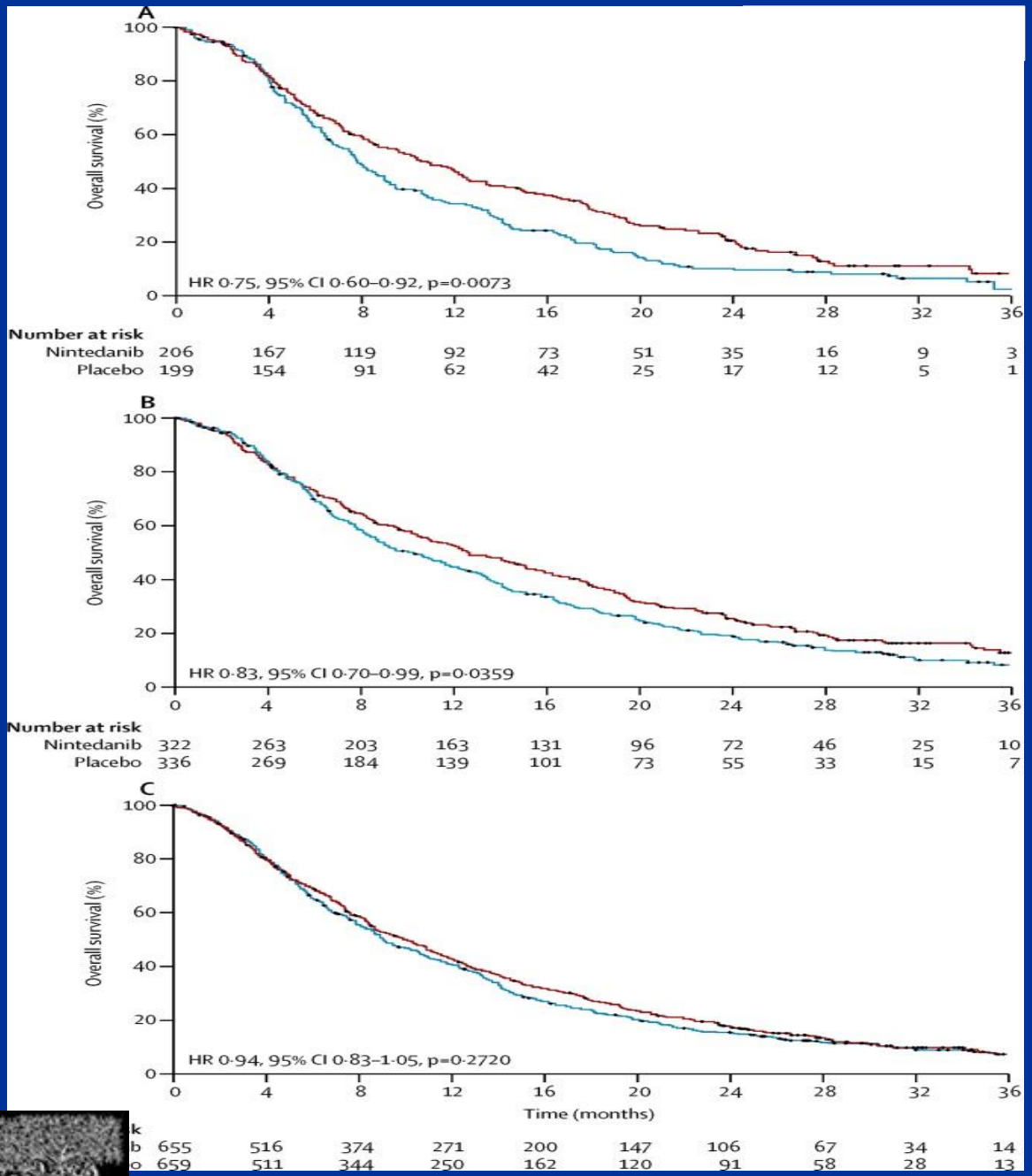
"The key secondary endpoint was OS, analysed by ITT after 1121 events,

in a prespecified stepwise order:

1. first in patients with ADK who progressed within 9 months after start of first-line TX,

2. then in all patients with ADK,

3. then in all patients"



405 pts,
HR=0.75,
p=0.007

658 pts,
HR=0.83,
p=0.03

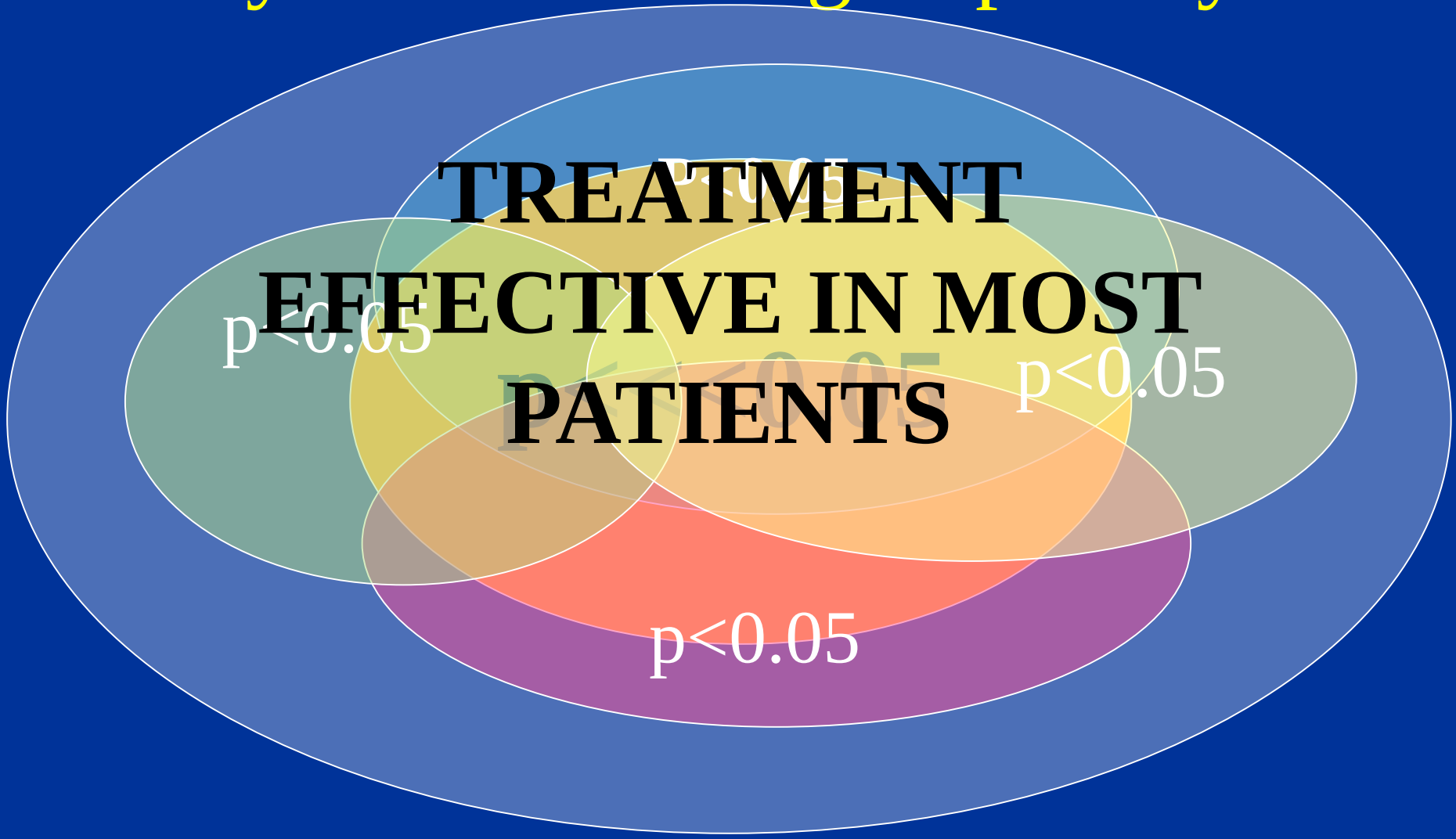
1314 pts,
HR=0.94,
p=0.27

The Lancet Oncology 2014 15, 143-155DOI: (10.1016/S1470-2045(13)70586-2)

Reversed Subgroup Analysis

- It is NOT a closed test procedure
  - The results of 2nd & 3rd analysis are not independent from those of the 1st
- Statistically speaking, it's nonsense
  - If the H01 is false, H02 and H03 are false
- Using it in a smart way, you can demonstrate almost everything

Tricky Reversed subgroup analysis



What to do when a prominent effect in a subgroup is suspected?

1. Closed test procedure – No multiplicity – useless because overall effect required

2. Standard subgroup analysis -> Multiplicity -> Loss of power -> **Enrichment**

3. **Subroup analysis: Primary analysis – Then other subgroups**

Low power = Less precision
(Im)Precision Medicine?

- Smaller trials -> Larger sampling error **(wider CL)**
- Bayesian Statistics -> Reliance on prior estimates (assumptions=subjectivity?)
- Uncontrolled trials -> Bias

Uncertainty in the Era of Precision Medicine

David J. Hunter, M.B., B.S., Sc.D.

N ENGL J MED 375;8 NEJM.ORG AUGUST 25, 2016

The New England Journal of Medicine

"...will precision medicine usher in an age of diagnostic and prognostic certainty?"

"The new tools for tailoring treatment ***will demand a greater tolerance of uncertainty*** *and greater facility for calculating and interpreting probabilities than we have been used to as physicians and patients"*

Is it possible to reconcile

-

a) the need of objectivity and statistical precision (EBM)

-

with

-

b) the growing demand of personalised medicine?

Issues in subgroup analyses

- *Appropriate methods and their use*
- *Risks when not used*
- *Problems when used*
- Fancy recent approaches to subgroup analyses

The NEW ENGLAND JOURNAL of MEDICINE
ESTABLISHED IN 1812
AUGUST 12, 2010 VOL. 363 NO. 8

Improved Survival with Ipilimumab in Patients with Metastatic Melanoma

IMMUNO CANCER THERAPY!?!

