Efficacy Claims and Subset Analyses in Phase III: Some thoughts and examples

Kevin J Carroll

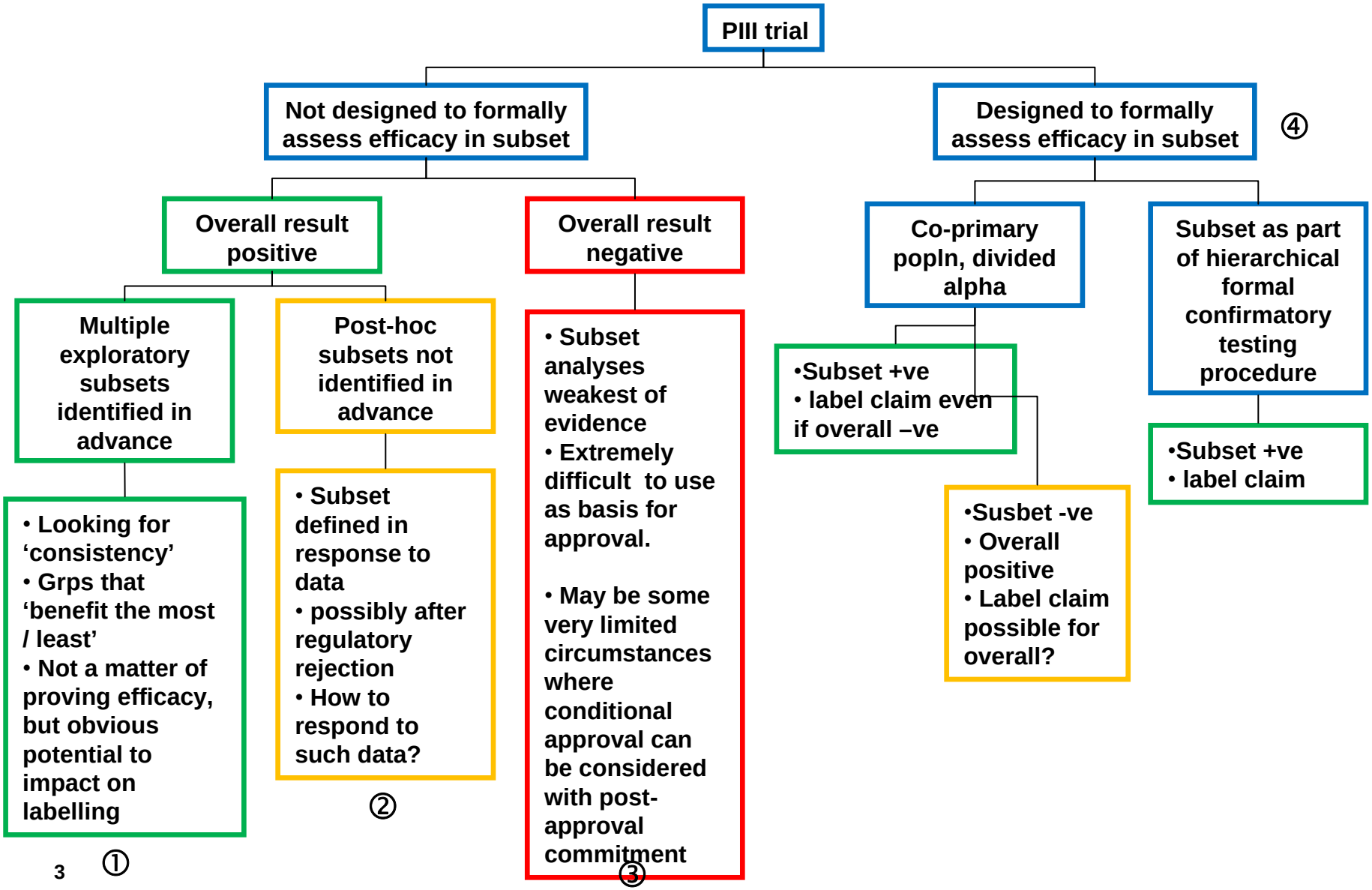
My personal views, not AZ's

Why do we do them?

- Fundamentally, because we can, and we feel compelled to do so.
- Yet, we have no agreed framework for the conduct and interpretation of such analyses, nor for their potential implications on product labelling and drug approval.
- Of interest to consider the use and utility of subgroup analyses in confirmatory Phase III trials where such analyses are intended:

1. To provide an assessment of internal 'consistency' in an trial with an overall positive outcome.
2. By design to formally assess a hypothesis of efficacy in a predefined subset.
3. To salvage a negative trial.

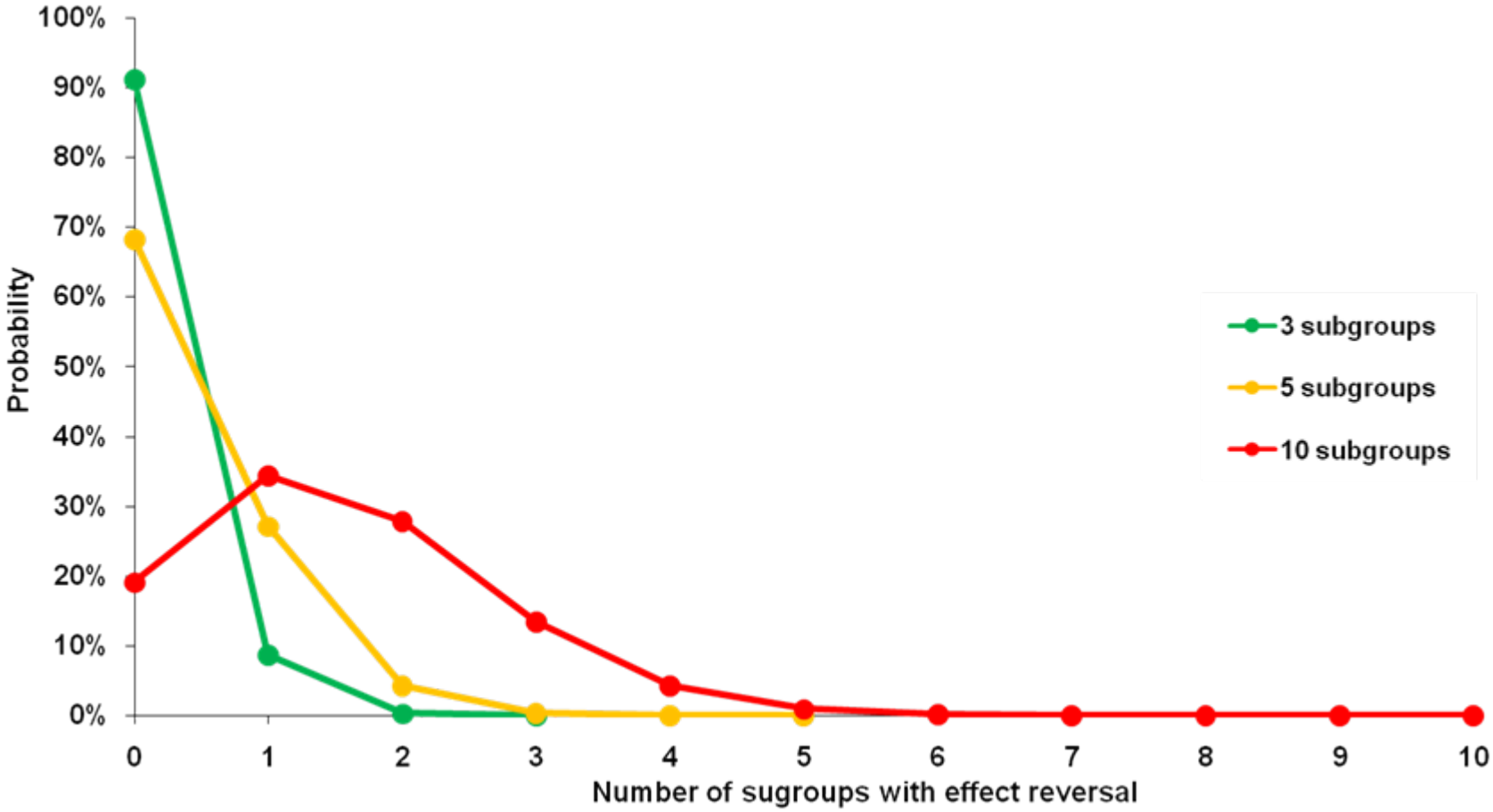
The persuasiveness of subset analyses in confirmatory trials



1. Pre-planned, descriptive subgroup analyses to assess internal 'consistency' in a positive PIII trial

- What do we mean by this? What is 'consistency'? How is this assessed statistically? Consistency within subset clusters or overall? What is the role of interaction tests?
- Genuine attempt to check that no group is 'disadvantaged'? Or to look for the group that 'benefits most'?
- Inevitable that extremes will be observed in subgroup analyses even when the treatment effect is truly consistent.
- Assumption that treatment effects truly different in different groups has fundamental implications for design and analysis.

“Effect Reversals”, where the treatment effect is positive overall but numerically negative in some subgroups are to be expected in Phase III subgroup analyses



Senn, S and Harrell, F. On wisdom after the Event. *J Clin Epidemiol*, 1997. 50(7):749-751.

2. Post-hoc subgroup analyses in a positive PIII trial

- Why would this ever be needed?
- At request of regulatory to narrow down what otherwise might be considered to be too broad an approval?
- By the Sponsor to improve the prospects of reimbursement?
- By the Sponsor in the face of a negative regulatory opinion despite meeting the primary endpoint of the trial?

