Statistical evaluation and analysis of regional interactions: The PLATO trial case study[1]

Kevin J Carroll

[1]Kevin J Carroll & Thomas R Fleming (2013). *SBR* Vol 5(2): 91-101

PLATO[1]

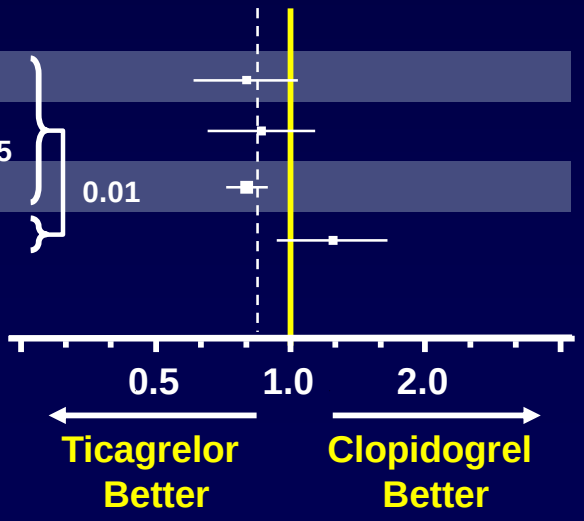
- Randomized double-blind study comparing BRILINTA (N=9333) to clopidogrel (N=9291), both given in combination with aspirin, in patients with acute coronary syndromes.
- Primary endpoint was time to first occurrence of CV death, MI or stroke.
- Randomisation across 41 countries.
- Primary endpoint met for BRILINTA 9.8% vs 11.7% events HR = 0·84 95% CI 0·77–0·92]; p=0·0003.
- Benefit also seen in overall mortality 4.5% vs 5.9% events HR = 0·78 95% CI 0·69–0·89]; p=0·0003.

[1] Wallentin, L et al. (2009) *N Engl J Med* Vol 361(11):1045-57.

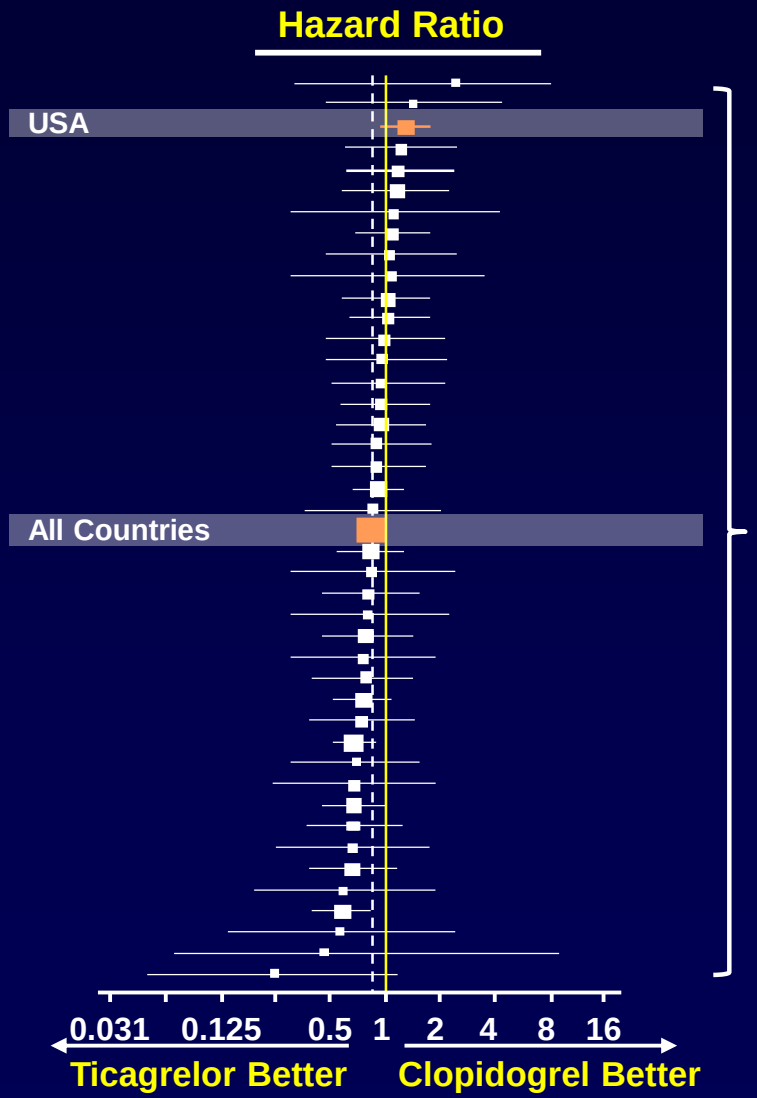
…However, the treatment effect was inconsistent across pre-defined geographic regions

- **31 pre-specified descriptive subgroup analyses 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)**

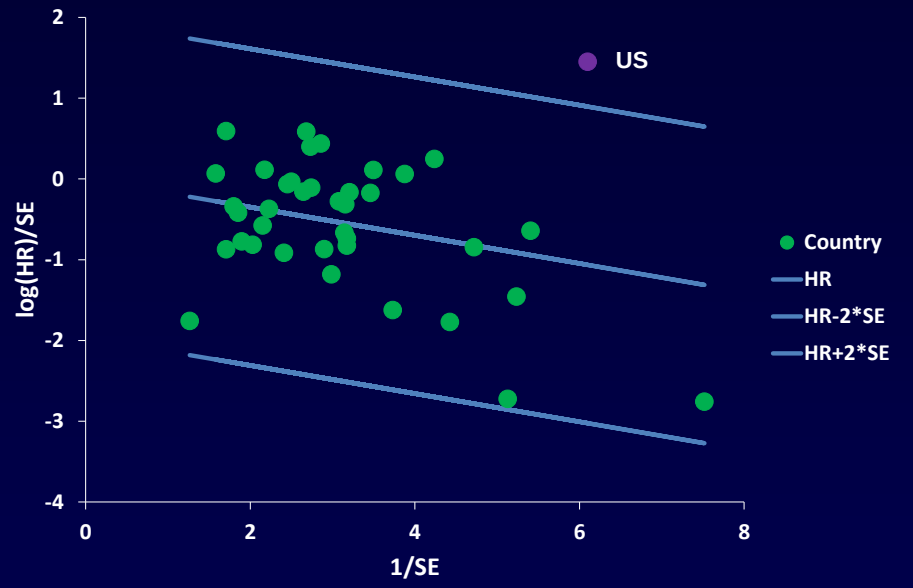
Characteristic	Total Patients	KM at Month 12 Tic	KM at Month 12 Clop	HR (95% CI)	Interaction p-values
Geographic Region					
Asia / Australia	1714	11.4	14.8	0.80 (0.61, 1.04)	0.045
Cent / Sth America	1237	15.2	17.9	0.86 (0.65, 1.13)	
Euro / Md E / Afr	13859	8.8	11.0	0.80 (0.72, 0.90)	0.01
North America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