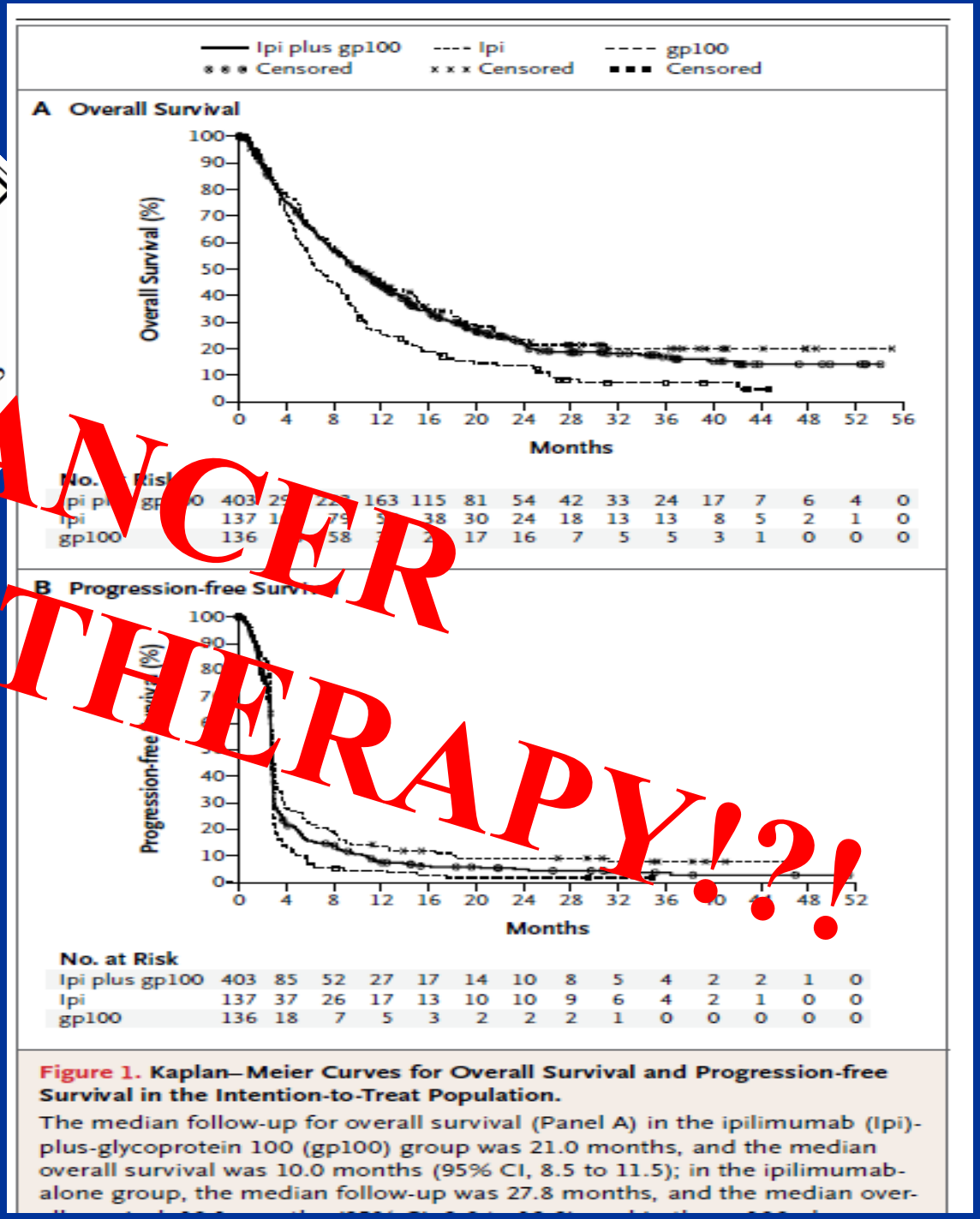


Figure 1. Kaplan–Meier Curves for Overall Survival and Progression-free Survival in the Intention-to-Treat Population.

The median follow-up for overall survival (Panel A) in the ipilimumab (Ipi)-plus-glycoprotein 100 (gp100) group was 21.0 months, and the median overall survival was 10.0 months (95% CI, 8.5 to 11.5); in the ipilimumab-alone group, the median follow-up was 27.8 months, and the median over-

Immunotherapy in cancer

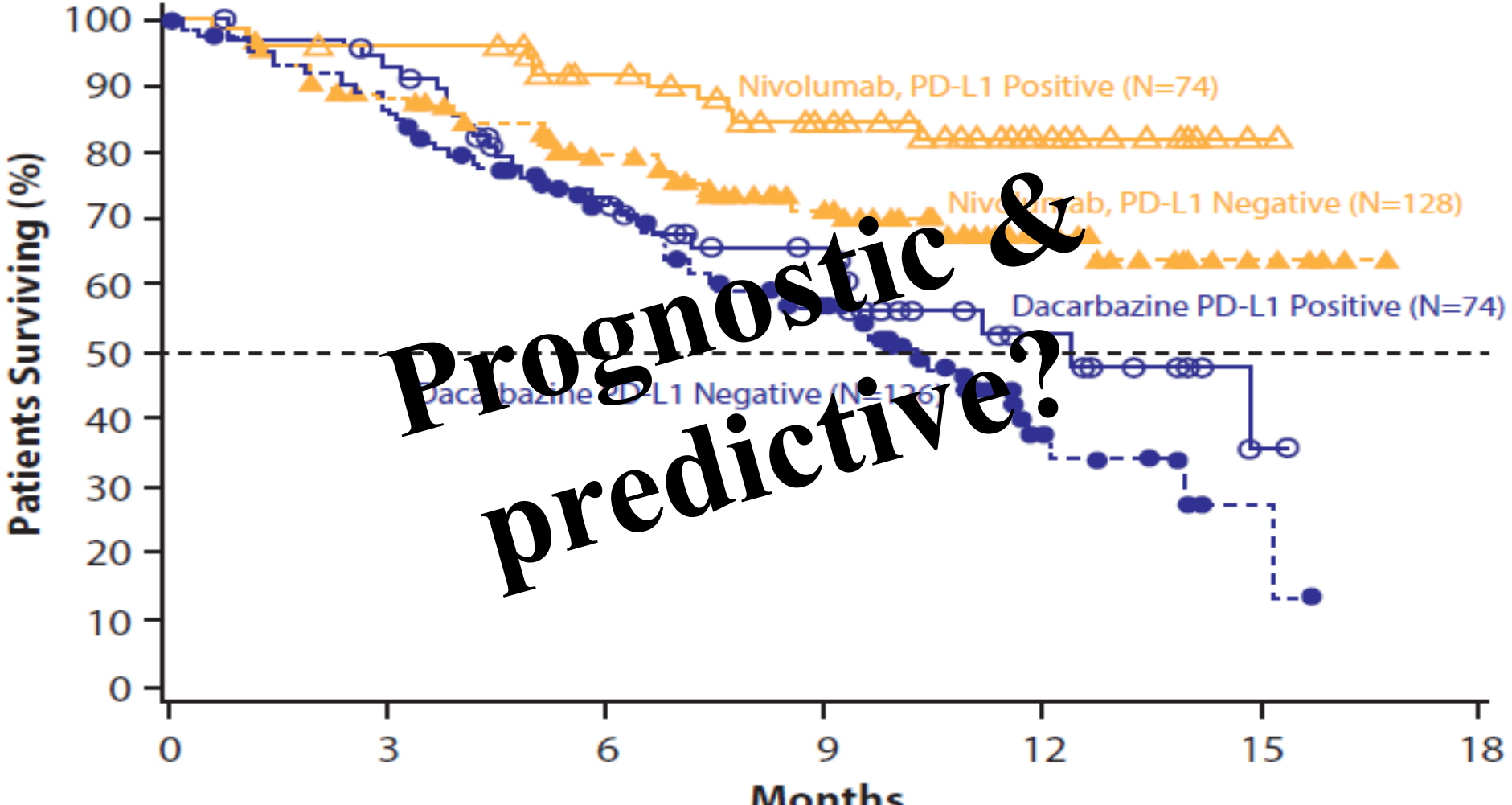
- Outstanding benefits
  - Effects in refractory cancers
  - Effects on OS
  - Long term effects in a fraction of patients
    - Anti-CTL4 : 10-20%
    - Anti-PD1/PDL1: 10-30%
- Outstanding Costs

Cancer Immunotherapy
Predictive factors

- *Histology (NSCLC, Squamous vs ADK)*
- *Mutational Burden*
- **PD-L1 expression level?**

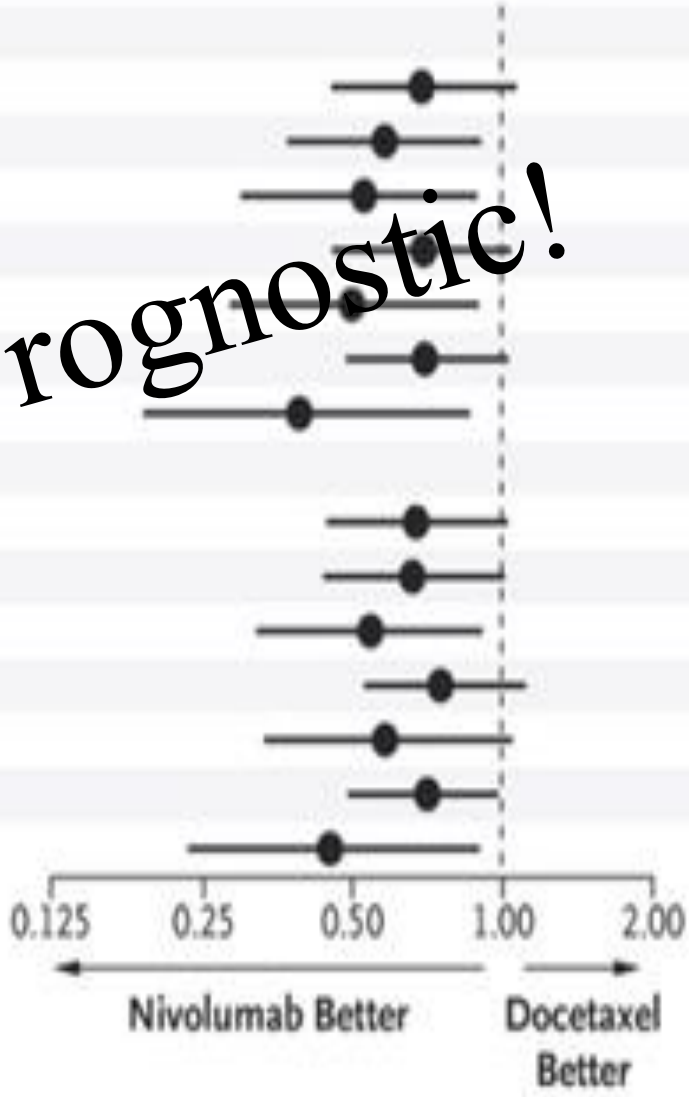
Metastatic Melanoma by PD-L1

All patients	**418**	**50/210**	**96/208**	**N.A.**	**10.84**	**(9.33, 12.09)**	**0.42**	**(0.30, 0.59)**
PD-L1 Status								
Positive	148	11/74	29/74	N.A.	12.39	(9.17, N.A.)	0.30	(0.15, 0.60)
Negative/Indeterminate	270	39/136	67/134	N.A.	10.22	(7.59, 11.83)	0.48	(0.32, 0.71)