3. Subgroup findings in a negative PIII trial

- Hypothesis generating. Extremely unlikely to support a claim of efficacy.
- Possible circumstances where a claim *might* be supported:
  a) **Strong, a-priori biologic plausibility linked to MOA.**
  b) **Statistical replication.**
  c) **External, independent randomised evidence of an effect in subgroup from drugs of similar class.**
  d) **Rare / serious disease with high morbidity / mortality, high unmet need with no therapeutic options.**
- If a)-d) are all met, is there some possibility for conditional approval with requirement for prospective confirmatory trial?

4. Confirmatory trials designed to prospectively evaluate a subgroup

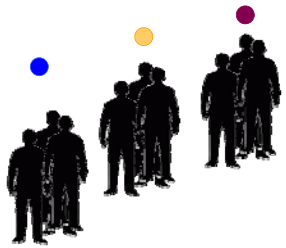
- When and why might you consider such a pre-specification strategy?
- Some examples of pre-specified subgroup approaches vs post-hoc defined subsets

When and why might you consider a subgroup pre-specification strategy?

1. When the biology tells you too – e.g herceptin
2. When you have seen encouraging ‘signals’ in PII suggesting a subgroup may benefit to a greater extent.
3. When you have seen external evidence from a similar class of agents suggesting an enhanced benefit in a clinical or biologically defined subgroup.

- 1 is clearly most secure.
- 2 might be pursued as a ‘no lose’ strategy with the overall population as co-primary.
- 3 may encourage toward a greater risk-taking strategy to accelerate development.

Considerations very closely linked to Personalised Medicine strategies



The right treatment at the right dose for the right person at the right time for the right outcome

Expectations raised for:

Safer, 'More Effective' Drugs

No more 'one-size-fits-all' drugs. New drugs will be safe and effective for specific populations defined by –omic characteristics.

Faster Developments at Less Cost and Less risk

Speedier clinical trials based on high 'responder' populations, higher R&D success rates and lower overall development costs.

Cost-Effective Healthcare

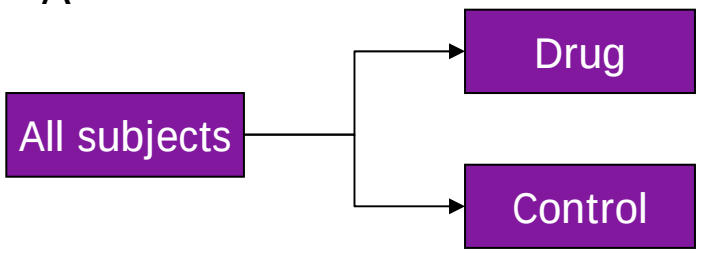
Reduced costs, due to avoidance of futile treatments in large populations who do not benefit, reimbursement challenges reduced

But it is not that simple

- How sure are we the defined subgroup will benefit to a greater extent?
- How have we defined the subgroup and do we know that non-subgroup patients will not benefit?
- If biologically defined, is the profile of the subgroup reproducible?
- Do we have a reliable test with acceptable specificity and sensitivity to identity the subgroup? A poor test can seriously impact the value of a pre-specified subgroup approach.
- How sure are we the characteristic(s) defining the subgroup is predictive for an enhanced treatment effect as opposed to merely being prognostic marker of disease?
- The real danger is we too often think we know the answer before we have the data...

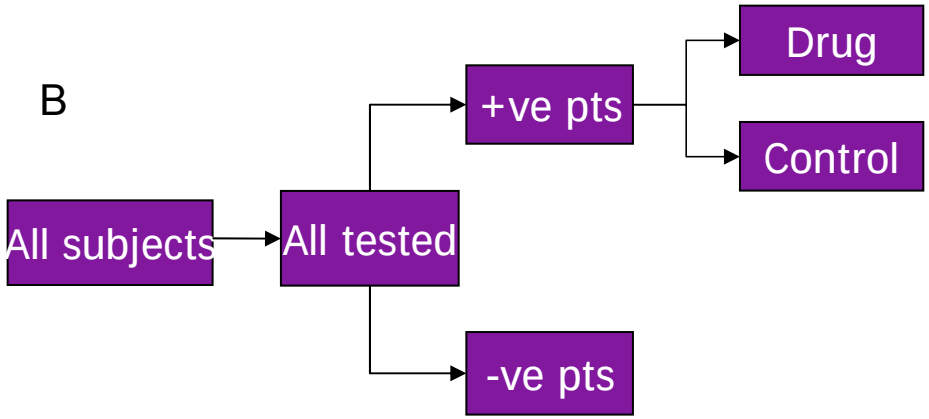
Which is the best approach in Phase III?

A



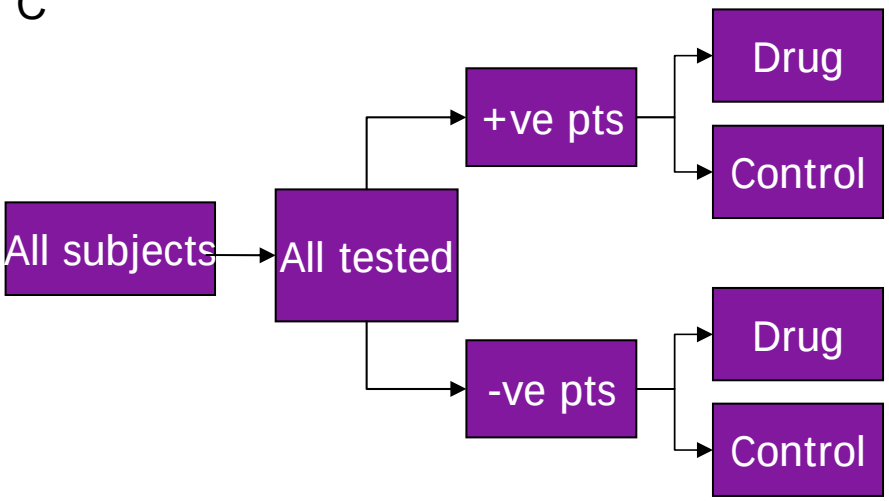
Unselected Design

B



Select +ve pts only; labelling in subgroup

C



Select +ve and –ve pts;
Further assess the predictive value of the subgroup

labelling in *either* the overall population *or* in subgroup patients only

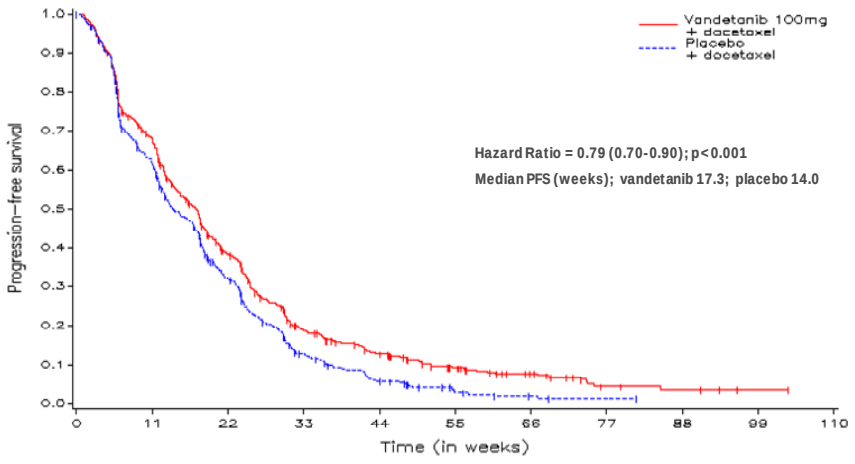
Some Examples

Confirmatory trials designed to prospectively evaluate a specific subgroup

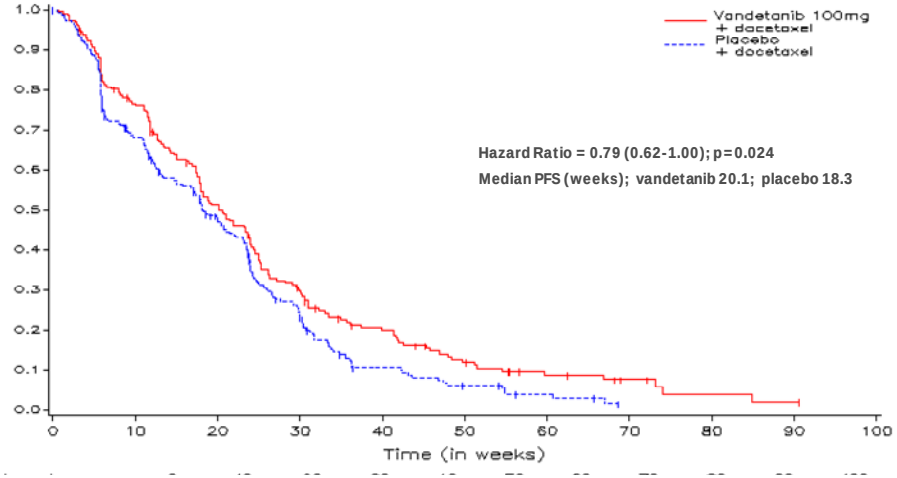
Vandetanib in advanced NSCLC

- Suggestion based on subgroup analysis in PII that females may achieve better effect – similar observations seen with other EGRF TKIs
- Phase III Primary end-point = Progression-Free Survival (PFS)
- Co-primary analysis population – Females
  - critical significance level for 'overall' and 'females' = 0.025