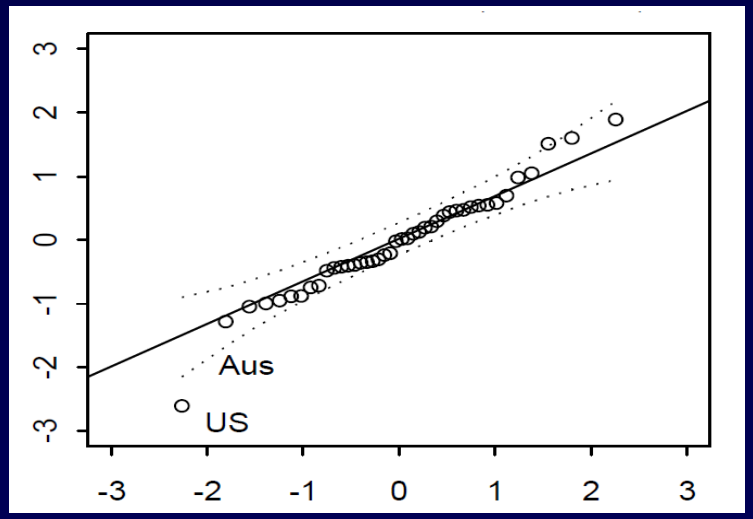


While global interaction test=NS, the US result stands out in Galbraith and Normal Probability plots



Derived from Mahaffey (2011)





Chance or A Real Difference Between Regions?

Possible Explanations:

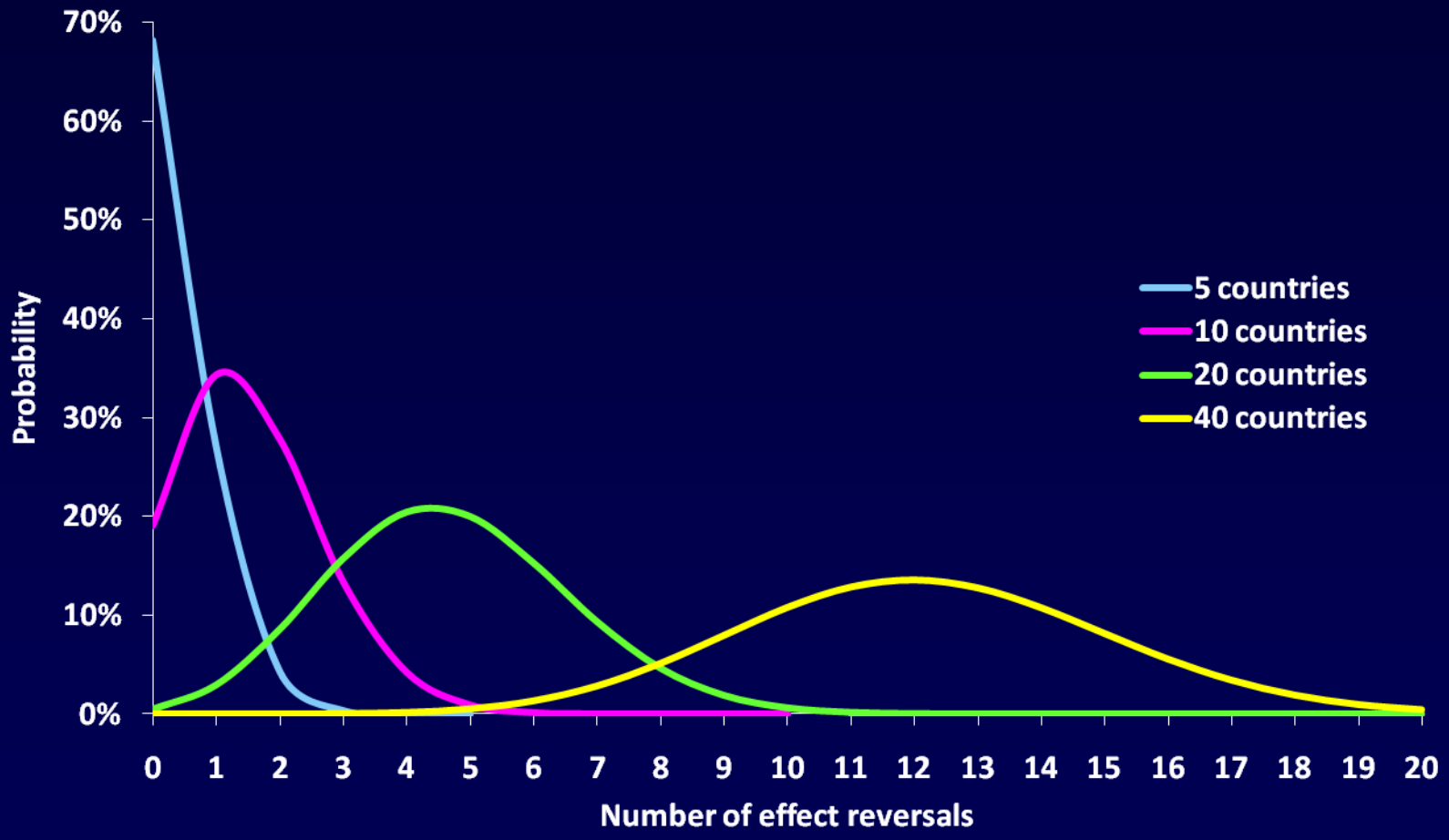
1. **Systematic issues in trial conduct at US sites**
   - **Ruled out**
2. **Play of chance**
   - **Plausible**
3. **Difference between US and non-US populations in important baseline characteristics or aspects of clinical management**
   - **Requires extensive investigation**

2. PLATO: Could the US Observation Be Due to Play of Chance?

- Yes
- Observed treatment-by-region interaction is of marginal statistical significance:
  - One of 31 descriptive interaction tests.
  - Adjustment for multiplicity would render the interaction p=NS.
- Switching of just one event in the NA cohort from ticagrelor to clopidogrel would render the regional interaction p=NS
- Given the overall PLATO result and distribution of patients and events across the 4 pre-specified regions[1]:
  - 32% chance of observing a HR>1 in at least one region.
  - 10% chance of observing HR>1 in the US while favouring ticagrelor in the other 3 regions.