Squamous NSCLC by PD-L1

PD-L1 Expression Level	Nivolumab	Docetaxel	Unstratified Hazard Ratio (95% CI)
	no. of patients		
Overall survival			
≥1%	63	56	0.69 (0.45–1.05)
<1%	54	52	0.58 (0.37–0.92)
≥5%	42	39	0.53 (0.31–0.89)
<5%	75	69	0.70 (0.47–1.02)
≥10%	36	33	0.50 (0.28–0.89)
<10%	81	75	0.70 (0.48–1.01)
Not quantifiable at baseline	18	29	0.39 (0.19–0.82)
Progression-free survival			
≥1%	63	56	0.67 (0.44–1.01)
<1%	54	52	0.66 (0.43–1.00)
≥5%	42	39	0.54 (0.32–0.90)
<5%	75	69	0.75 (0.52–1.08)
≥10%	36	33	0.58 (0.33–1.02)
<10%	81	75	0.70 (0.49–0.99)
Not quantifiable at baseline	18	29	0.45 (0.23–0.89)



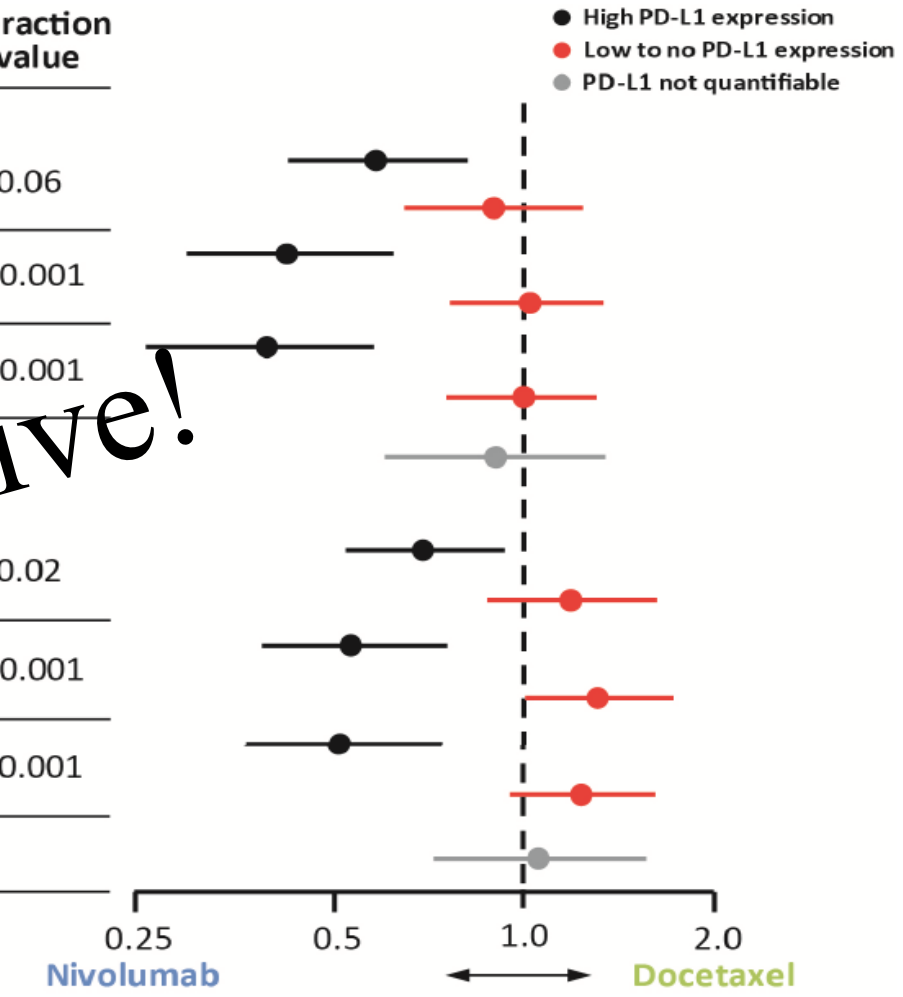
Only Prognostic!

Non-Squamous NSCLC by PD-L1

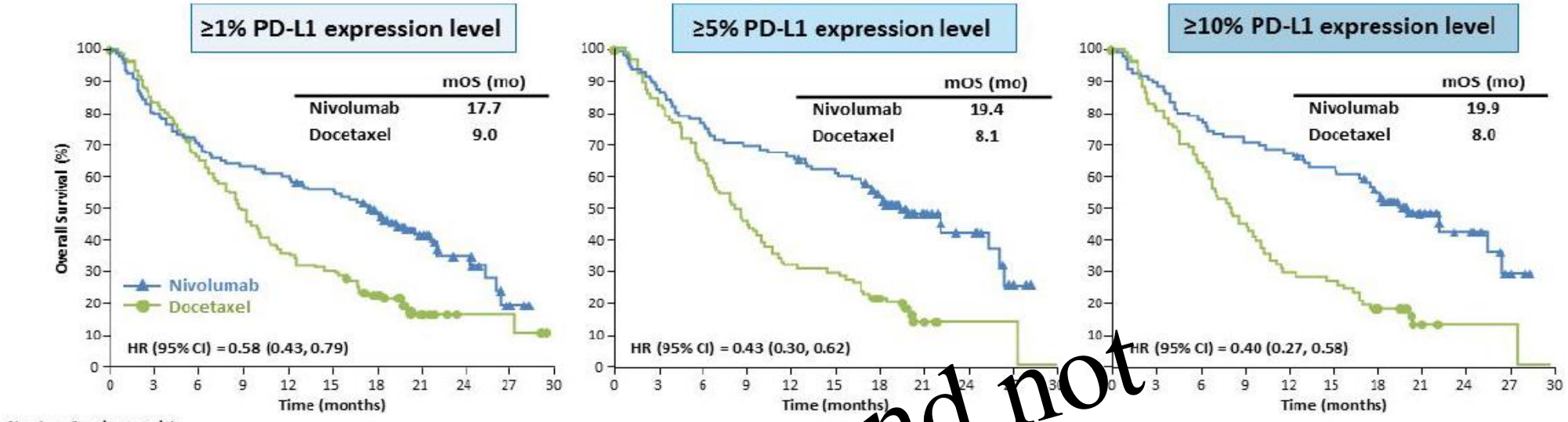
Figure S7. Plot of Overall Survival and Progression-free Survival Hazard Ratios by PD-L1 Expression at Baseline.

PD-L1expression level	Nivolumab n	Docetaxel n	Unstratified HR (95% CI)	Interaction *P*-value
OS				
≥1%	123	123	0.59 (0.43, 0.82)	0.06
<1%	108	101	0.90 (0.66, 1.24)	
≥5%	95	86	0.43 (0.30, 0.63)	<0.001
<5%	136	138	1.01 (0.77, 1.34)	
≥10%	86	79	0.40 (0.26, 0.59)	<0.001
<10%	145	145	1.00 (0.76, 1.31)	
Not quantifiable at baseline	61	66	0.91 (0.61, 1.35)	
PFS				
≥1%	123	123	0.70 (0.53, 0.94)	0.02
<1%	108	101	1.19 (0.88, 1.61)	
≥5%	95	86	0.54 (0.39, 0.76)	<0.001
<5%	136	138	1.31 (1.01, 1.71)	
≥10%	86	79	0.52 (0.37, 0.75)	<0.001
<10%	145	145	1.24 (0.96, 1.61)	
Not quantifiable at baseline	61	66	1.06 (0.73, 1.56)	

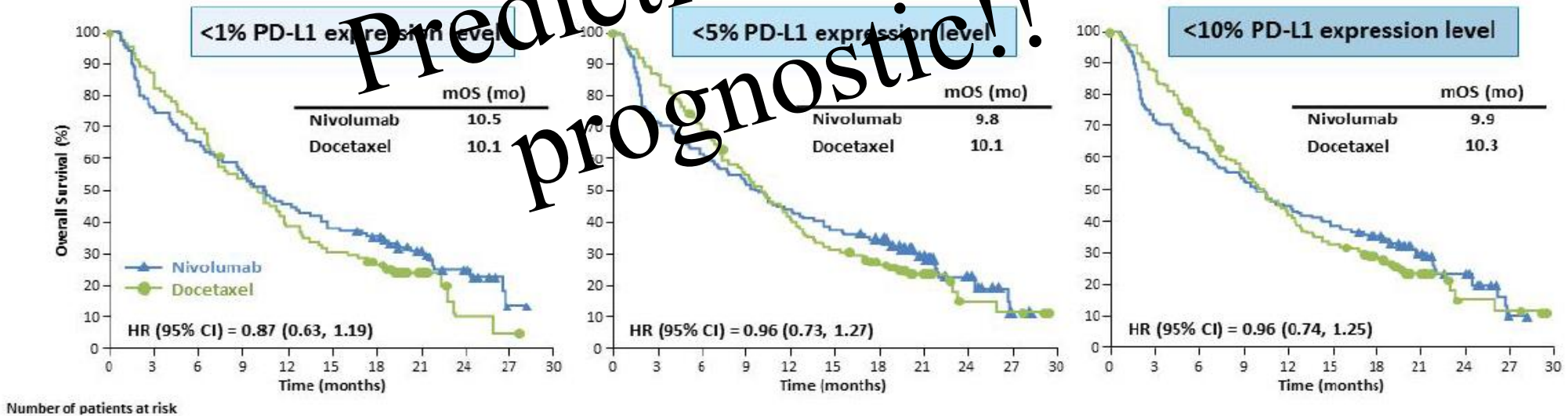
Predictive!