Overall population

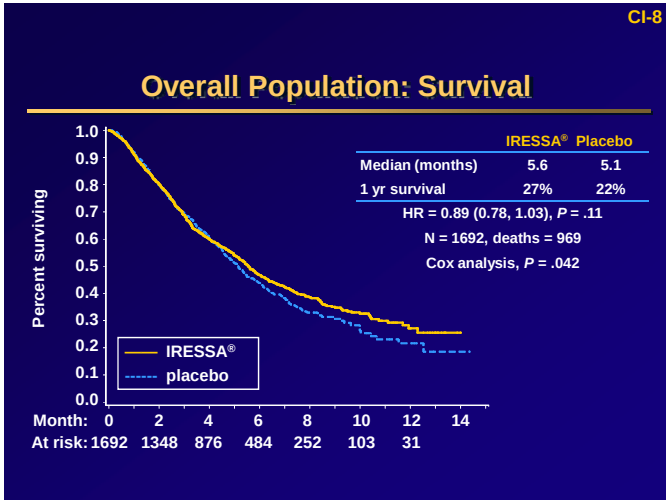


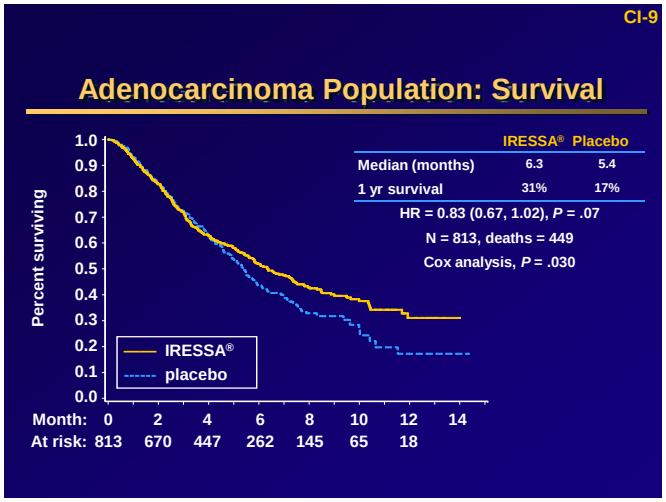
Females only population



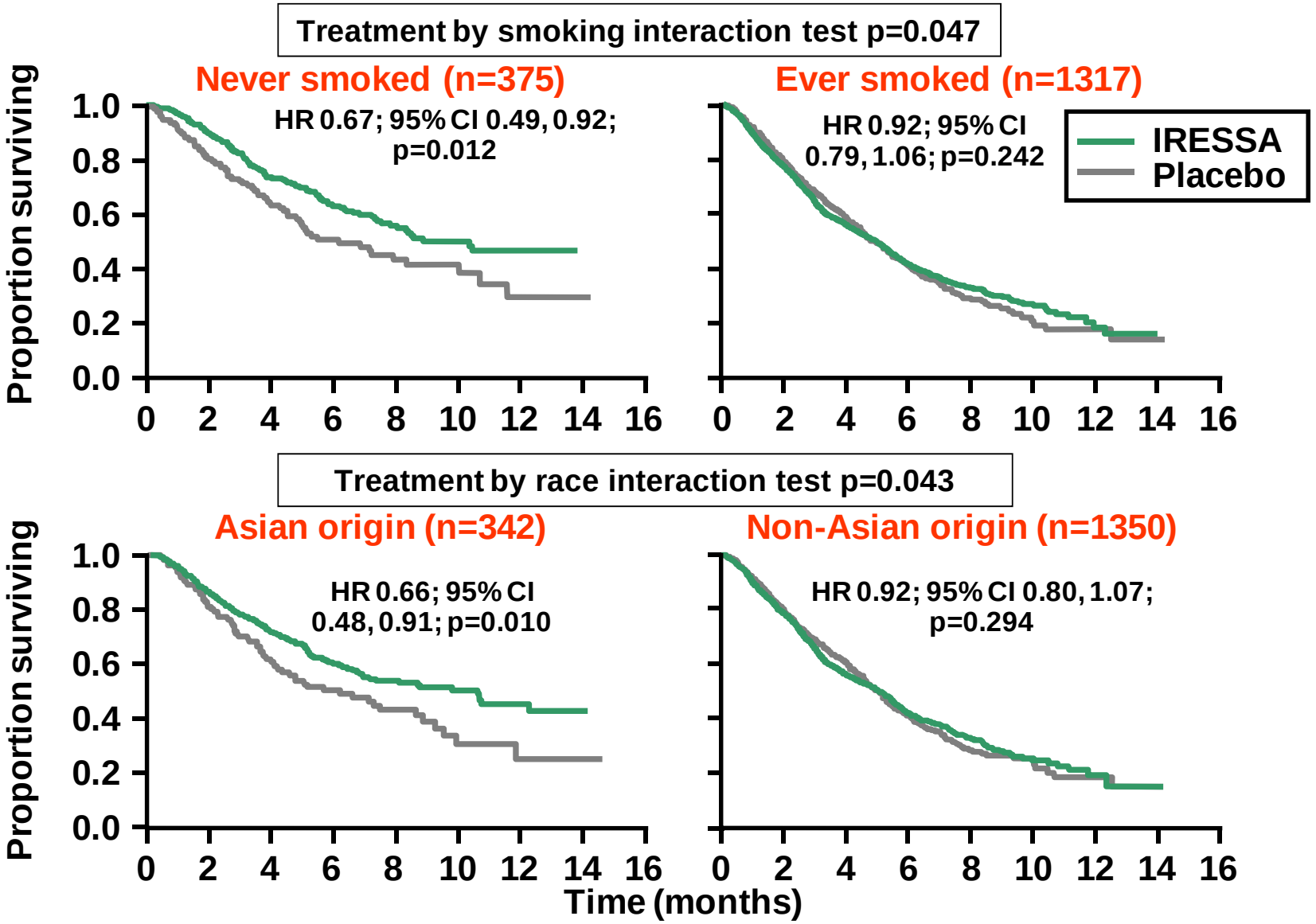
Iressa vs BSC in refractory NSCLC: ISEL Study

- Early PII data suggested variability in response to Iressa with females, never smokers and Asian ethnicity having higher response rates
  - Stratified log rank test primary, Cox regression supportive
- Primary endpoint overall survival
- Overall and Adenocarinoma populations co-primary
  - At least 900 deaths for 90% power
  - Hochberg procedure to control Type I error
- Several pre-planned subgroup analyses defined by clinical and biologic factors



