Module 6
Evaluation of consistency

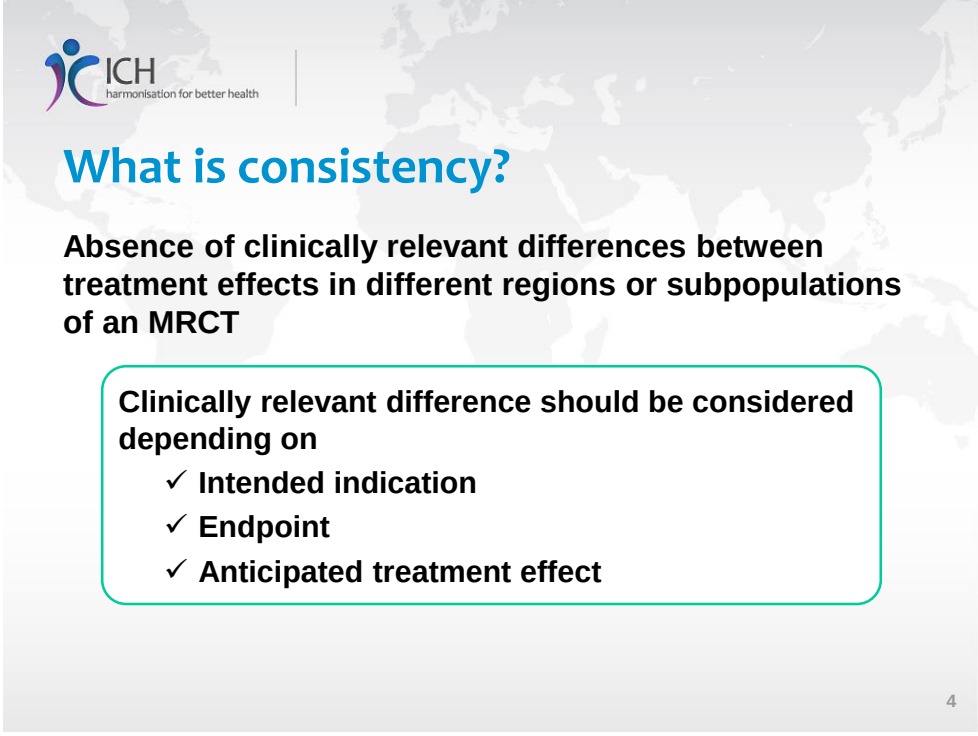
ICH E17: General principles for planning and design of Multi-Regional Clinical Trials

International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use

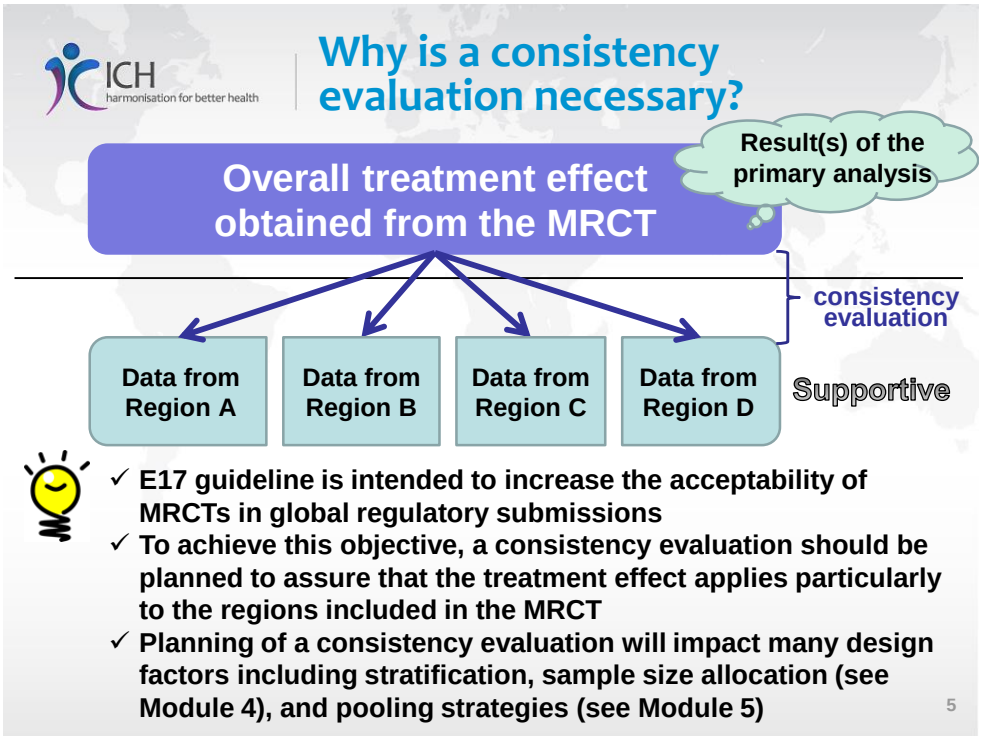
Legal Notice

Outline

- What is consistency?
- Why is a consistency evaluation necessary?
- How should a consistency evaluation be done in an MRCT?
- Examination of regional consistency
- Case study: the PLATO trial
- Concluding remarks