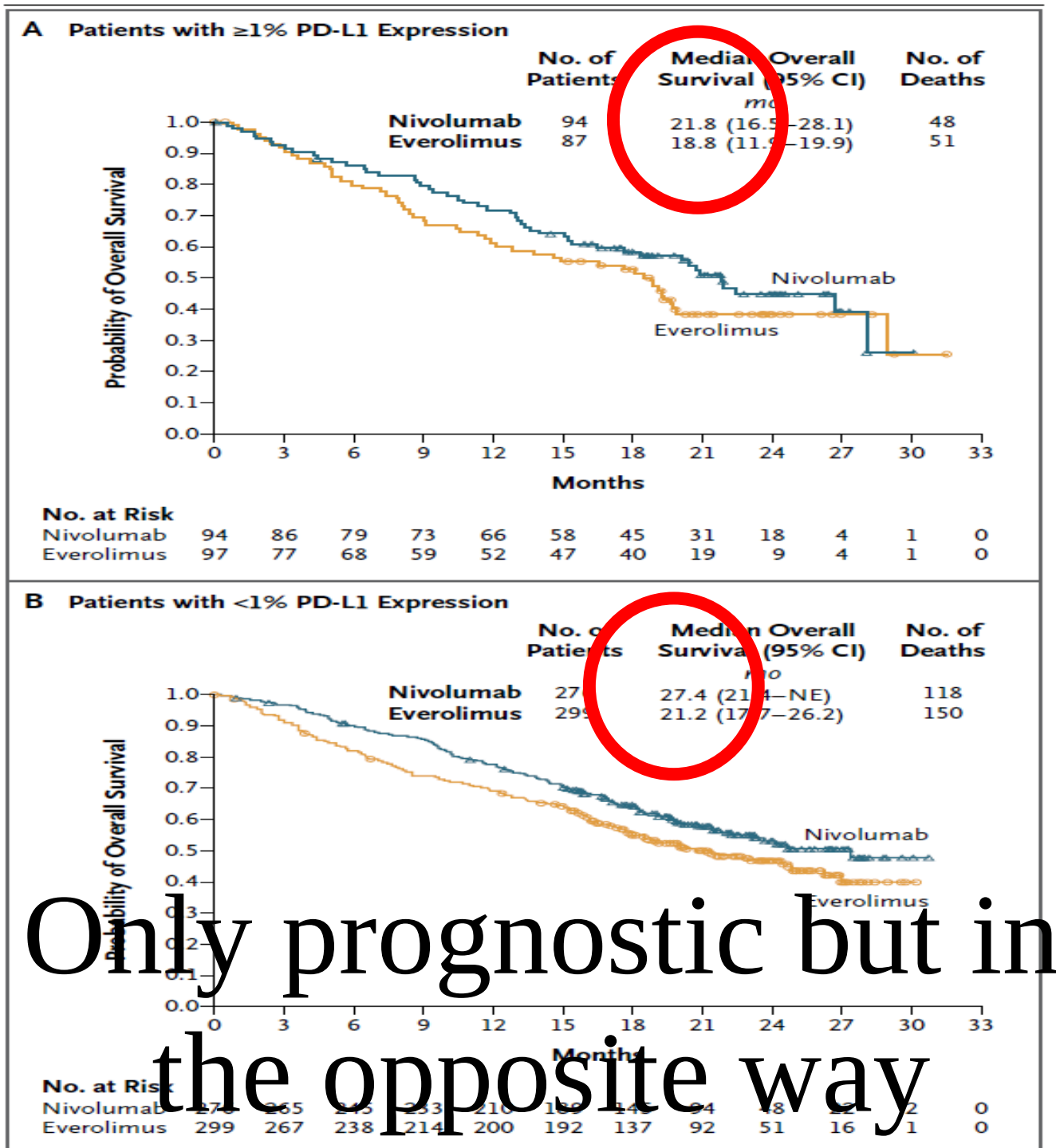
Non-Squamous NSCLC - OS



Predictive and not prognostic!!