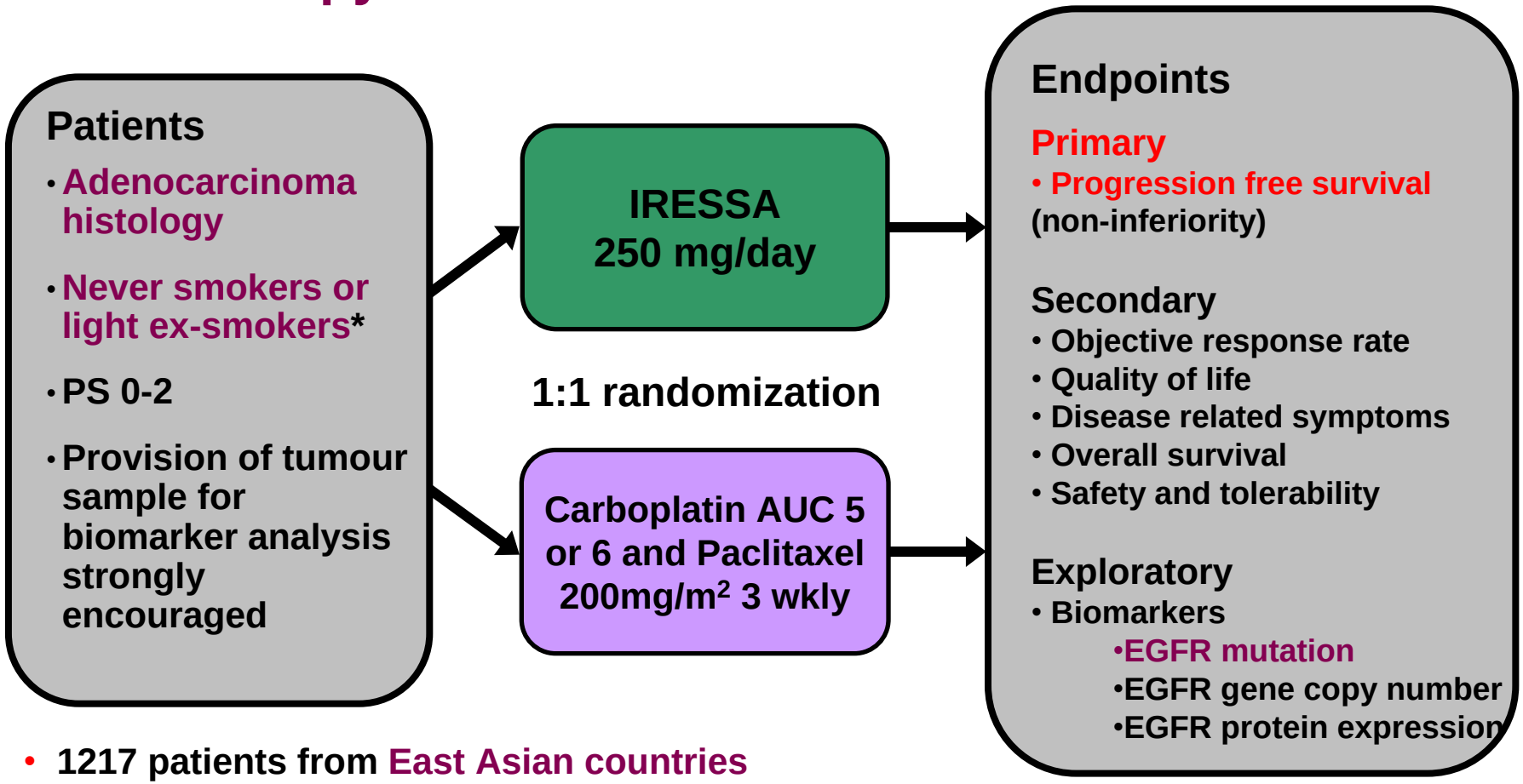


ISEL subgroups by smoking status and histology



Thatcher 2005, Chang 2006

IPASS: Phase III study of IRESSA versus doublet chemotherapy in first line NSCLC



- 1217 patients from East Asian countries

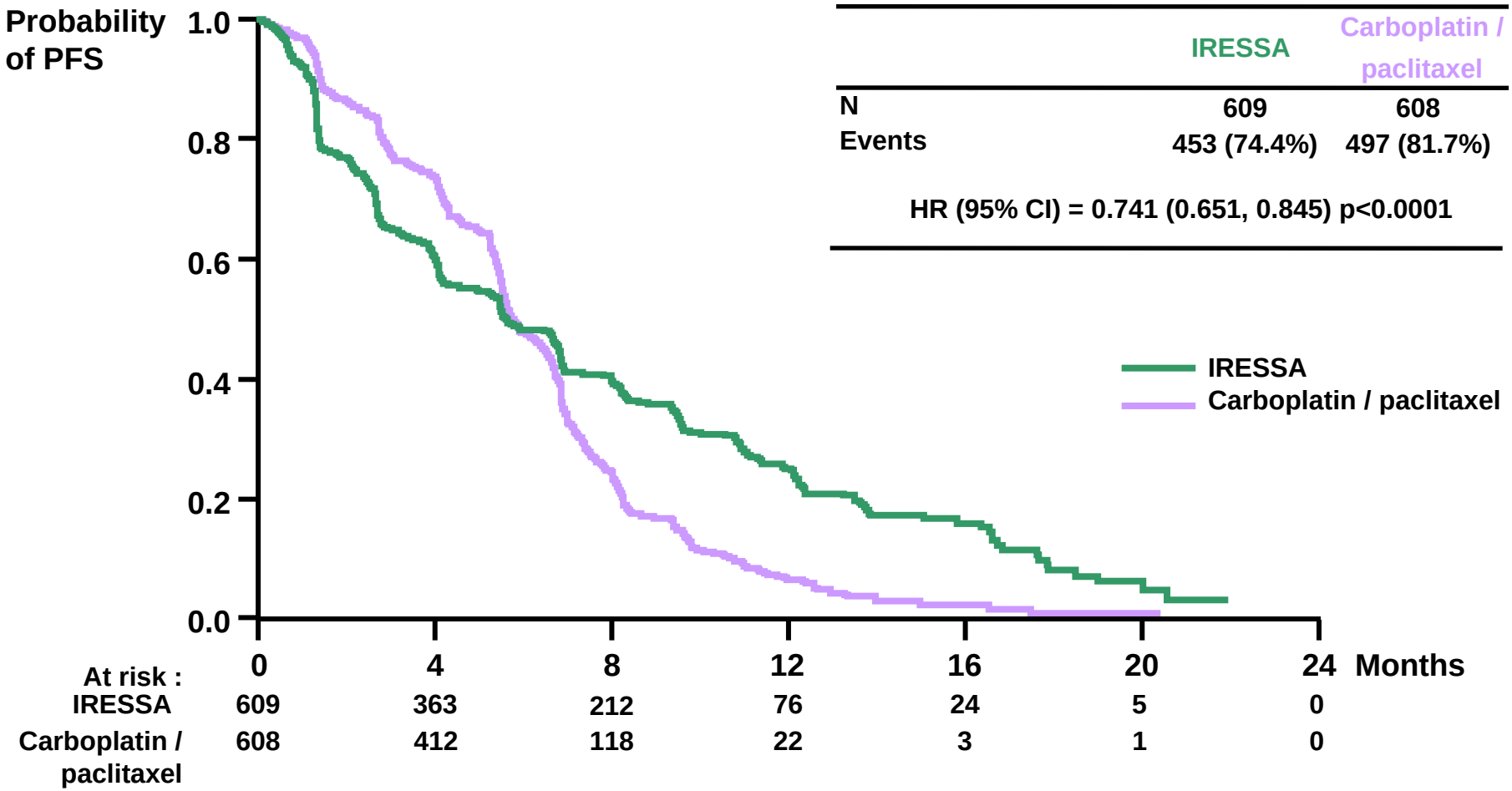
*Never smokers:<100 cigarettes in lifetime; light ex-smokers: stopped ≥15 years ago and smoked ≤ 10 pack yrs

Carboplatin/paclitaxel was offered to IRESSA patients upon progression

PS, performance status; EGFR, epidermal growth factor receptor

Mok 2009

IPASS: Primary Endpoint of PFS exceeded – Superior not just non-inferior



	IRESSA	Carboplatin / paclitaxel
N	609	608
Events	453 (74.4%)	497 (81.7%)

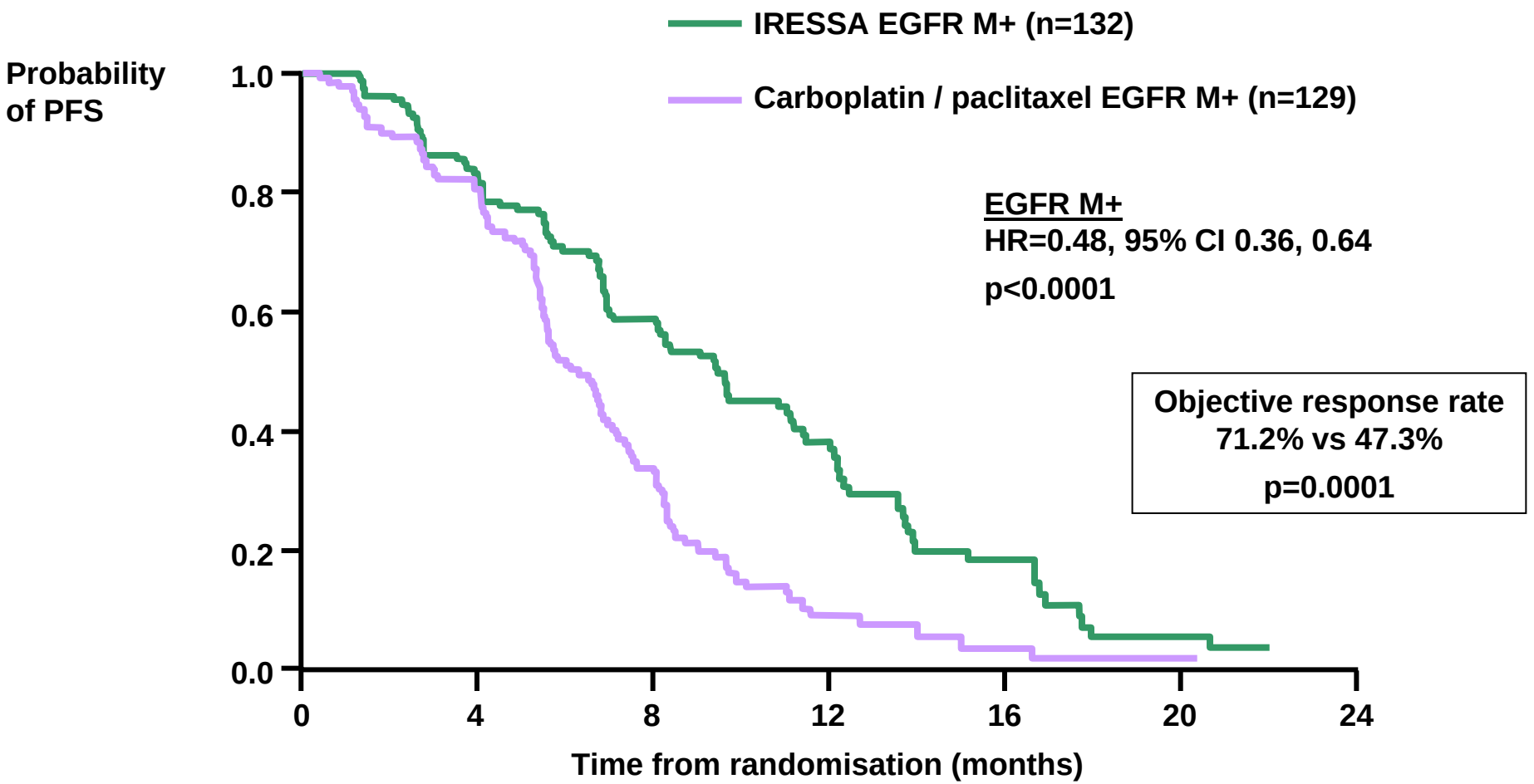
HR (95% CI) = 0.741 (0.651, 0.845) p<0.0001