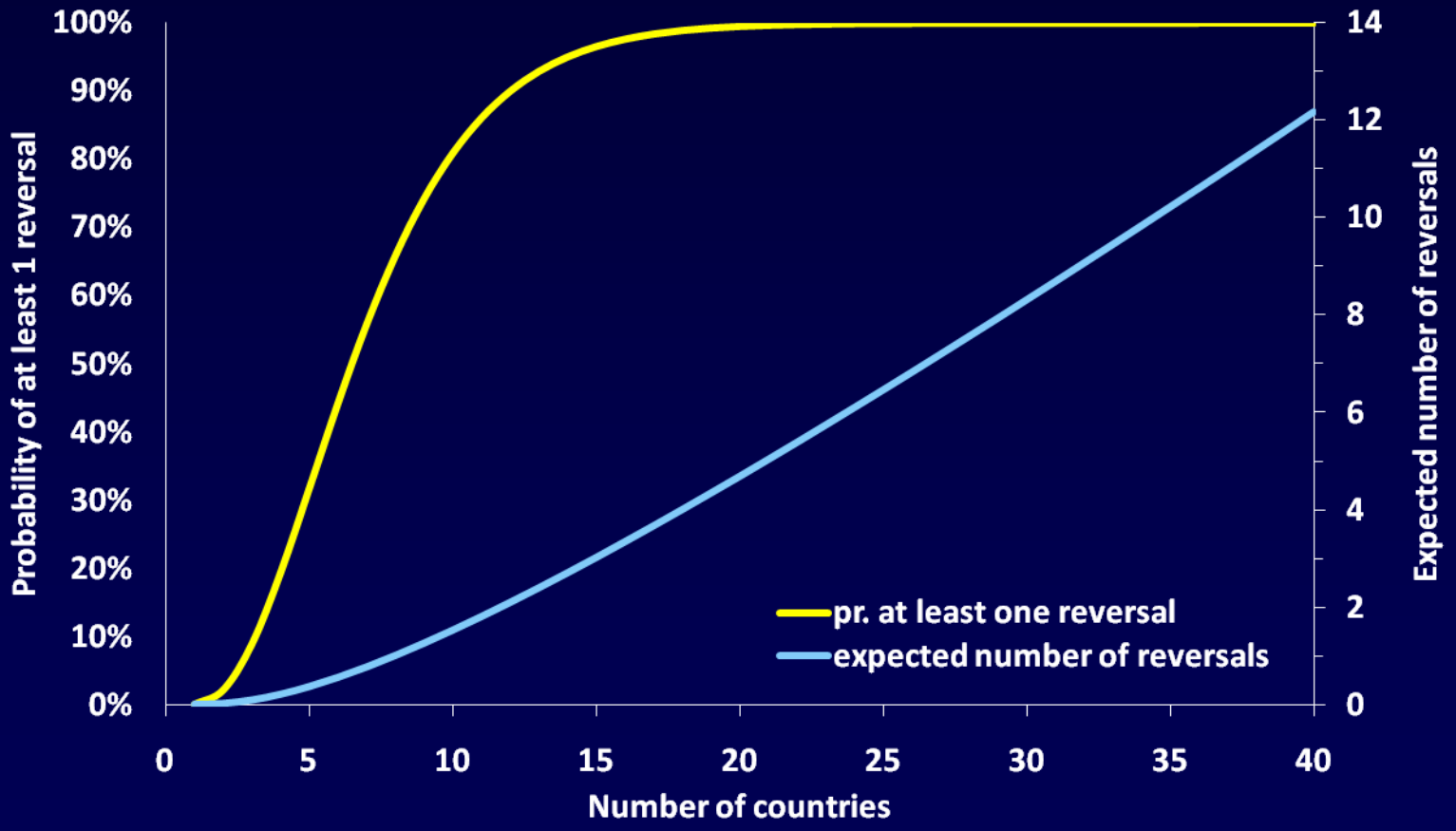
[1] Mahaffey et al. (2011) *Circulation*. 124(5):544-54.

“Effect Reversals”, where the treatment effect is positive overall but numerically negative in some regions, are to be expected in a large multiregional trials



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

“Effect Reversals”, where the treatment effect is positive overall but numerically negative in some regions, are to be expected in a large multiregional trials



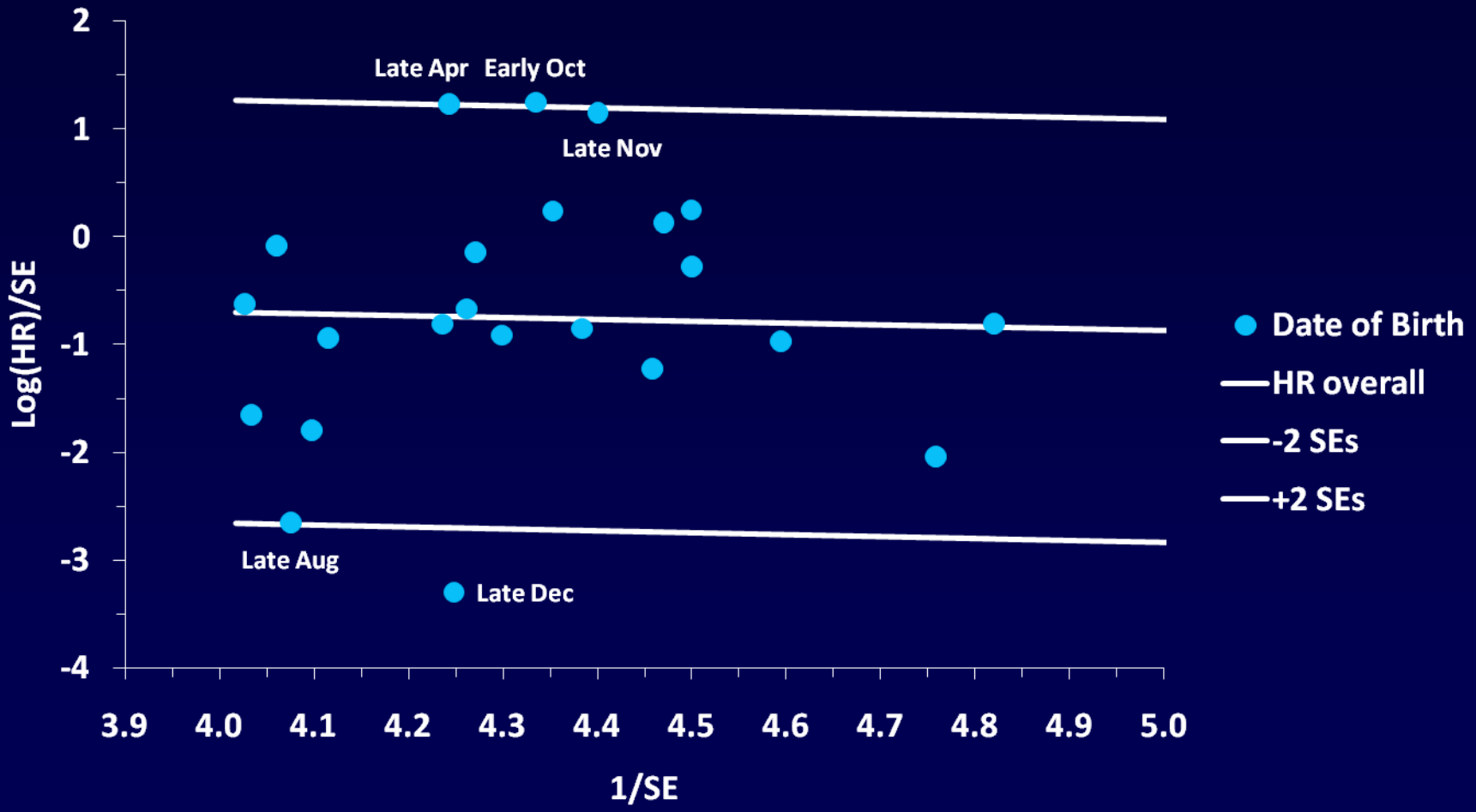
Senn, S and Harrell, F. On wisdom after the Event. *J Clin Epidemiol*, 1997. 50(7):749-751.