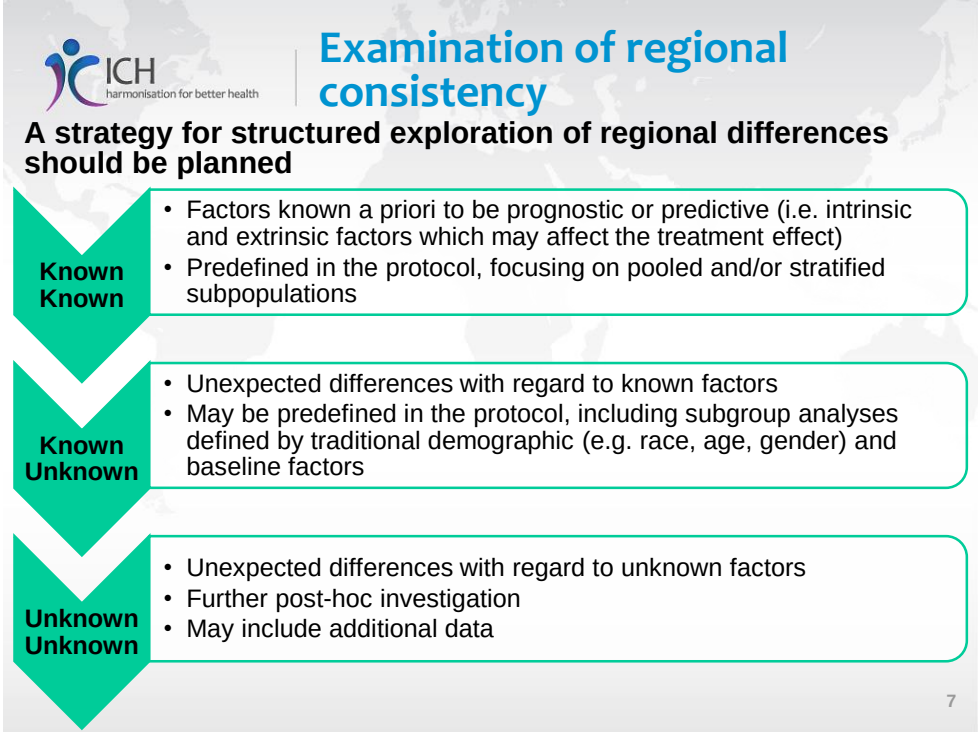
Evaluation of consistency examines the extent to which the overall treatment effect applies to the breadth of the trial population (see next slide).



A consistency evaluation in an MRCT, especially with an expected and/or unexpected inconsistency finding, offers an opportunity to understand the overall treatment effect and its potential modification by intrinsic/extrinsic factors across regions.

How should a consistency evaluation be done in an MRCT?

