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

Kevin J Carroll

PLATO

- 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.

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

- 31 pre-specified descriptive subgroup analyses conducted for consistency
- No $\alpha$-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)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total Patients</th>
<th>KM at Month 12</th>
<th>HR (95% CI)</th>
<th>Interaction p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Tic</td>
<td>Clop</td>
<td></td>
</tr>
<tr>
<td>Asia / Australia</td>
<td>1714</td>
<td>11.4</td>
<td>14.8</td>
<td>0.80 (0.61, 1.04)</td>
</tr>
<tr>
<td>Cent / Sth America</td>
<td>1237</td>
<td>15.2</td>
<td>17.9</td>
<td>0.86 (0.65, 1.13)</td>
</tr>
<tr>
<td>Euro / Md E / Afr</td>
<td>13859</td>
<td>8.8</td>
<td>11.0</td>
<td>0.80 (0.72, 0.90)</td>
</tr>
<tr>
<td>North America</td>
<td>1814</td>
<td>11.9</td>
<td>9.6</td>
<td>1.25 (0.93, 1.67)</td>
</tr>
</tbody>
</table>
While global interaction test = NS, the US result stands out in Galbraith and Normal Probability plots.

Hazard Ratio

USA

Ticagrelor Better
Clopidogrel Better

All Countries

Ticagrelor Better
Clopidogrel Better

Global interaction p = 0.95

Derived from Mahaffey (2011)

Chen (2013)

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 \( p = \text{NS} \).

- Switching of just one event in the NA cohort from ticagrelor to clopidogrel would render the regional interaction \( p = \text{NS} \).

- 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.

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

PLATO Pattern of effect reversals consistent with what would be expected in a large MRCT¹

<table>
<thead>
<tr>
<th>Expected no. countries with HR &gt;1</th>
<th>Actual no. countries with HR &gt;1</th>
<th>Expected no. countries with HR &gt;1.25</th>
<th>Actual no. countries with HR &gt;1.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.9</td>
<td>12</td>
<td>6.2</td>
<td>3</td>
</tr>
</tbody>
</table>

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

![Graph showing Log(HR)/SE vs 1/SE with data points for different months: Late Apr, Early Oct, Late Nov, Late Aug, Late Dec.](http://www.ceb-institute.org/bbs/wp-content/uploads/2011/10/20110916_Carroll_Evaluation-of-regional-effects-in-MRCTs_2.pdf)
FDA Summary Review\textsuperscript{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.

\textsuperscript{1} http://www.accessdata.fda.gov/drugsatfda_docs/nda/2011/022433Orig1s000TOC.cfm
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

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

¹ 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.

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. 

80-100% of regional interaction explained by ASA maintenance dose.
ASA maintenance dose in non-US patients is independent of randomised treatment\(^1\)

\[\text{Region} = \text{non US}\]

![Graph showing ASA maintenance dose distribution](image)

The same is true for US patients but the distribution of ASA dose is very different\textsuperscript{1}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{chart}
\caption{Percentage of patients with different median ASA doses in the US region.}
\end{figure}

\textsuperscript{1}Kevin J Carroll & Thomas R Fleming (2013). \textit{SBR} Vol 5(2): 91-101, Supplm Appendix
Event rates increase for both Ticagrelor and Clopidogrel 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

1FDA 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\(^1\)

![Bar chart showing % primary events for Clopidogrel and Ticagrelor in different regions.]

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

- “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