PLATO Pattern of effect reversals consistent with what would be expected in a large MRCT[1]

Expected no. countries with HR >1	Actual no. countries with HR >1	Expected no. countries with HR >1.25	Actual no. countries with HR >1.25
12.9	12	6.2	3

[1]Basel Biometric Section Sep 16, 2011:
http://www.ceb-institute.org/bbs/wp-content/uploads/2011/10/20110916_Carroll_Evaluation-of-regional-effects-in-MRCTs_2.pdf

Ticagrelor is indicated for patients born in late Summer or over the Christmas Holidays?



[1]Basel Biometric Section Sep 16, 2011:
http://www.ceb-institute.org/bbs/wp-content/uploads/2011/10/20110916_Carroll_Evaluation-of-regional-effects-in-MRCTs_2.pdf

FDA Summary Review[1], 8 July 2011

- **"…the finding suggests that the overall result might not apply to the US—and, in fact, appears to be adverse. In such a case, I believe that part of due diligence, on the part of the review team and the sponsor, is to evaluate such a finding to see how credible it is."**
- **Dr Stockbridge, Director Division of Cardiovascular and Renal Products.**

[1] http://www.accessdata.fda.gov/drugsatfda_docs/nda/2011/022433Orig1s000TOC.cfm

hst saca; dep stngth

3. Are There Imbalances in Baseline Characteristics or Clinical Management That Might Explain the US vs Non-US Regional Interaction?[1]

Factors evaluated in exploratory analyses

- Race
- Index event
- Weight#
- Troponin
- BMI #
- Age #
- Compliance
- ASA at rand.
- Invasive or med man
- Smoking status
- Waist circumference
- ACE at rand.
- NSAID at rand.
- Gender
- CCB at rand.
- Time index to 1st dose #
- CYP3A at rand.
- Heparin use
- PCI <24h of rand.
- Lipid low at rand.
- Stent use
- ARB at rand.
- BB at rand.
- PPI at rand.
- GPI at rand.
- Pre index anti-plat.
- Diabetes hist.
- Prior MI
- Prior CABG
- Prior PCI
- Cath lab access
- Clop loading dose
- TIMI risk score
- ASA loading dose
- ASA maintenance dose #

- # Some factors defined in different ways, e.g age: <65 vs ≥ 65 and age <75 vs ≥ 75.
- ASA dose defined for patients who had (i) at least 5 days or (ii) at least 2 days of ASA; and (iii) as agreed with FDA, for patients with at least 1 maintenance dose to avoid the biasing influence of high ASA loading dose.
- ASA loading dose considered separately.

[1]FDA Cardio Renal Advisory Committee, 2010.
http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/ucm221382.htm

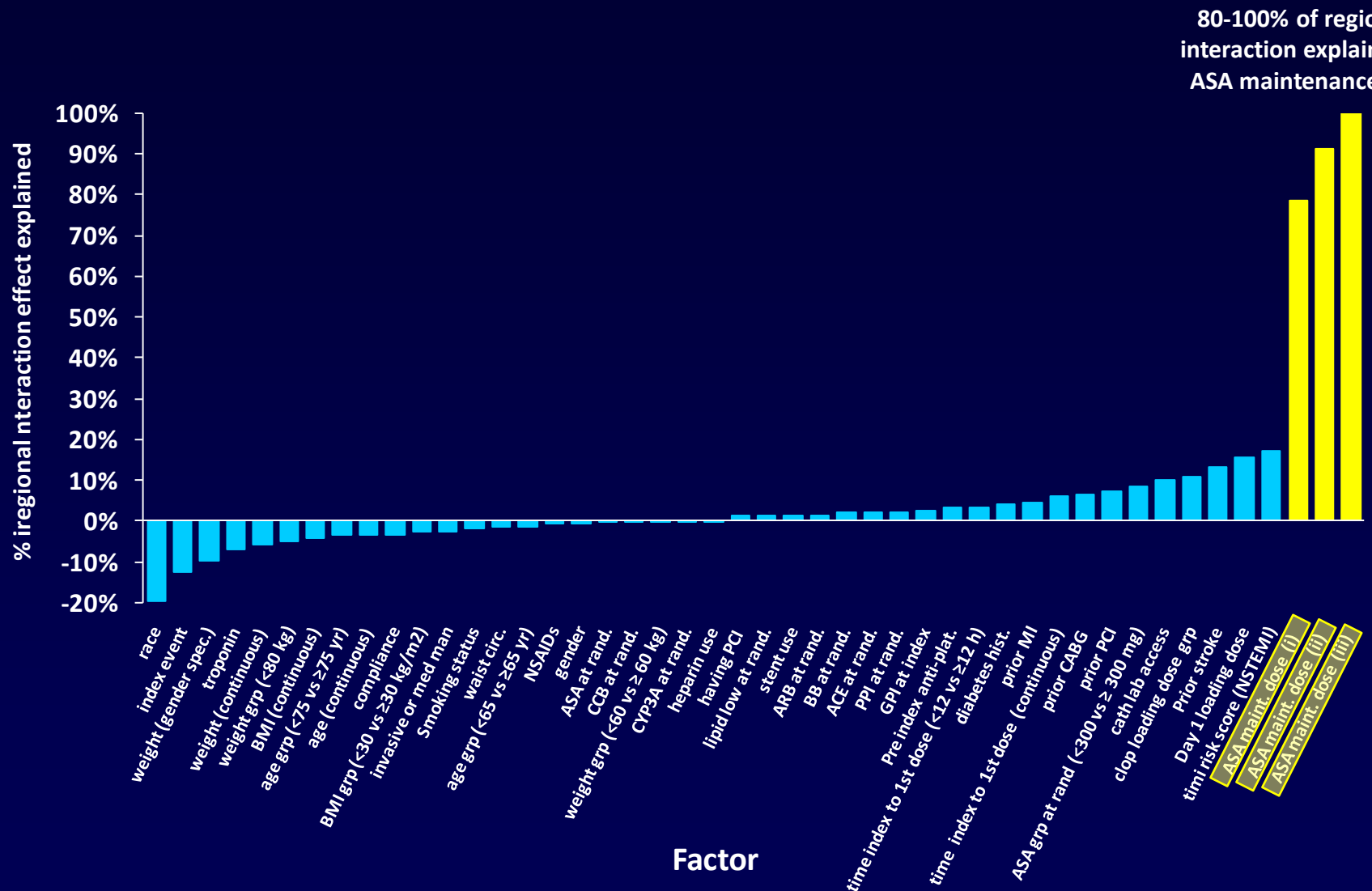
What kind of factors or patient characteristics might be 'effect modifying' and possibly explain the US vs non-US result?