IRESSA
Carboplatin / paclitaxel

Objective response rate 43% vs 32% p=0.0001

Mok 2009

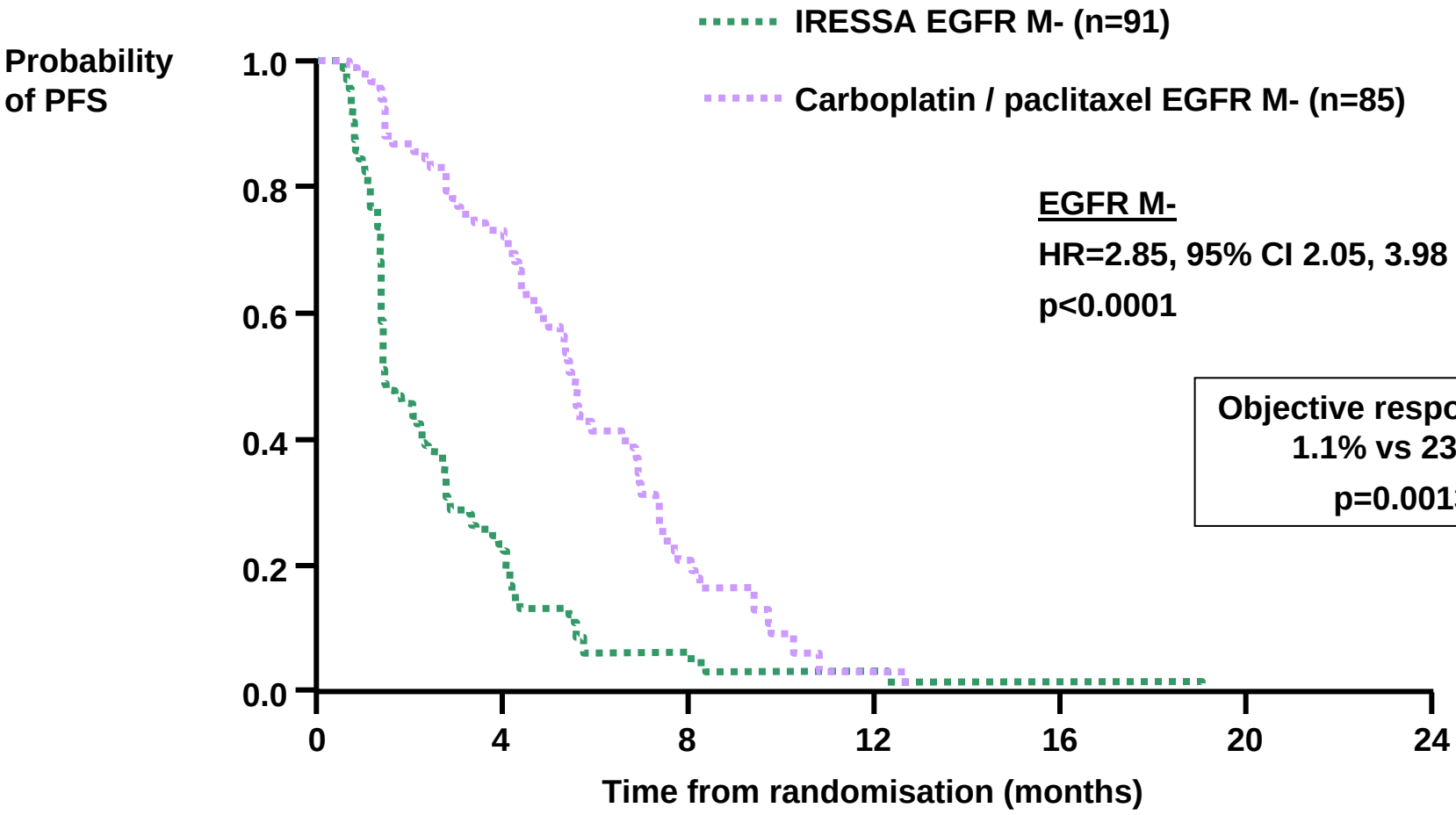
IPASS: Superior PFS and response rate for IRESSA in EGFR mutation positive patients



Mok 2009

M+, mutation positive

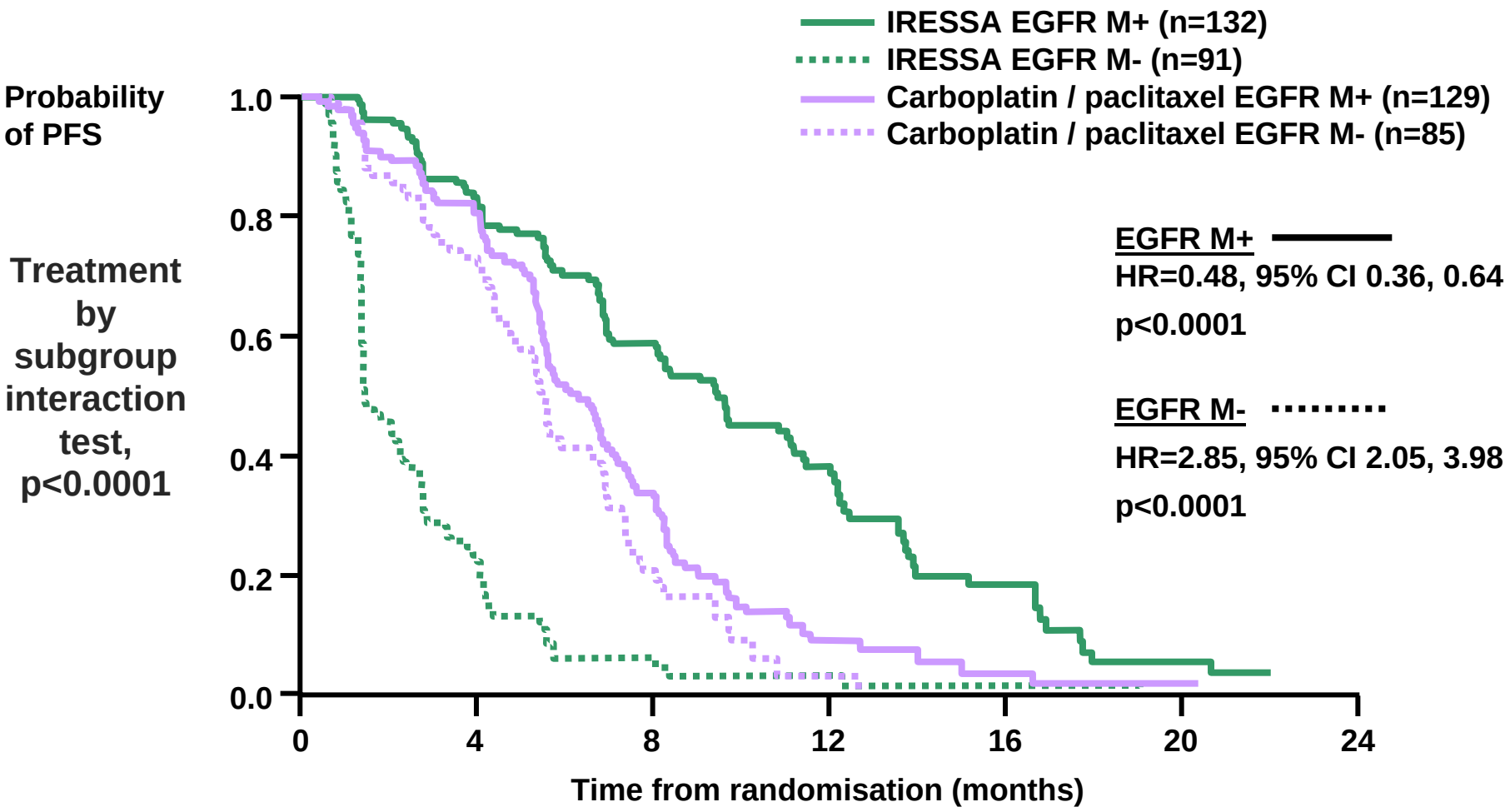
IPASS: Superior PFS and response rate for chemotherapy in EGFR mutation negative patients



Mok 2009

M-, mutation negative

IPASS: EGFR mutation is a strong predictor for differential PFS benefit between IRESSA and doublet chemotherapy



Mok 2009

M+, mutation positive; M-, mutation negative

European Indication – approved June 2009

IRESSA is indicated for the treatment of adult patients with locally advanced or metastatic non-small cell lung cancer (NSCLC) with activating mutations of EGFR-TK.