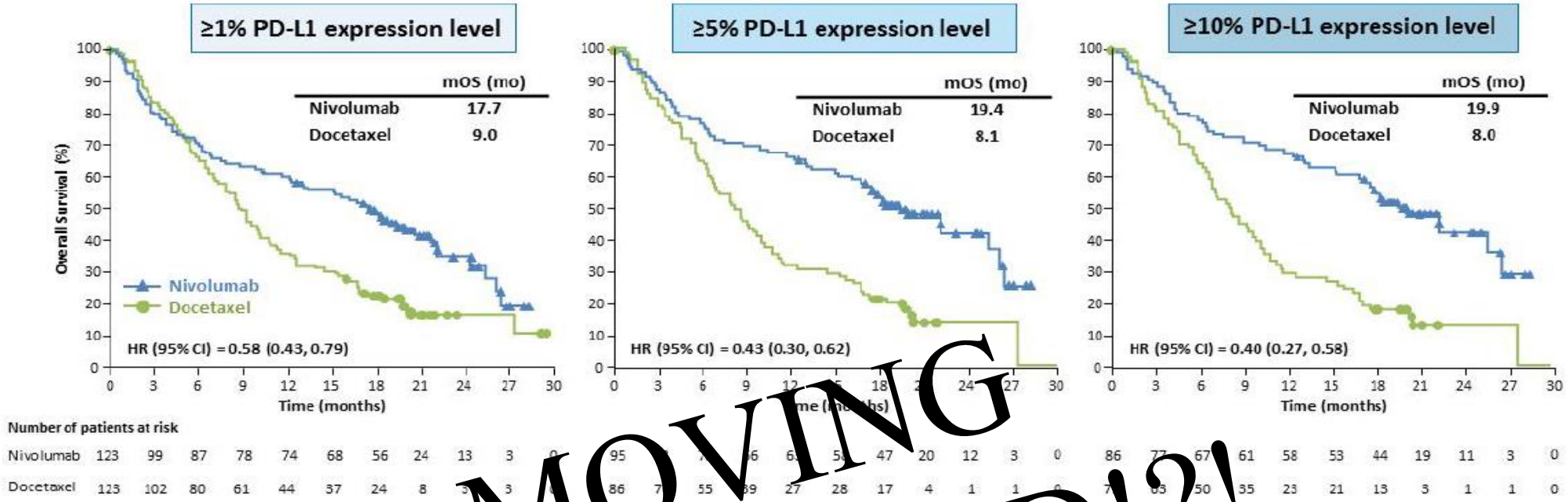


Kidney cancer

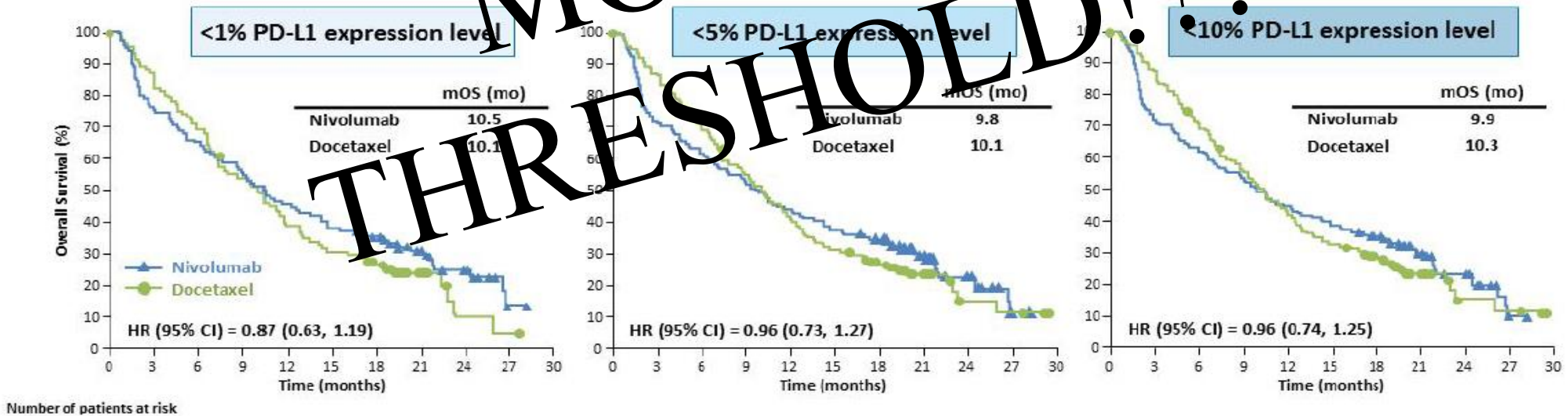


Only prognostic but in the opposite way

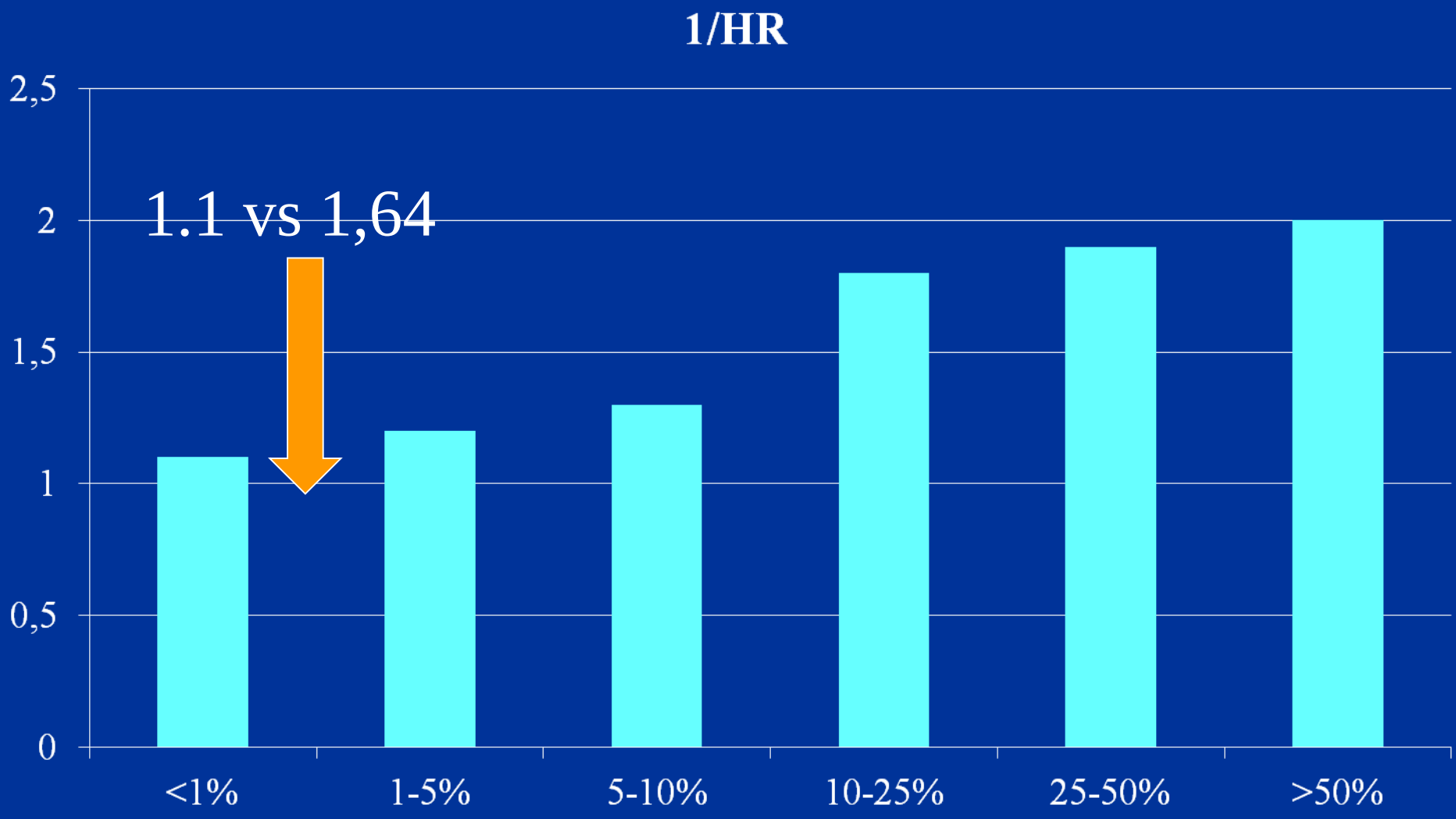
SUBGROUP ANALYSES?



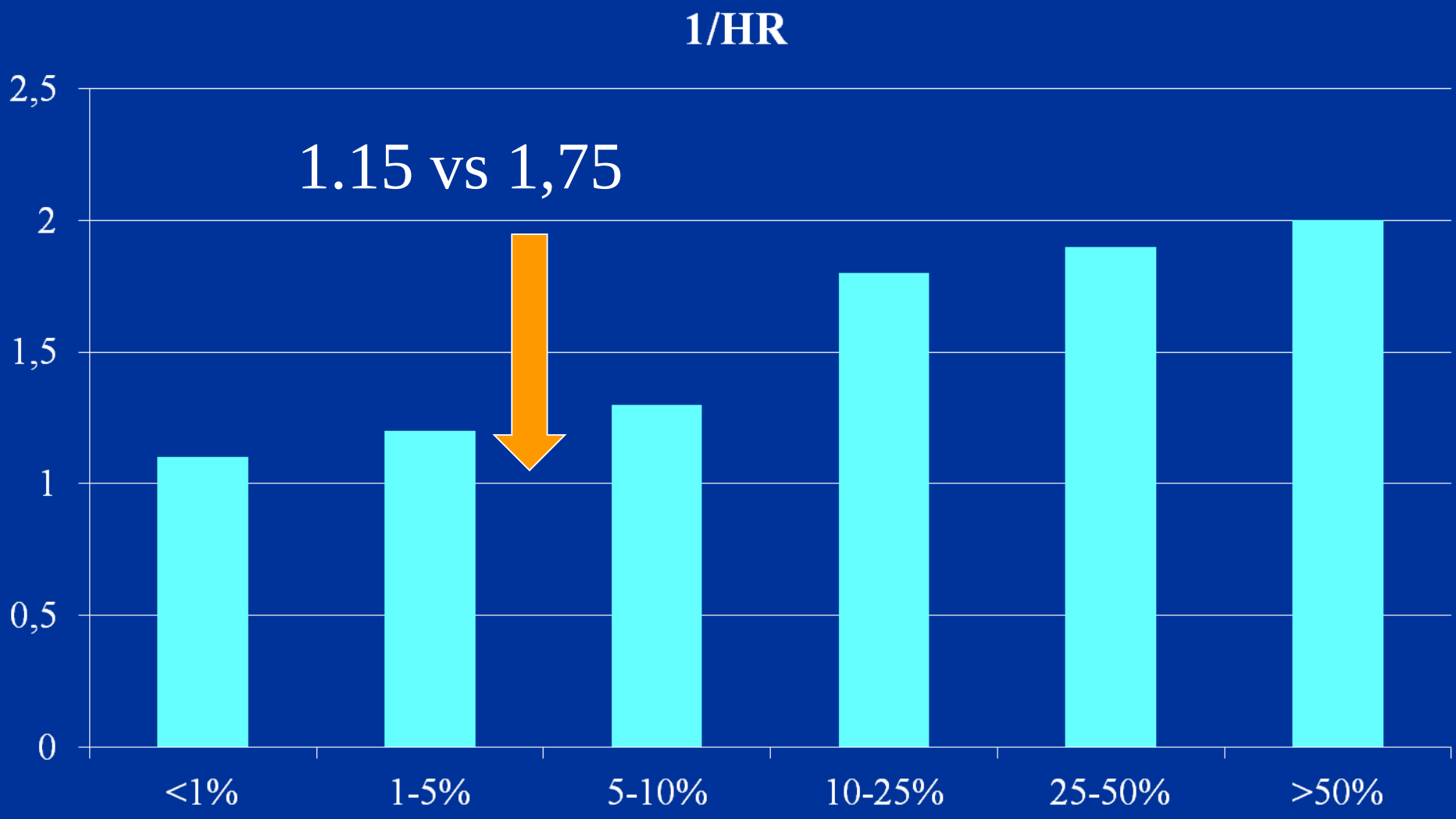
MOVING THRESHOLD!?!