- To explain a meaningful fraction of the US/non-US interaction, a factor is needed that simultaneously:
  - (i) has a strong qualitative interaction with randomized treatment for the primary endpoint and
  - (ii) is strongly imbalanced between US and non US settings

- Weakly imbalanced prognostic factors will likely not be sufficient to explain the US result

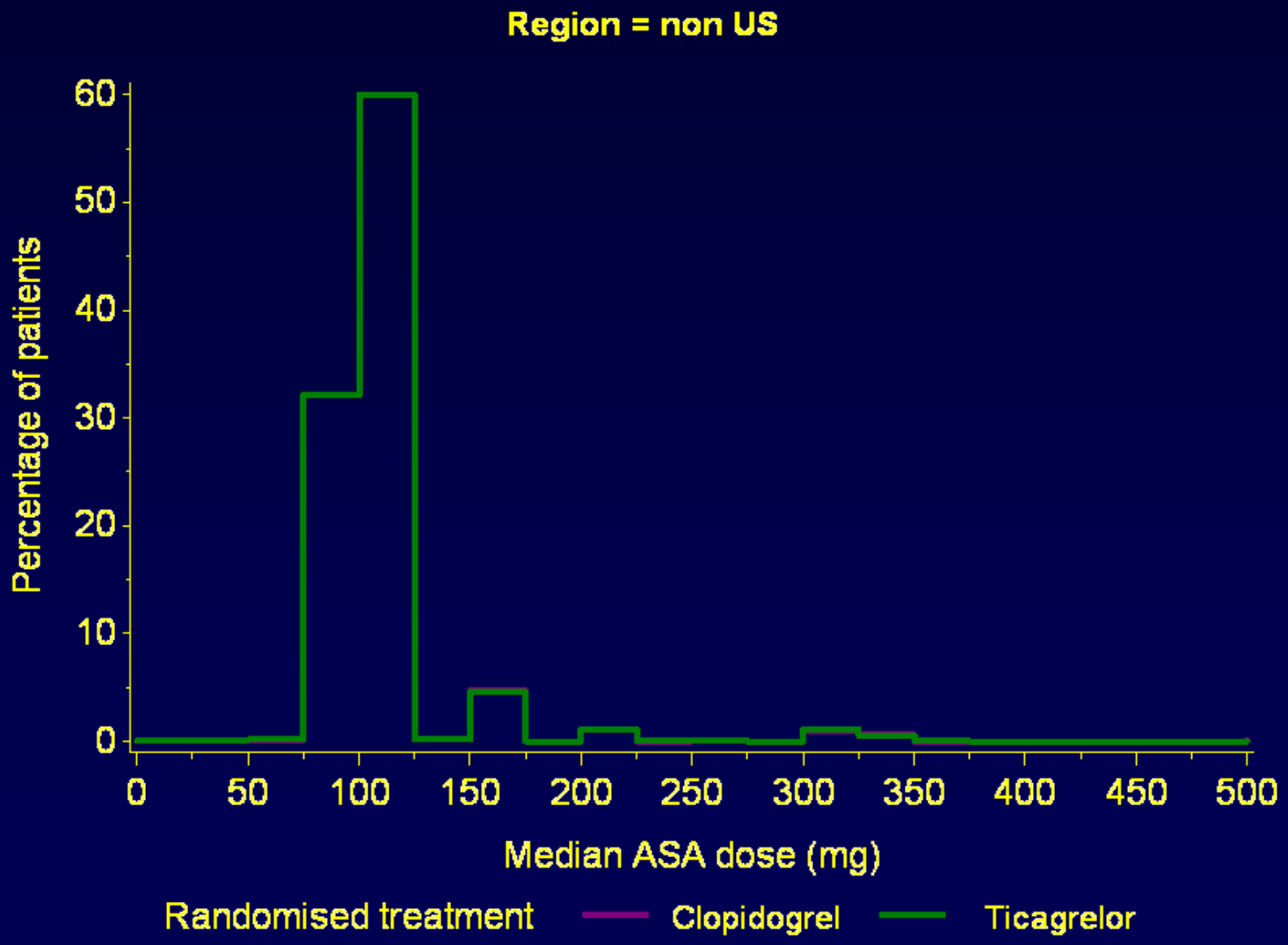
- How can we achieve a robust analysis to explore which factors, if any, might be driving the US interaction?

No factor potentially accounts for the regional interaction with the exception of aspirin maintenance dose[1]



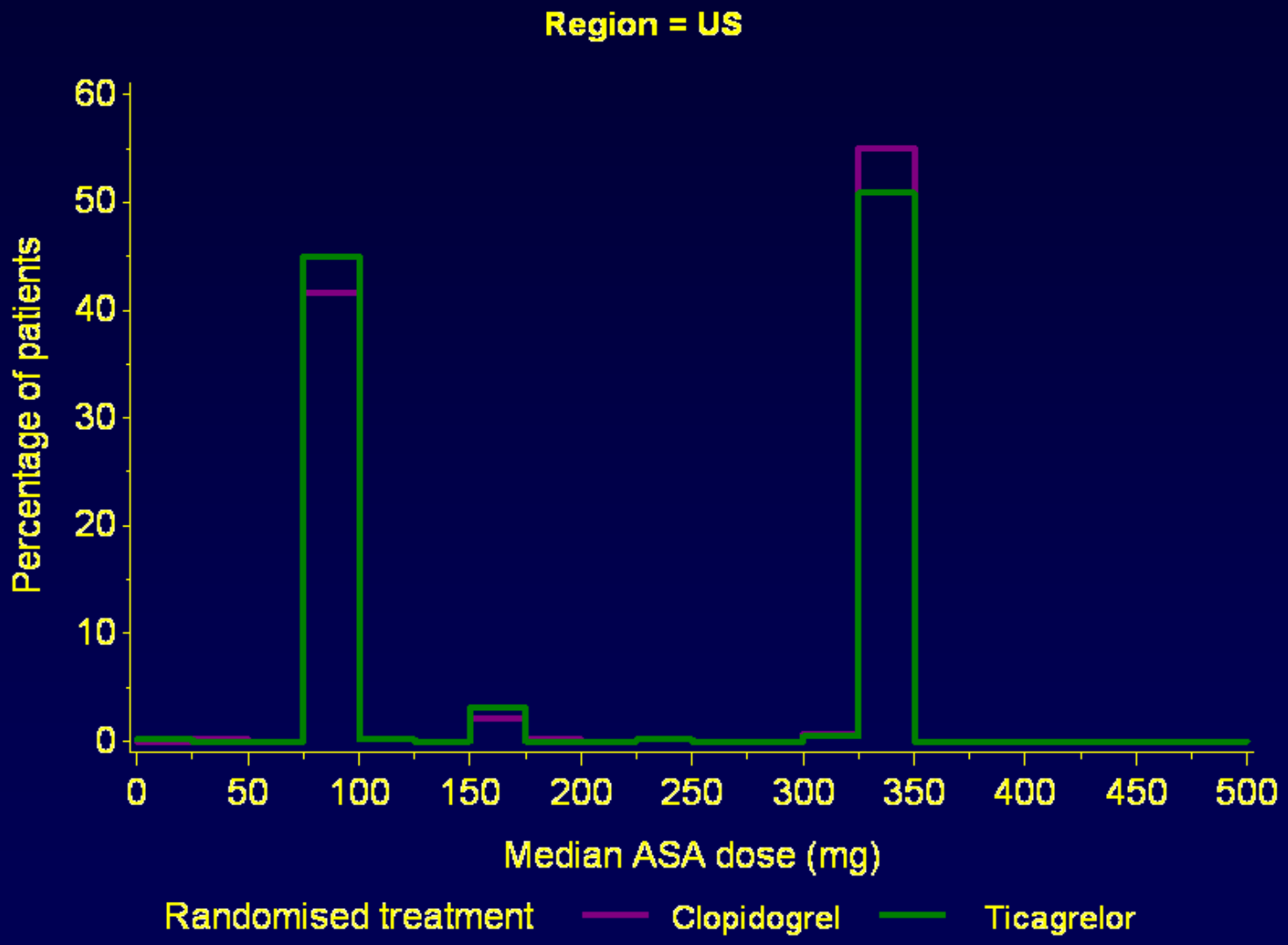
[1]FDA Cardio Renal Advisory Committee, 2010.
http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/ucm221382.htm