Post-hoc subgroup analyses in a positive PIII trial

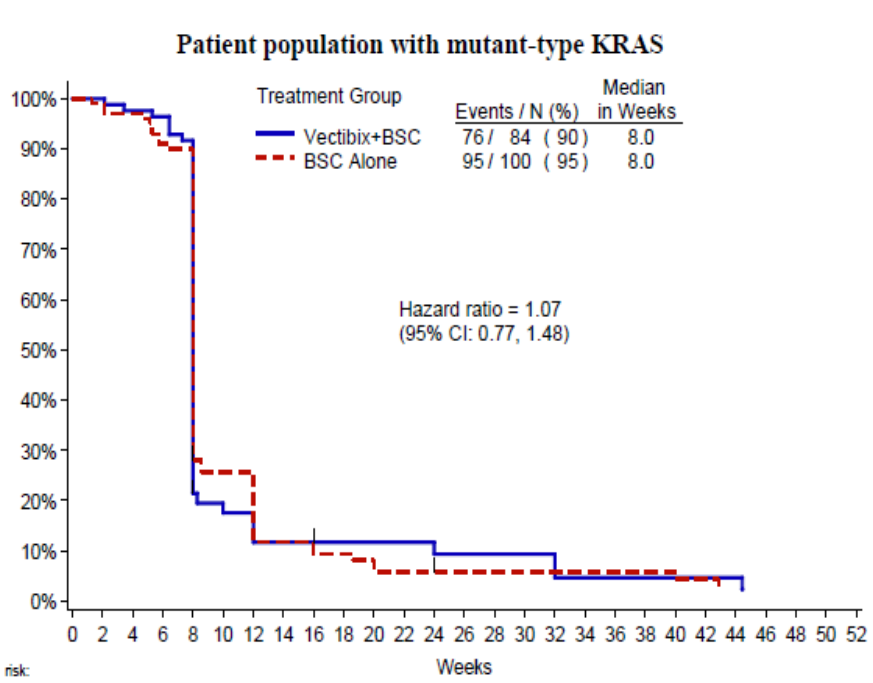
Panitumumab in metastatic colorectal cancer

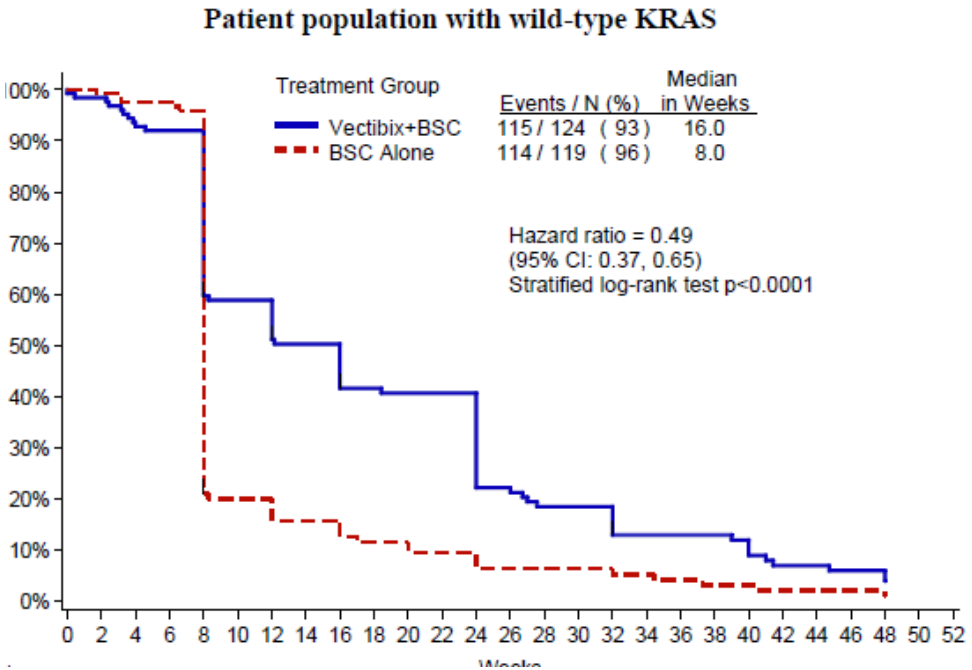
- Open label PIII trial vs best supportive care in n=463 patients
- PFS primary endpoint, HR= 0.54 95% CI(0.44 – 0.66). However, difference in median PFS times only around 1 week and p=0.60 for survival.

- EPAR:
- *In conclusion, the risk/benefit of panitumumab ...is not considered favourable due to the following grounds:*
  - *Only a very small effect on progression-free survival and no favourable effect has been shown in terms of overall survival or other clinical benefit endpoint.*
  - *The clinical efficacy observed is too small to constitute a clinical benefit and does not outweigh the risks associated to treatment with panitumumab.*
- *CHMP Recommendation:*
- *Risk-benefit balance unfavourable and therefore did not recommend the granting of the marketing authorisation.*

Re-examination by the CHMP

- *The Applicant has identified a biomarker (KRAS) which allows selecting patients who will not benefit from panitumumab treatment.*





- Note, tissue samples to assess KRAS status available in only around ½ patients randomised

CHMP revised recommendation

- ***Recommendation***
- *Based on the CHMP review of data on quality, safety and efficacy, the CHMP considered by majority decision that the risk/benefit balance of Vectibix as monotherapy for the treatment of patients with EGFR expressing metastatic colorectal carcinoma with non-mutated KRAS after failure of chemotherapy was favourable and therefore recommended the granting of the conditional marketing authorisation, subject to the following specific obligations: to provide results of ongoing studies 20050181 and 20050203.*

Xigris in sepsis FDA review 2001

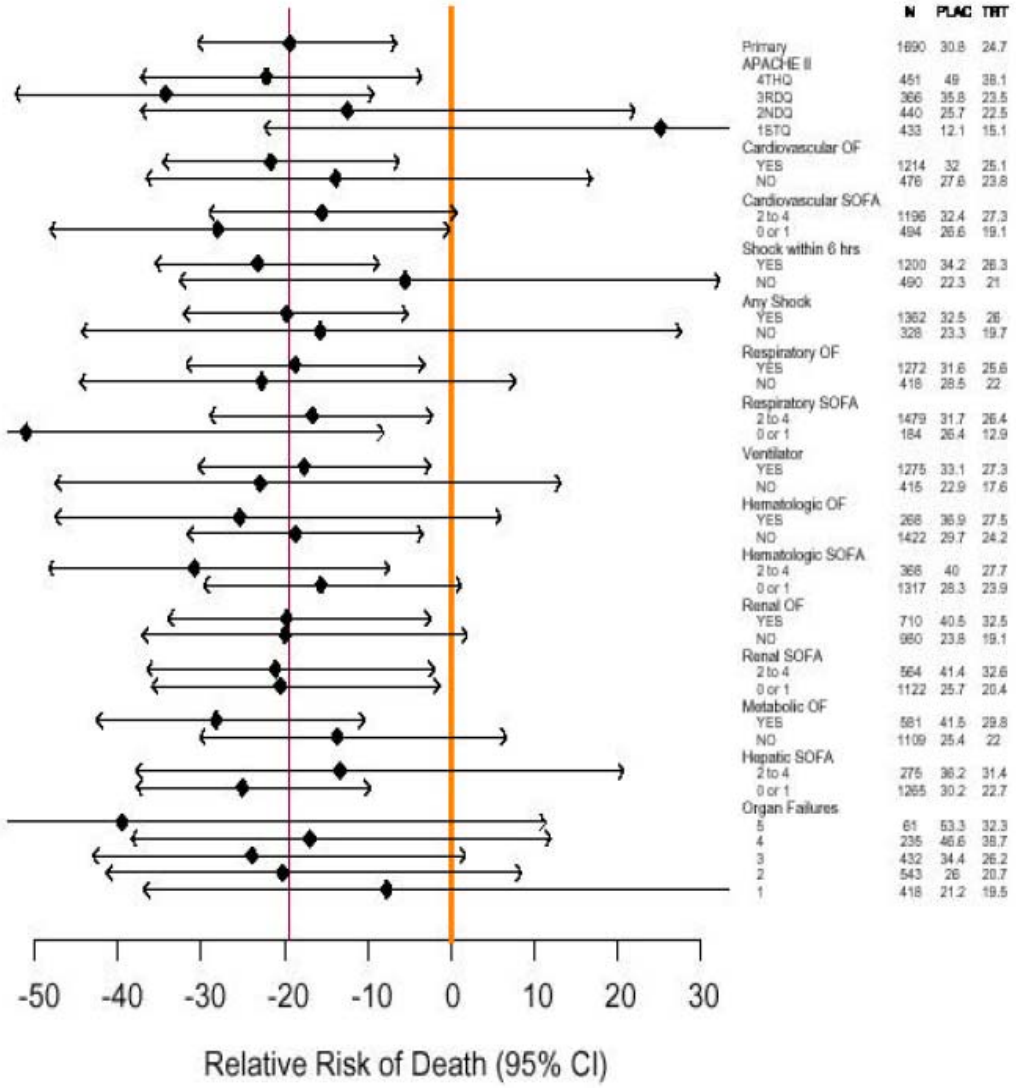
- Primary endpoint met with 28 day mortality 25% vs 31%, HR=0.81 (0.70, 0.93), p=0.0054.
- The data 'show some evidence of an interaction between treatment effect, mortality and APACHE II quartiles; p=0.09 for the interaction.'

Table 20. 28-day all-cause mortality for all patients and for subgroups defined by APACHE II as a measure of disease severity

	rhAPC		Placebo		Absolute Mort diff (%)	RR	95% CI for RR
	Total N	N (%)	Total N	N (%)			
Overall	850	210 (24.7)*	840	259 (30.8)	-6.1	0.81	0.70, 0.93
Apache II quartile (score)							
1st + 2nd (3-24)	436	82 (19)	437	83 (19)	0	0.99	0.75, 1.30
3rd + 4th (25-53)	414	128 (31)	403	176 (44)	-13	0.71	0.59, 0.85

- **INDICATIONS AND USAGE**
- Xigris is indicated for the reduction of mortality in adult patients with severe sepsis who have a high risk of death (e.g., as determined by APACHE II).