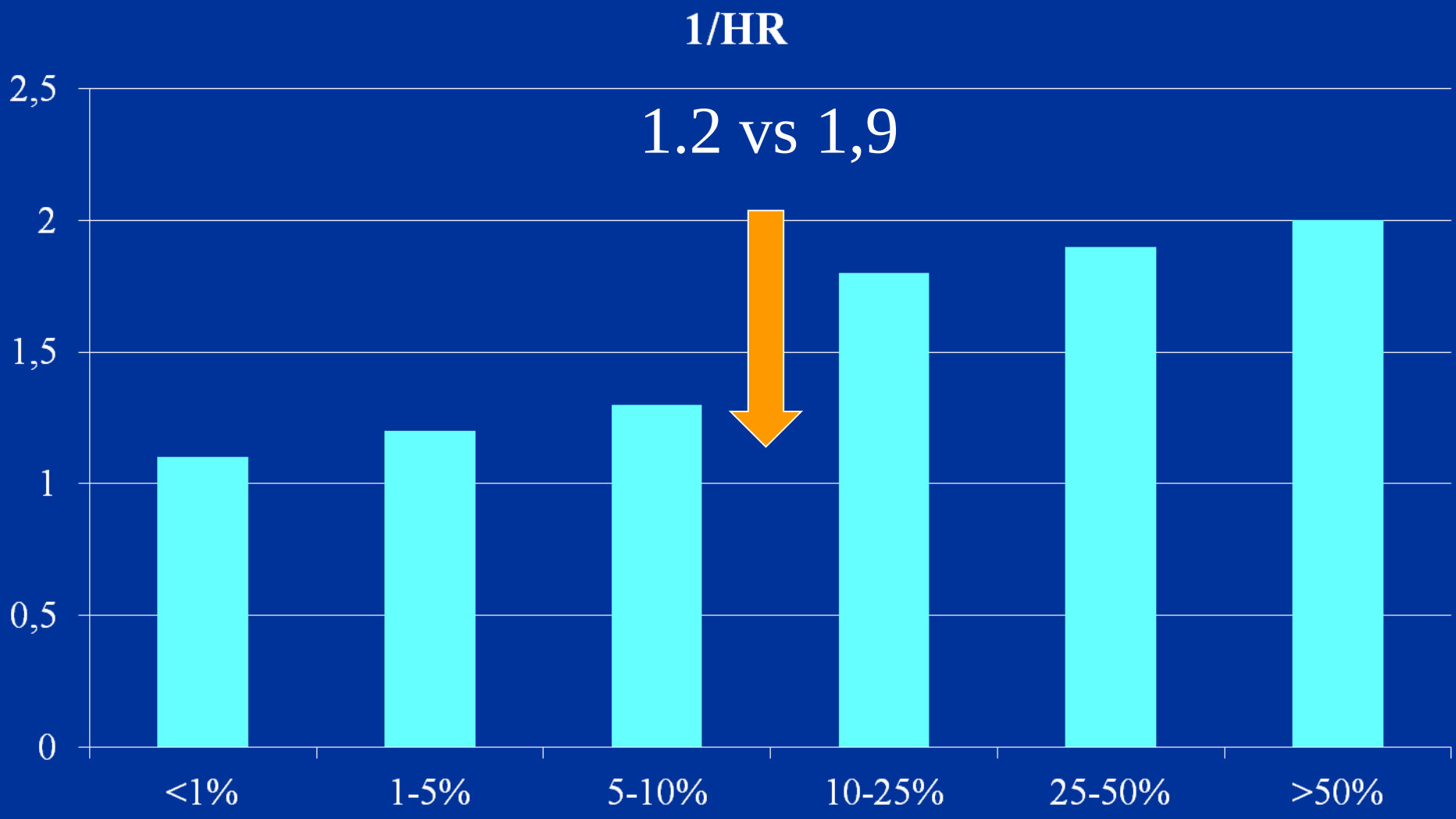
MOVING THRESHOLD



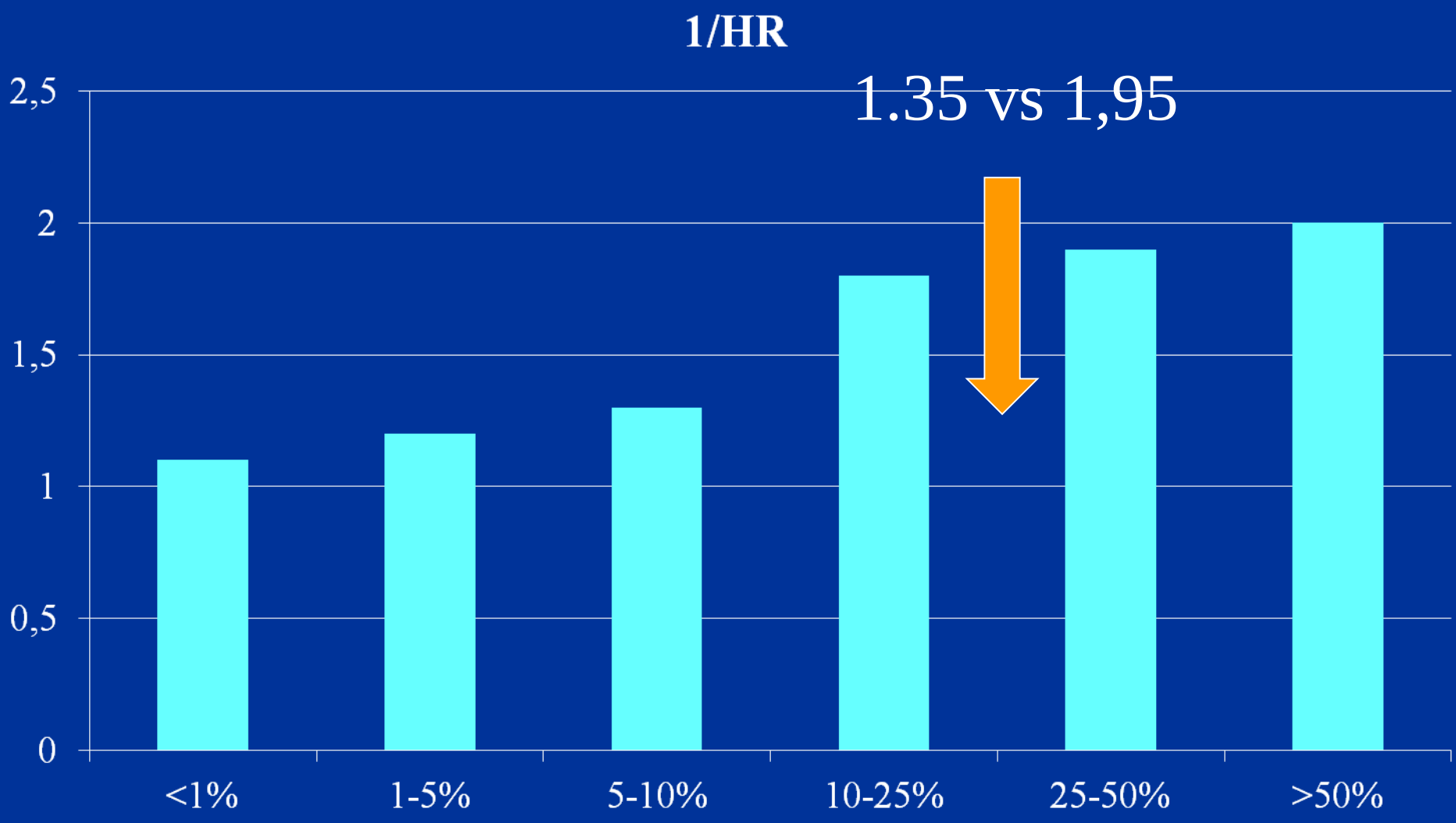
MOVING THRESHOLD



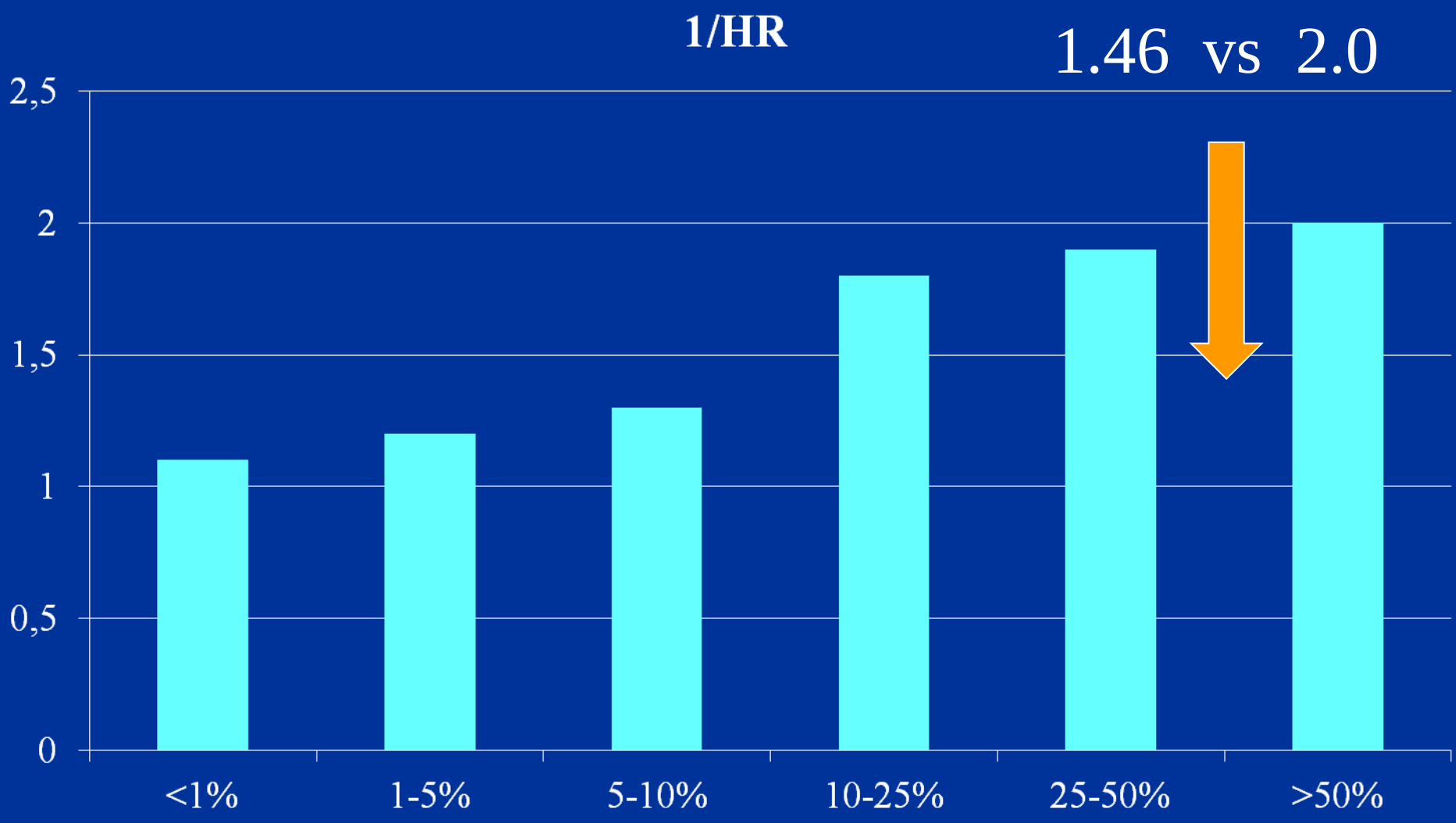
MOVING THRESHOLD



MOVING THRESHOLD



MOVING THRESHOLD



MOVING THRESHOLD

% Expression	1/HR	1/HR Below/=	1/HR Above
<1%	1.1	1.1	1.64
1-5%	1.2	1.15	1.75
5-10%	1,3	1.2	1.9
10-25%	1,8	1.35	1.95
25-50%	1,9	1.46	2.0
>50%	2,0	1.55	-

MOVING THRESHOLD

- The moving threshold technique is fine when the dependent variable is binary (= ROC curve), to identify the cut-off with the most desirable combination of sensitivity and specificity
- When the dependent variable is a survival (OS, PFS) it is not efficient to clarify the relationship between covariate values and HR and should not be used.