ASA maintenance dose in non-US patients is independent of randomised treatment[1]



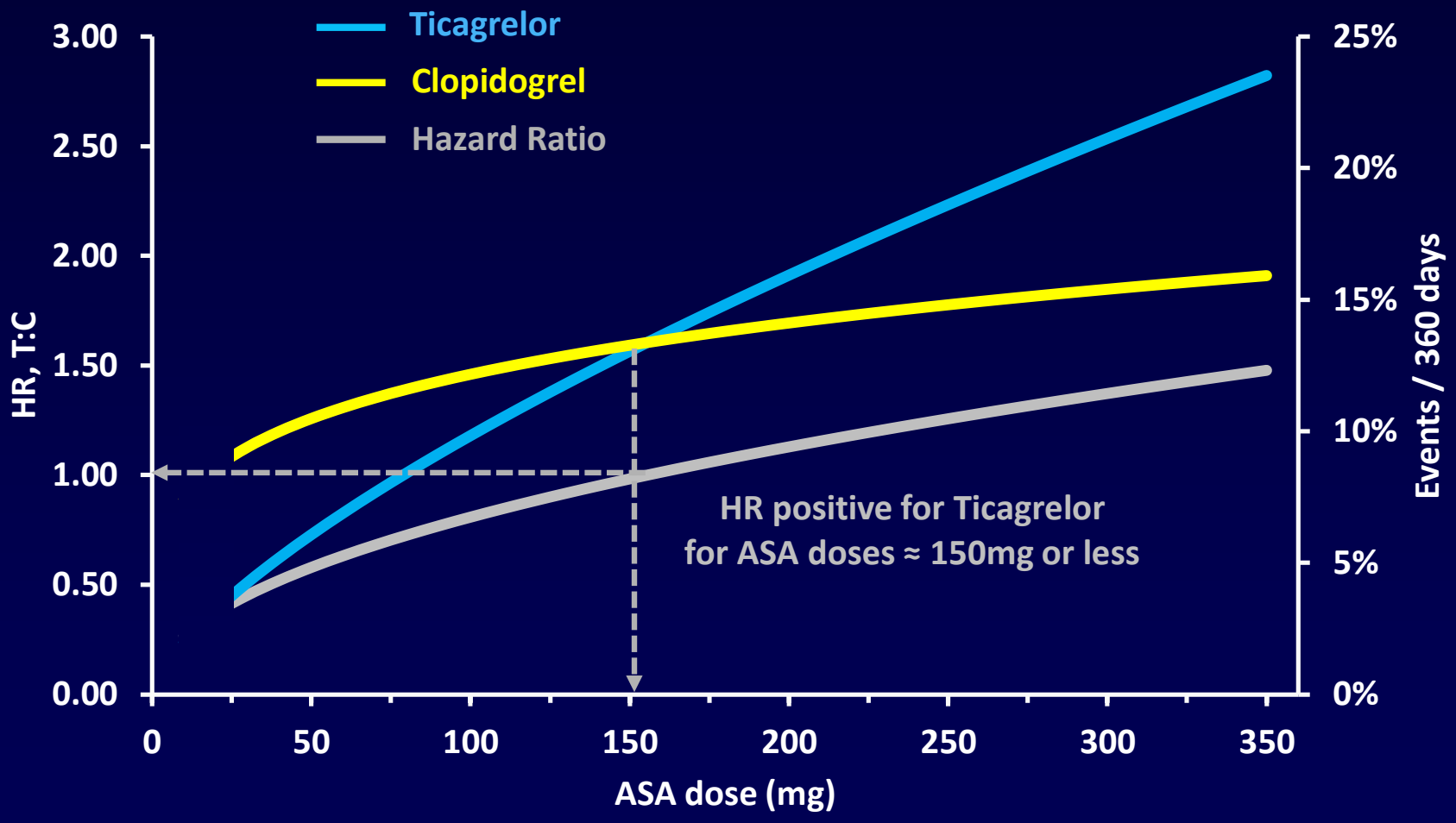
[1]Kevin J Carroll & Thomas R Fleming (2013). *SBR* Vol 5(2): 91-101, Supplm Appendix

The same is true for US patients but the distribution of ASA dose is very different[1]



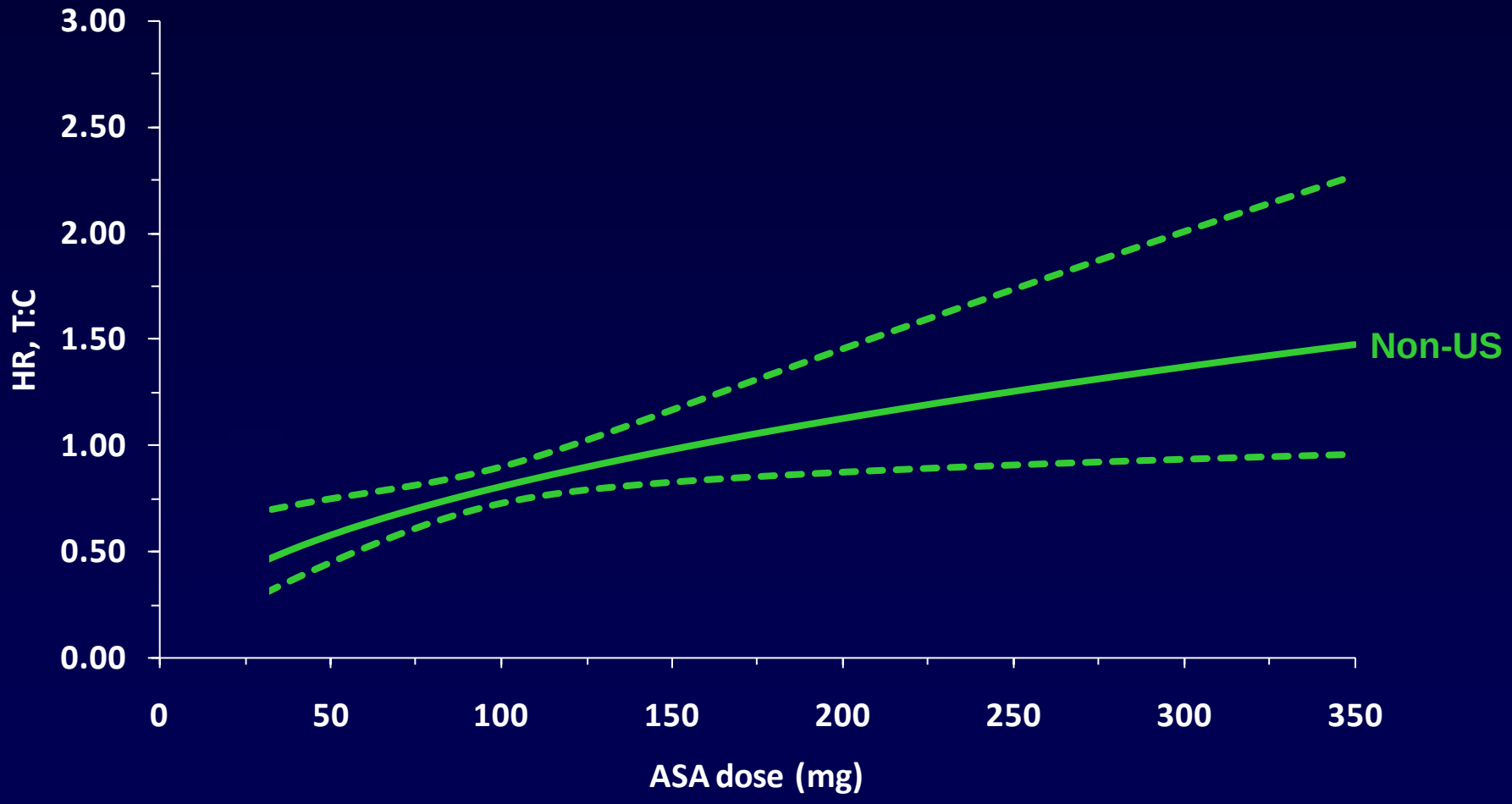
[1]Kevin J Carroll & Thomas R Fleming (2013). *SBR* Vol 5(2): 91-101, Supplm Appendix