...copious amount of subgroup analyses performed by FDA

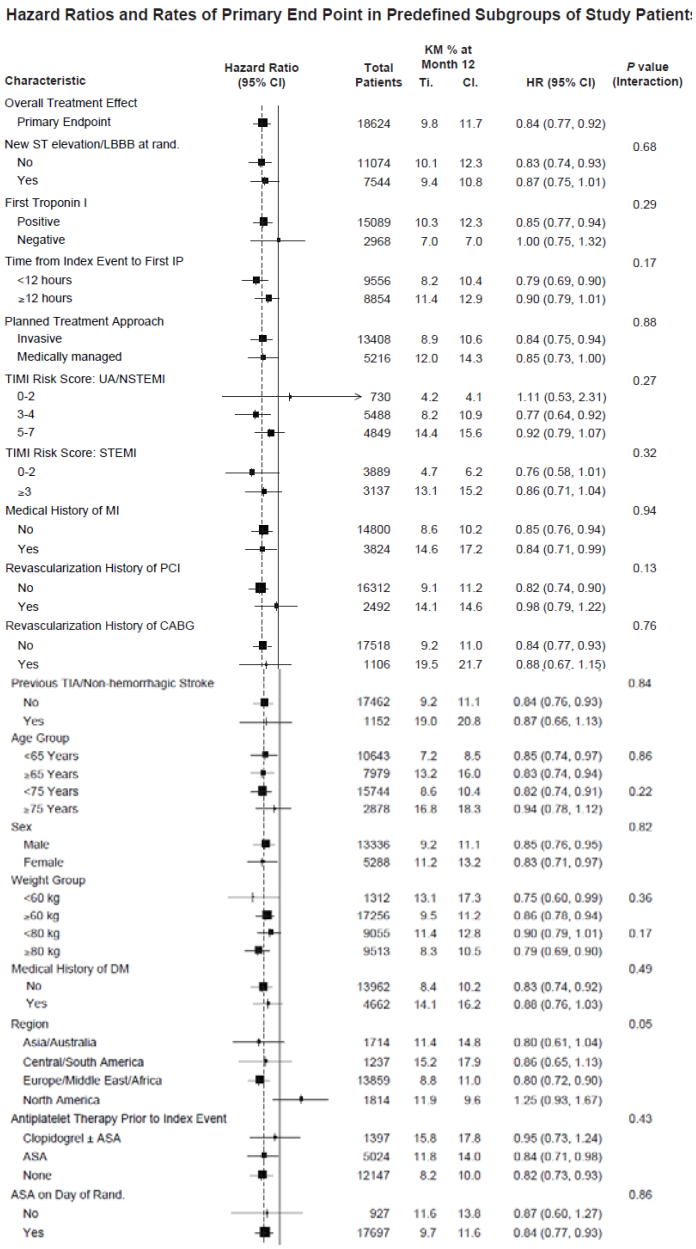


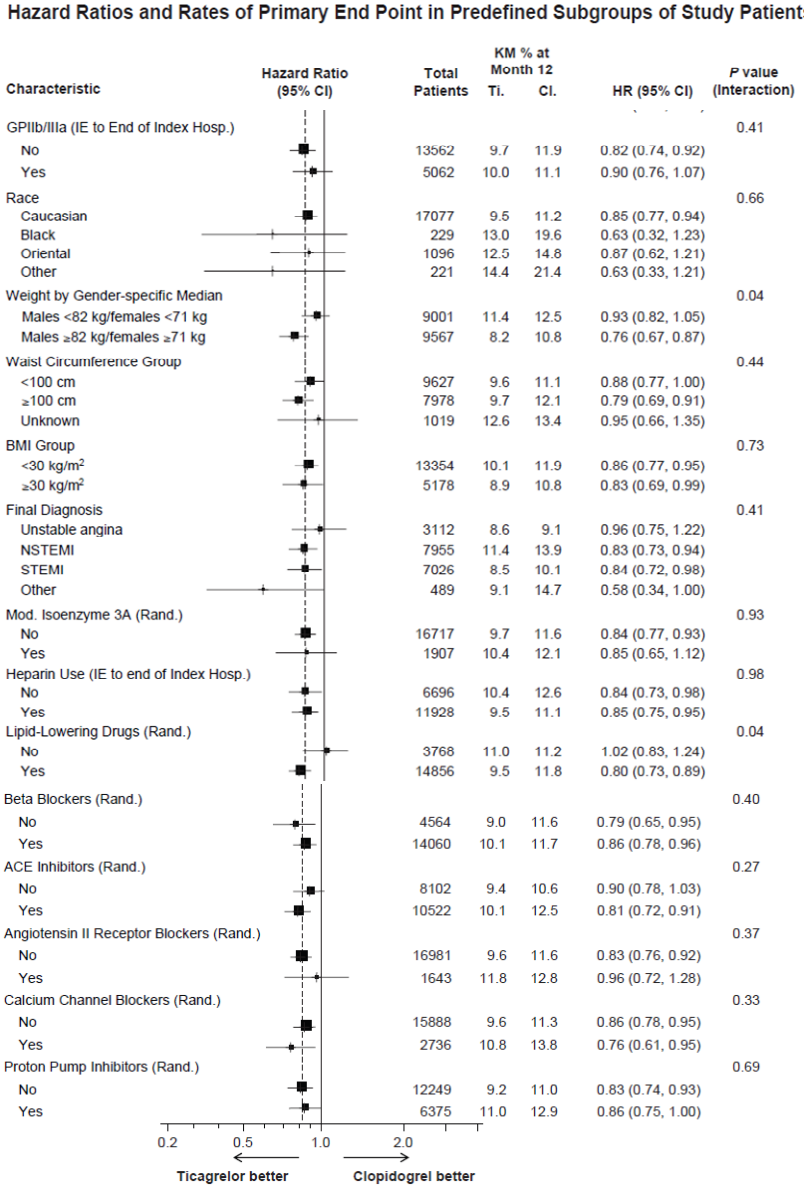
	N	PLAC	TRT
Primary	1690	30.8	24.7
APACHE II			
4THQ	451	49	38.1
3RDQ	366	35.8	23.5
2NDQ	440	25.7	22.5
1STQ	433	12.1	15.1
Cardiovascular OF			
YES	1214	32	25.1
NO	476	27.6	23.8
Cardiovascular SOFA			
2 to 4	1196	32.4	27.3
0 or 1	494	26.6	19.1
Shock within 6 hrs			
YES	1200	34.2	26.3
NO	490	22.3	21
Any Shock			
YES	1362	32.5	26
NO	328	23.3	19.7
Respiratory OF			
YES	1272	31.6	25.6
NO	418	28.5	22
Respiratory SOFA			
2 to 4	1479	31.7	26.4
0 or 1	184	26.4	12.9
Ventilator			
YES	1275	33.1	27.3
NO	415	22.9	17.6
Hematologic OF			
YES	268	36.9	27.5
NO	1422	29.7	24.2
Hematologic SOFA			
2 to 4	368	40	27.7
0 or 1	1317	28.3	23.9
Renal OF			
YES	710	40.5	32.5
NO	980	23.8	19.1
Renal SOFA			
2 to 4	564	41.4	32.6
0 or 1	1122	25.7	20.4
Metabolic OF			
YES	581	41.5	29.8
NO	1109	25.4	22
Hepatic SOFA			
2 to 4	276	36.2	31.4
0 or 1	1265	30.2	22.7
Organ Failures			
5	61	53.3	32.3
4	235	46.6	38.7
3	432	34.4	26.2
2	543	26	20.7
1	418	21.2	19.5

25th October 2011, EMA communicates Xigris to be withdrawn due to lack of efficacy

- Xigris (drotrecogin alfa (activated)) to be withdrawn due to lack of efficacy,
- After the annual reassessment in 2007 the CHMP concluded that the initial efficacy results of the pivotal clinical study had not been reproduced in further studies.
- The results of the PROWESS-SHOCK study have now become available and they fail to meet the primary endpoint of a statistically significant reduction in 28-day all-cause mortality in patients treated with Xigris compared with placebo.