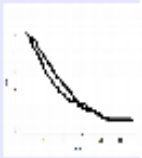
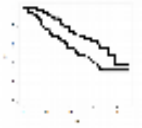
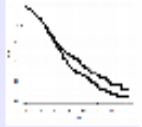
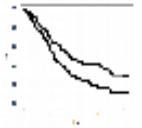
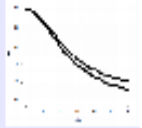
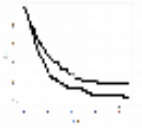
Factors predicting the efficacy of inmunotherapy

- *Moving threshold*

- Heterogeneity in Hazard Ratios

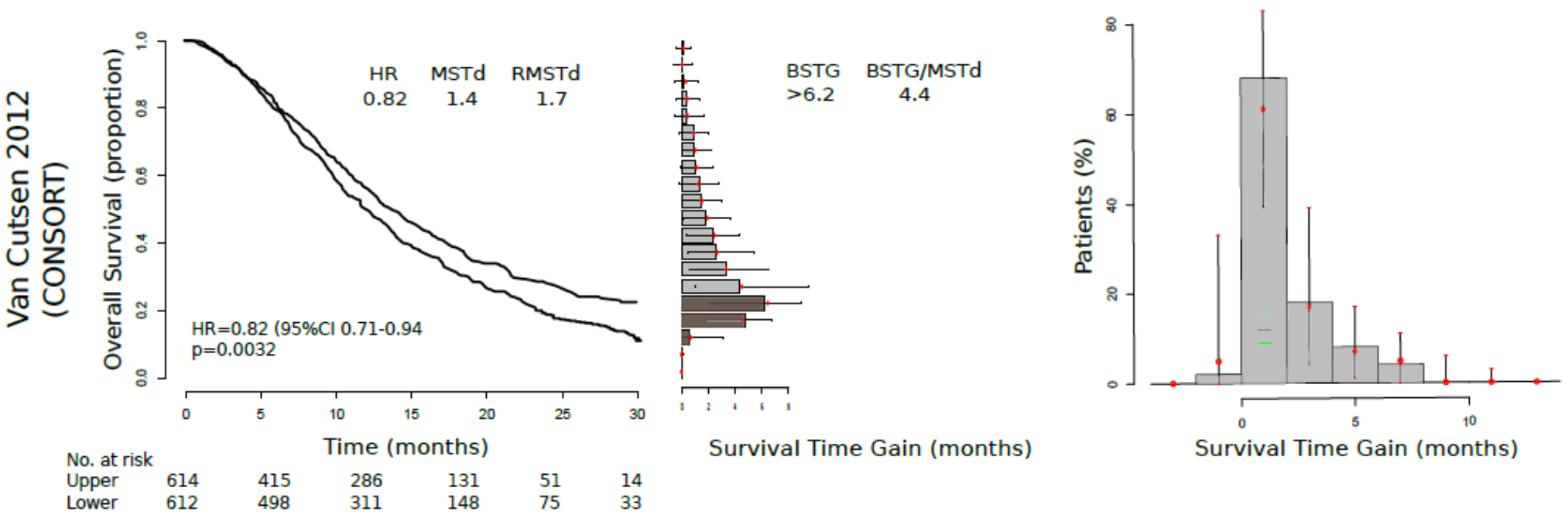
Is it appropriate to compare hazard ratios in different subgroups when the Hazard Ratio does not appear to be a good descriptor of the effect of the drug?

Summary of studies' characteristics

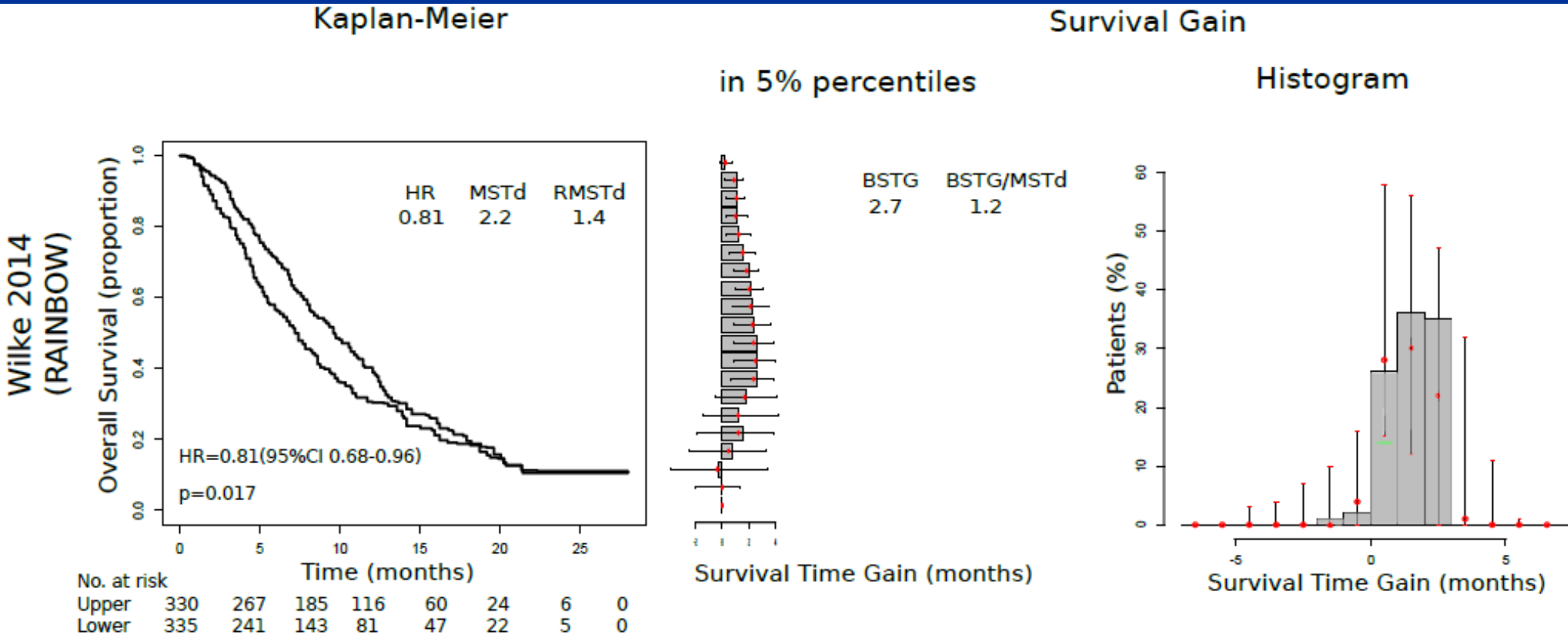
Study	Cancer Site	Treatments	End Point	Patient number	MST* [95%CI]	HR [95%CI]	Follow up* (Restrict)	MSTd *	RMSTd *	MSTd/ RMSTd	BSTG* [95%CI]	BSTG/ RMSTd	BSTG/ MSTd	KM (OS) shape
Wilke 2014 (RAINBOW)	Gastric	ramucirumab + paclitaxel placebo +paclitaxel	OS	330 335	9.6 [8.5-10.8] 7.4 [6.3-8.4]	0.81 [0.68-0.92]	27 (24)	2.2	1.4	1.6	2.7 [1.4-4.6]	1.9	1.2	
Gligorov 2014 (IMELDA)	Breast	capecitabine+ bevacizumab bevacizumab alone	OS	91 94	39 [32.3-n.r.] 23.7 [18.5-31.7]	0.43 [0.26-0.69]	43 (39)	15.3	6.6	2.3	14.4 [7.8-22.4]	2.2	0.9	
Heinemann 2014 (FIRE3)	Colon	FOLFIRI+ cetuximab FOLFIRI+ bevacizumab	OS	297 295	28.7 [24.0-33.6] 25.0 [22.7-27.6]	0.77 [0.62-0.96]	72 (70)	3.7	5.2	0.71	15.8 [7-22.1]	3.0	4.3	
Brahmer 2015 (NCT01642004)	Lung	nivolumab docetaxel	OS	135 137	9.2 [7.3-13.3] 6.0 [5.1-7.3]	0.59 [0.44-0.79]	24 (21)	3.2	3.3	0.97	>10.4 [6.5-13.1]	3.1	3.2	
Van Cutsen 2012 (CONSORT)	Colon	FOLFIRI+ aflibercept FOLFIRI+ placebo	OS	612 614	13.5 [12.5-14.9] 12.1 [11.1-13.1]	0.82 [0.71-0-94]	36 (30)	1.4	1.7	0.82	>6.2 [3.2-9.9]	3.6	4.4	
Hodi 2010 (NCT00094653)	Melanoma	ipilimumab alone glycoprotein 100 alone	OS	137 136	10.1 [8.0-13.8] 6.4 [5.5-8.7]	0.66 [0.51-0.87]	55 (44)	3.7	5.6	0.66	>26.7 [13.4-32.1]	4.8	7.2	

OS: Overall Survival; MST: Median Survival Time; CI: Confidence Interval; HR: Hazard Ratio; MSTd: difference of Median Survival Times; RMSTd: difference of Restricted Mean Survival Time; BSTG: Best (mean restricted) Survival Time in 5% percentiles of patients; KM(OS): Kaplan Meier curve for Overall Survival.
*: values in months