Event rates increase for both Ticagrelor and Clopidgrel with increasing ASA dose, but to a greater extent with Ticagrelor



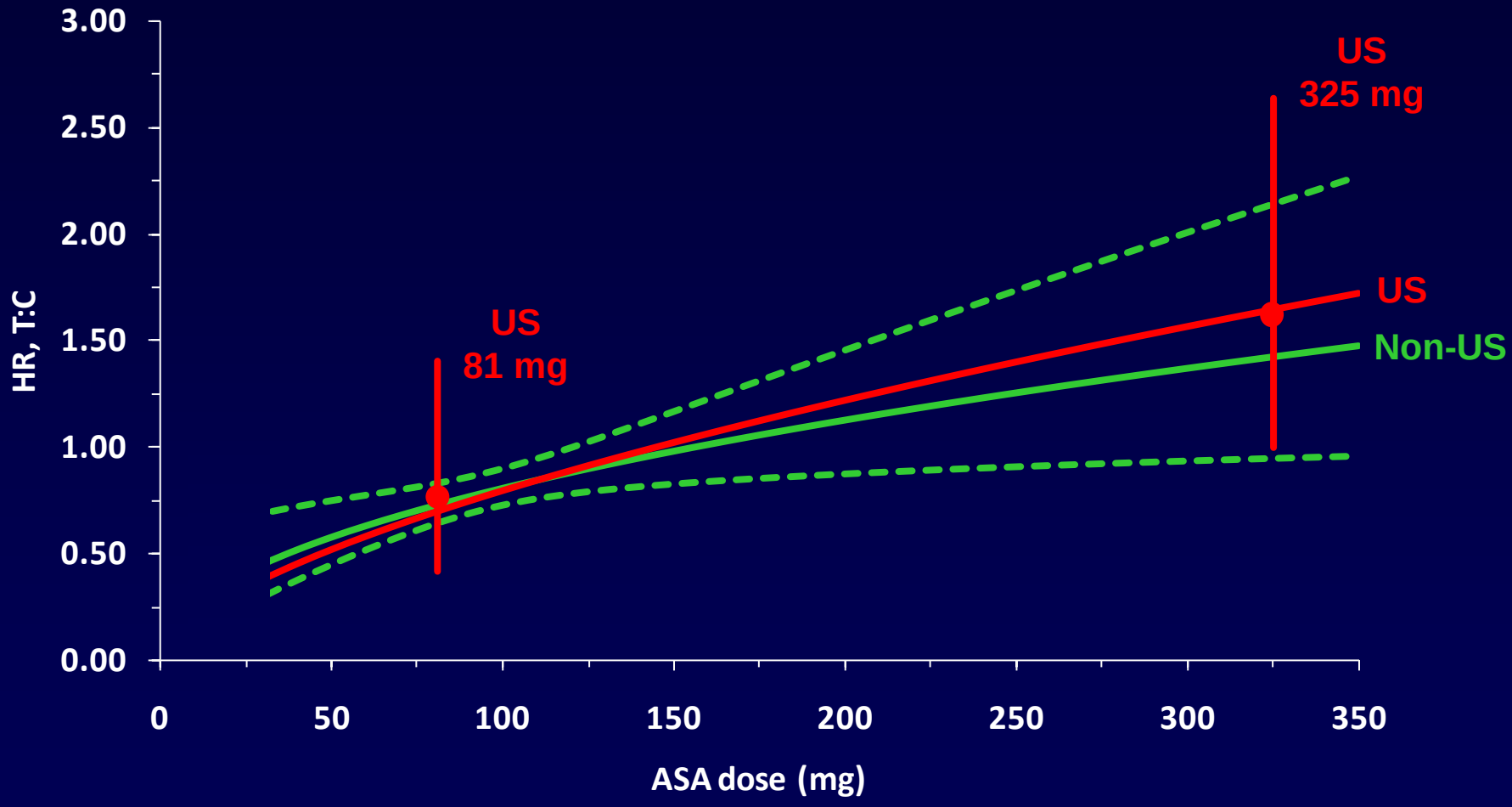
Weibull modelling; event rates are the integrated hazard over (0-360] days

The Relationship Between ASA Maintenance Dose and Treatment Effect is seen in Non-US patients