PLATO, NEJM 2009

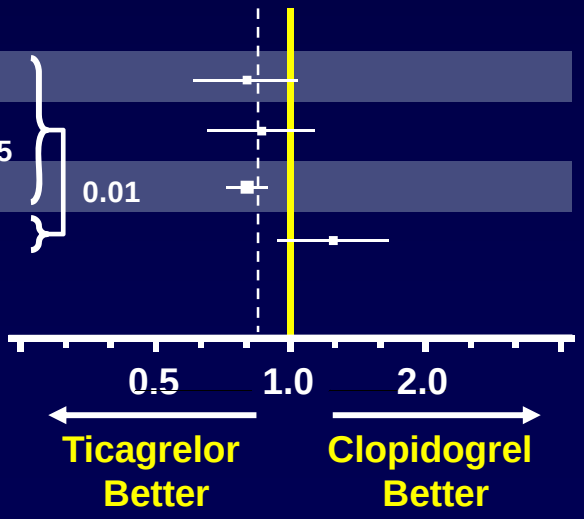




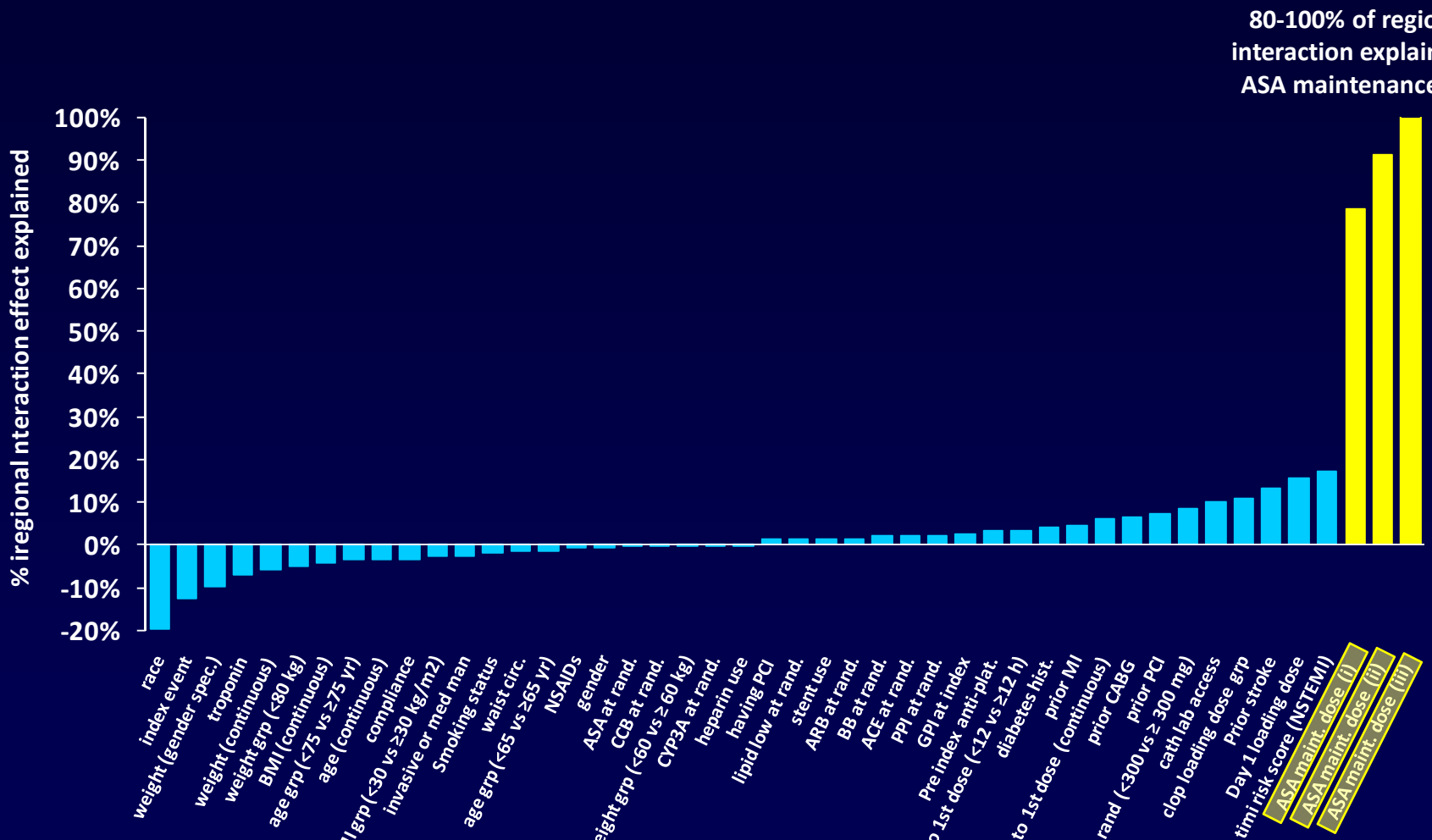
PLATO: Ticagrelor Effect Apparently Inconsistent Across Geographic Regions

- 31 pre-specified subgroup tests conducted for consistency
- No α-level adjustment for multiplicity
- Indication of qualitatively different outcomes by region
- Results in NA appear to be driven by US: HR 1.27 (0.92, 1.75)

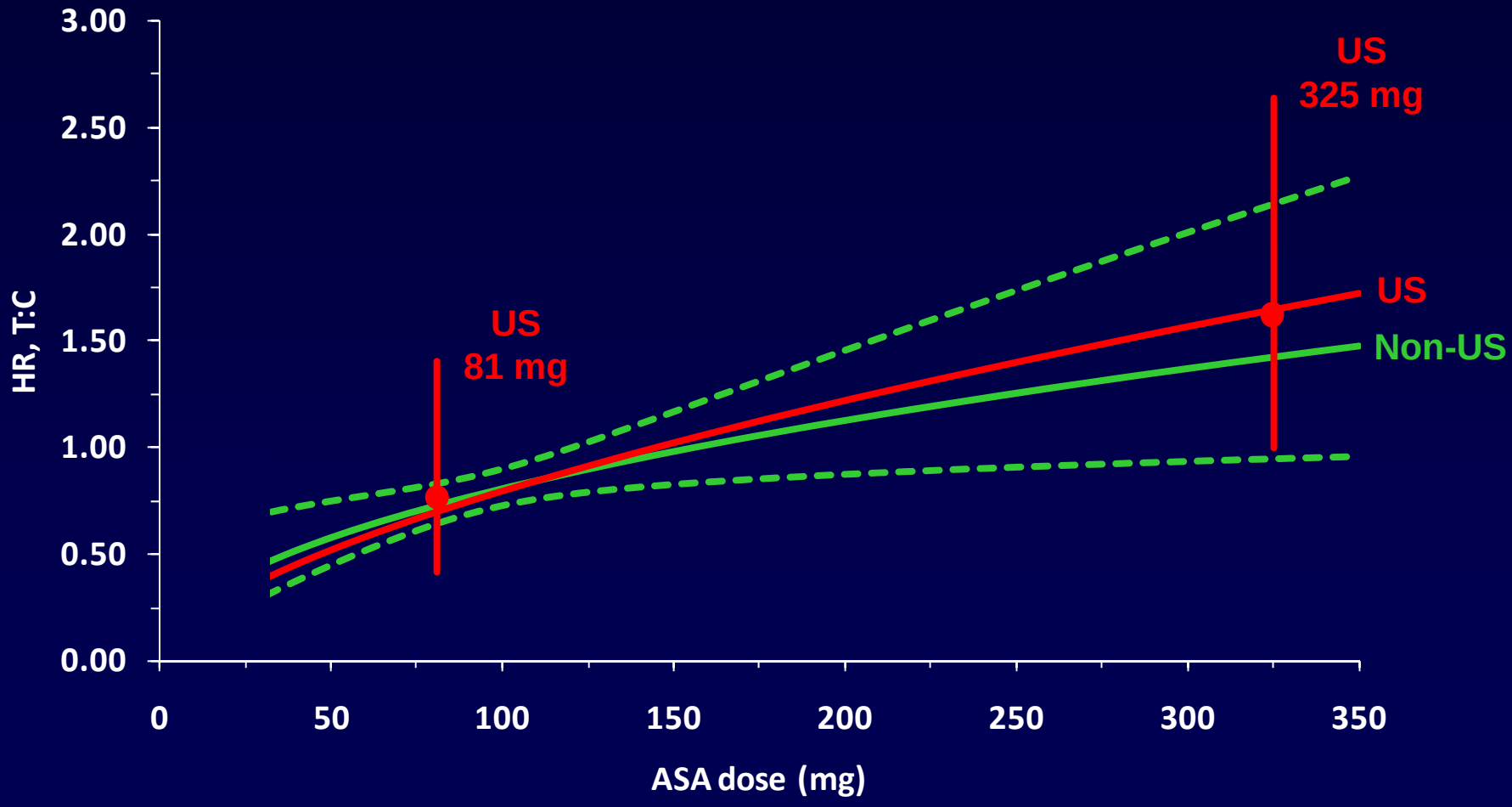
Characteristic	Total Patients	KM at Month 12 Tic	KM at Month 12 Clop	HR (95% CI)	Interaction p-values
Geographic Region					
Asia / Australia	1714	11.4	14.8	0.80 (0.61, 1.04)	0.045
Cent / Sth America	1237	15.2	17.9	0.86 (0.65, 1.13)	
Euro / Md E / Afr	13859	8.8	11.0	0.80 (0.72, 0.90)	0.01
North America	1814	11.9	9.6	1.25 (0.93, 1.67)	



PLATO: No Factor Potentially Accounts for the Regional Interaction with the Exception of ASA Maintenance Dose During Therapy



Relationship between aspirin dose and efficacy



Brilinta US label

- ***INDICATIONS AND USAGE***
- *BRILINTA has been studied in ACS in combination with aspirin. Maintenance doses of aspirin above 100 mg decreased the effectiveness of BRILINTA. Avoid maintenance doses of aspirin above 100 mg daily.*

Summary

1. Pre-planned, descriptive subgroup analyses to assess internal 'consistency' in a positive PIII trial.
2. Post-hoc subgroup analyses in a positive PIII trial.
3. Subgroup analyses in a negative PIII trial.
4. PIII trials prospectively designed to formally assess efficacy in a subgroup.

- Last provides the most secure route to an efficacy claim a specific subgroup of interest.
- Requirement for a set of statistical principles to guide the conduct and interpretation of subgroup analyses in confirmatory trials.
- Need for clarity on for the potential implications for product labelling and drug approval.