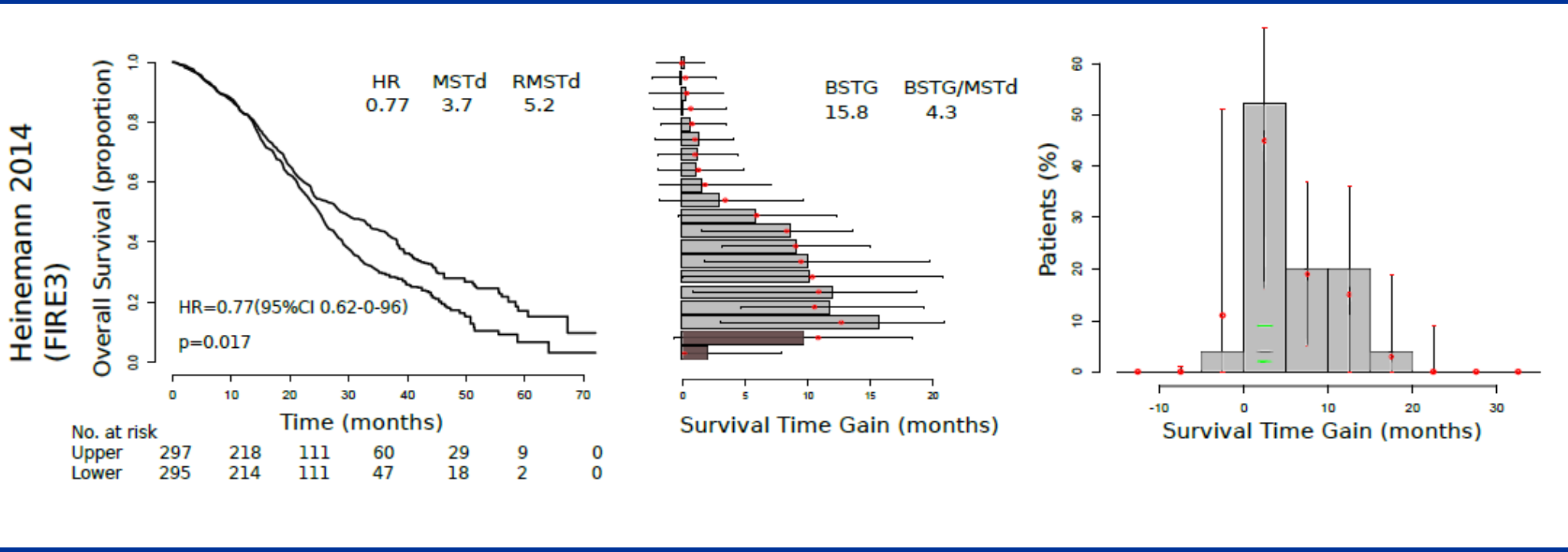
CRC: HR appropriate



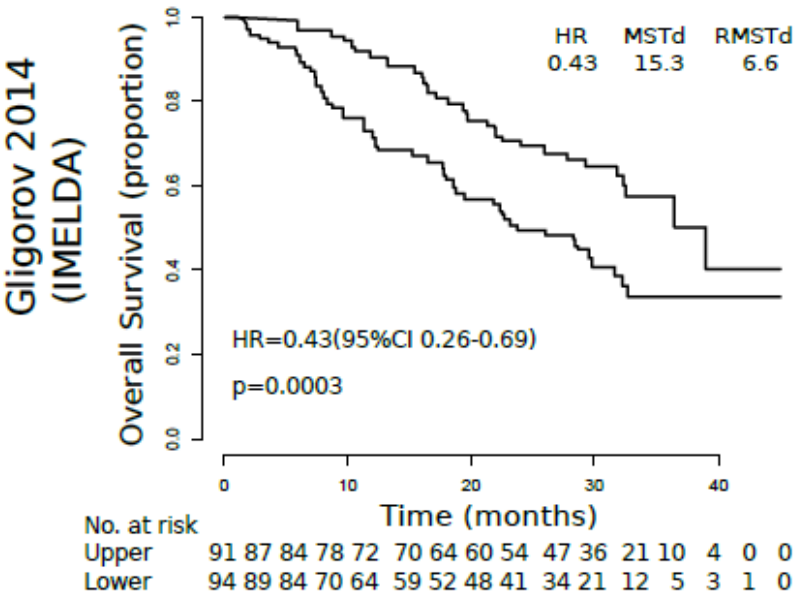
Gastric c.: Small for many

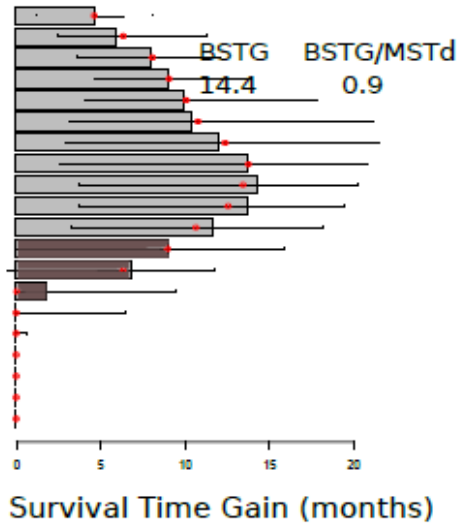


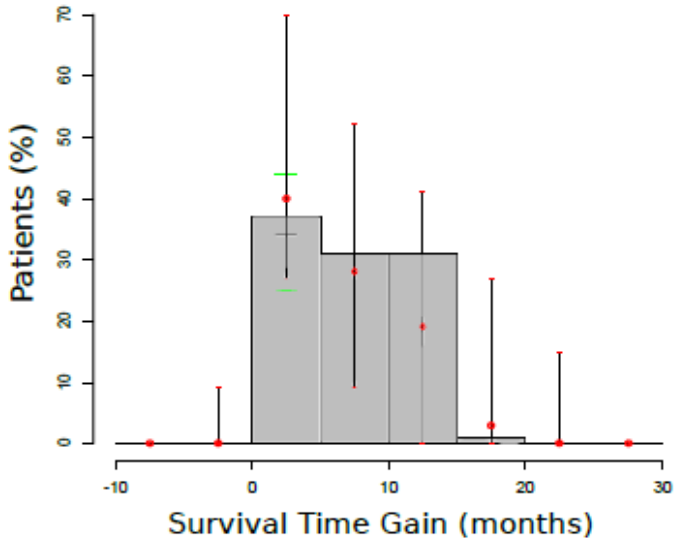
CRC: Heterogeneous benefits for 50% of patients



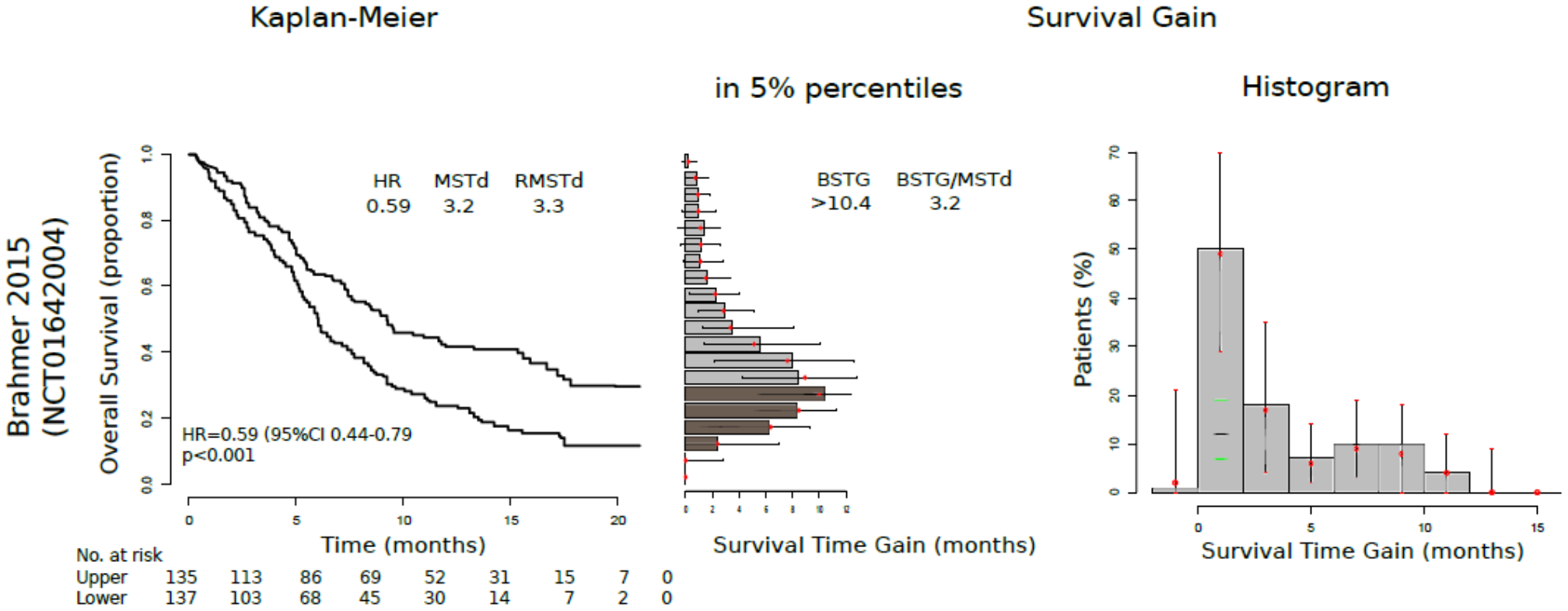
Breast c.: Variable for >50%



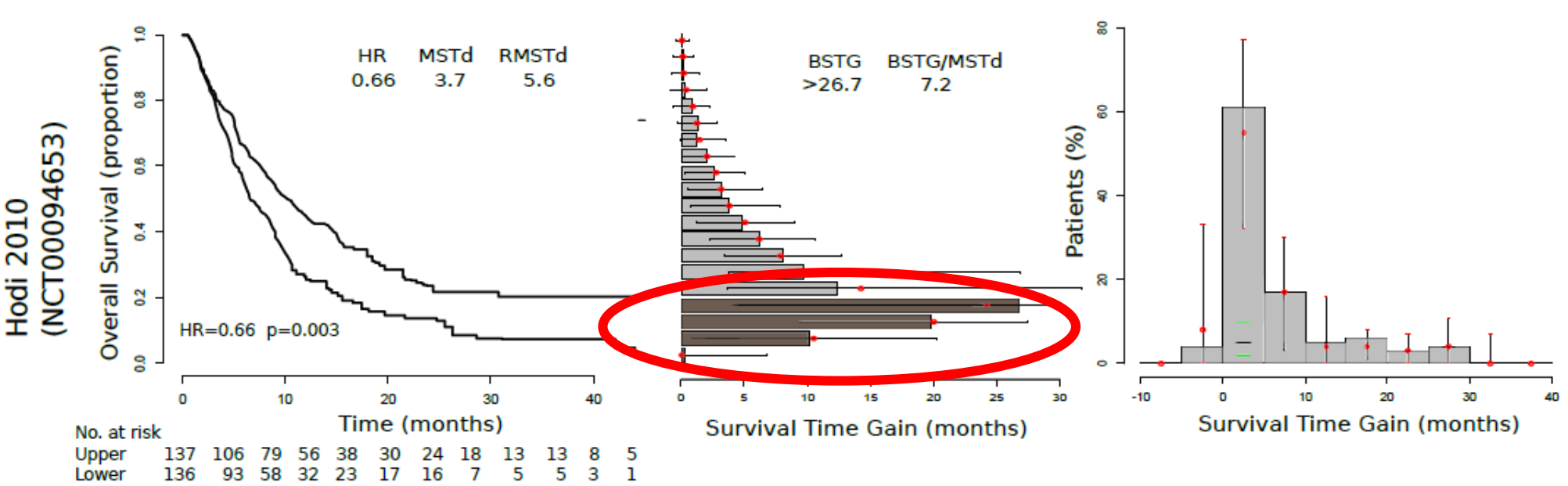




NSCLC – Too early to tell



Melanoma: Large for few



New type of subgroup analysis?

New types of subgroup analyses:

- Early Failures vs average PFS (Immunotherapy in NSCLC)

- Long term survivors vs average survivors (Breast c., Melanoma, others?)

Conclusions on subgroup analyses:

- Stick to the standard methods
- Use them wisely
- Explore new approaches with caution