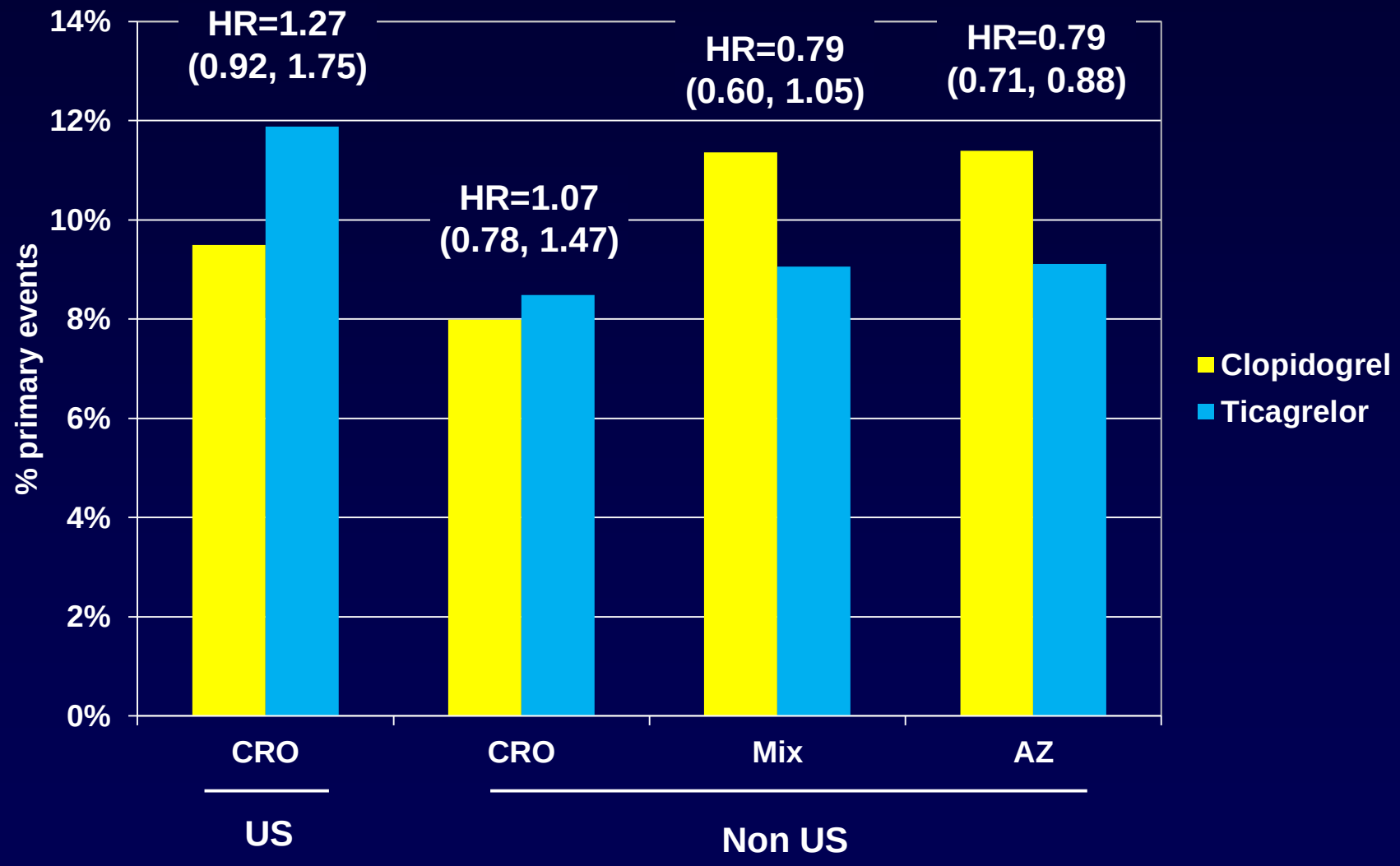
[1]FDA Cardio Renal Advisory Committee, 2010.
http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/ucm221382.htm

And this closely reflects that seen in US patients



[1]FDA Cardio Renal Advisory Committee, 2010.
http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/ucm221382.htm

Source of site monitoring largely confounded with region[1]



[1]Kevin J Carroll & Thomas R Fleming (2013). *SBR* Vol 5(2): 91-101, Supplm Appendix

FDA CRL 16 December 2010

- 13 new ASA definitions × 4 covariate classifications × 4 endpoints × 3 populations × 4 different imputation methods for missing ASA data.
- Each evaluated via 6 different, increasingly complex Cox regression models + categorical analysis.
- Categorical subset analyses for STEMI/NSTEMI, invasive/non invasive strategy by intent and early/no early intervention × 8 ASA definitions × 5 endpoints x 3 populations x 4 imputation methods.
- ASA dose on T vs C × 8 ASA definitions x 3 populations x 6 imputation methods for pts going to angioplasty, pts with and without a stent and by type of stent.
- FDA 'worst case' imputation: 13 new ASA definitions × 3 populations × 2 Cox regression models + categorical analysis
- Forest plots 13 ASA definitions × 3 populations × 4 imputation methods = 156 for primary endpoint
- HR vs ASA dose plots 13 ASA definitions × 1 endpoints × 3 populations × 4 imputation methods = 156 for primary endpoint

Full response and analyses submitted 20 January 2011

FDA Summary Review[1], 8 July 2011

- "… post-randomization dose of aspirin does appear to account for regional differences, at least in the statistical sense. The Agency issued a Complete Response letter on 16 December 2011. **I interpret the Agency's position with regard to approval to have been critically dependent upon the persuasiveness of the aspirin hypothesis.** Had the Agency been ready to accept the regional disparity in results as a chance finding, it would have approved Brilinta in the first cycle"

[1] http://www.accessdata.fda.gov/drugsatfda_docs/nda/2011/022433Orig1s000TOC.cfm

soaftallwrk ws wrth it

“The most likely identified factor distinguishing US and non-US subjects is aspirin dose”

Imputations						
If a subject has...	...some data, impute...	Zero	X	X		
		Previous value			X	X
	...no data, impute...	Zero		X		X
		Country median	X		X	
		Worst case				

Analyses	Metric	For analyses of events... ...on days 1-30, start with later of... ...Day	... or days before censoring	...on days >30, start with later of ...Day	... or days before censoring		M1	M2	M3	M4
	Mean	1	5	1	5	A1	■			■
	Mean	1	10	1	10	A2	■	■		■
	Mean	1	30	1	30	A3	■	■		■
	Median	1	5	1	5	A4	■	■	■	■
	Median	1	10	1	10	A5	■	■	■	■
	Median	1	30	1	30	A6	■	■	■	■
	Last	1	30	1	30	A7	■	■	■	■
	Mean	1	30	1	Any	A8				
	Median	1	30	1	Any	A9	■			
	Maximum	1	30	1	Any	A10				
	Median	1	30	31	Any	A11	■	■	■	■
	Median	1	30	2	Any	A12	■	■	■	■
	Last	1	30	1	Any	A13				

US Label

WARNING: ASPIRIN DOSE AND BRILINTA EFFECTIVENESS

- **Maintenance doses of aspirin above 100 mg reduce the effectiveness of BRILINTA and should be avoided. After any initial dose, use with aspirin 75-100 mg per day**

- **"Like any unplanned subset analysis, especially one where the characteristic is not a true baseline characteristic (but may be determined by usual investigator practice), the above analyses must be treated with caution. It is notable, however, that aspirin dose predicts outcome in both regions with a similar pattern, and that the pattern is similar for the two major components of the primary endpoint, CV death and non-fatal MI."**
- **"Despite the need to treat such results cautiously, there appears to be good reason to restrict aspirin maintenance dosage accompanying ticagrelor to 100 mg. Higher doses do not have an established benefit in the ACS setting, and there is a strong suggestion that use of such doses reduces the effectiveness of BRILINTA."**

Summary

- PLATO met its primary endpoint but a qualitative regional interaction was observed between US and non-US regions
- Issues related to trial conduct ruled out
- Chance cannot be ruled out entirely
- Extensive evaluation of the data revealed ASA maintenance dose was strongly imbalanced across US and non-US regions, and statistically accounted for 80-100% of the observed interaction
- Data suggest the regional interaction is, in fact, an underlying interaction with ASA maintenance dose
- Evaluation of unexpected regional interactions in MRCTs requires very extensive, consistent and clinically persuasive analyses
- Statistical arguments that appeal to chance alone are unlikely to be successful