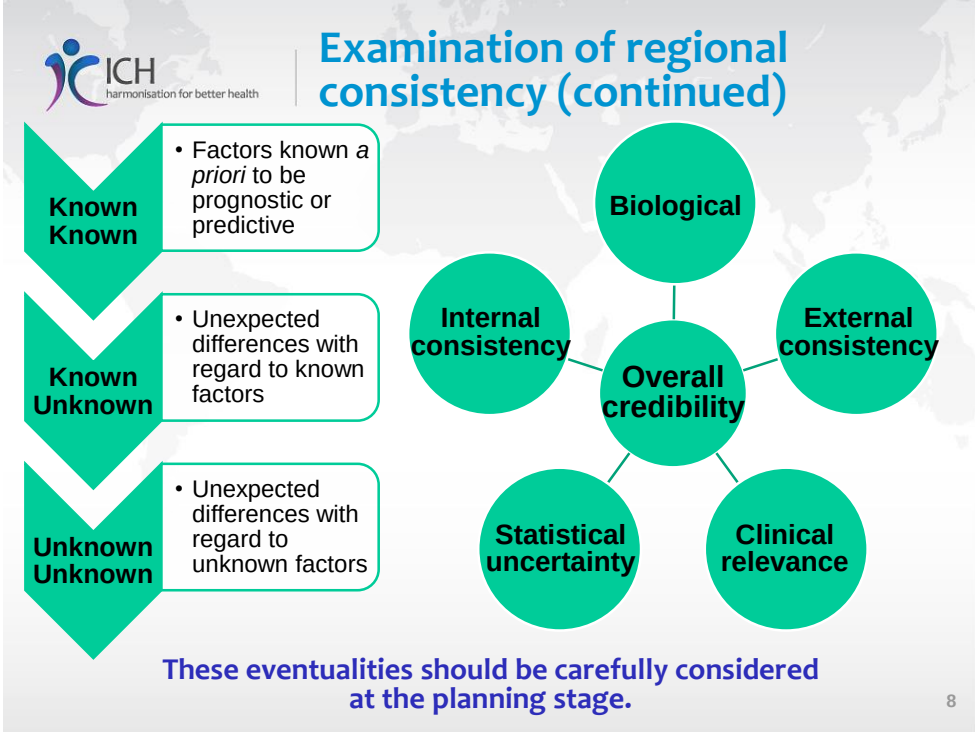
- **A structured exploration of regional differences should be planned**
  - Intrinsic and extrinsic factors which may affect the treatment effect (see Module 2) should be identified and evaluated by region
- **The potential eventualities of the trial results should be carefully considered at the planning stage**
  - These may include expected and/or unexpected potential differences across regions
  - This consideration is to ensure a comprehensive evaluation and to minimize unnecessary post-hoc analyses
- **Evidence for consistency of treatment effects across regions should be evaluated holistically**



This slide shows structured steps based on careful consideration of the intrinsic and extrinsic factors.

An example of an Unknown/Unknown is an unexpected finding despite careful planning (e.g., Aspirin dose in the PLATO trial).

When results are unexpected, a comprehensive holistic exploration of the plausibility of the finding is needed (see next 2 slides). Sometimes, additional studies to investigate may be needed (e.g., PEGASUS trial).

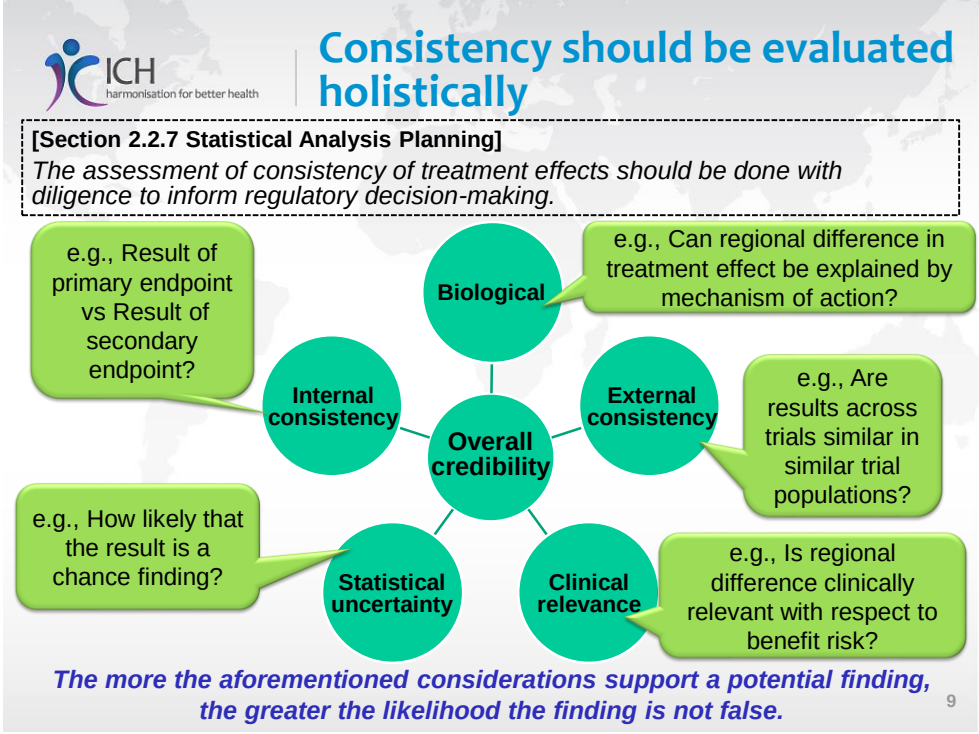


- Eventuality is to "keep the end in mind", to ultimately aim for robust evaluation of the overall treatment effect under a multi-regional context. With proper planning, the MRCT should be capable of answering questions on the overall treatment effect and the influence by intrinsic and extrinsic factors.

- This slide further explains the planning for the eventuality of study findings and the holistic evaluation

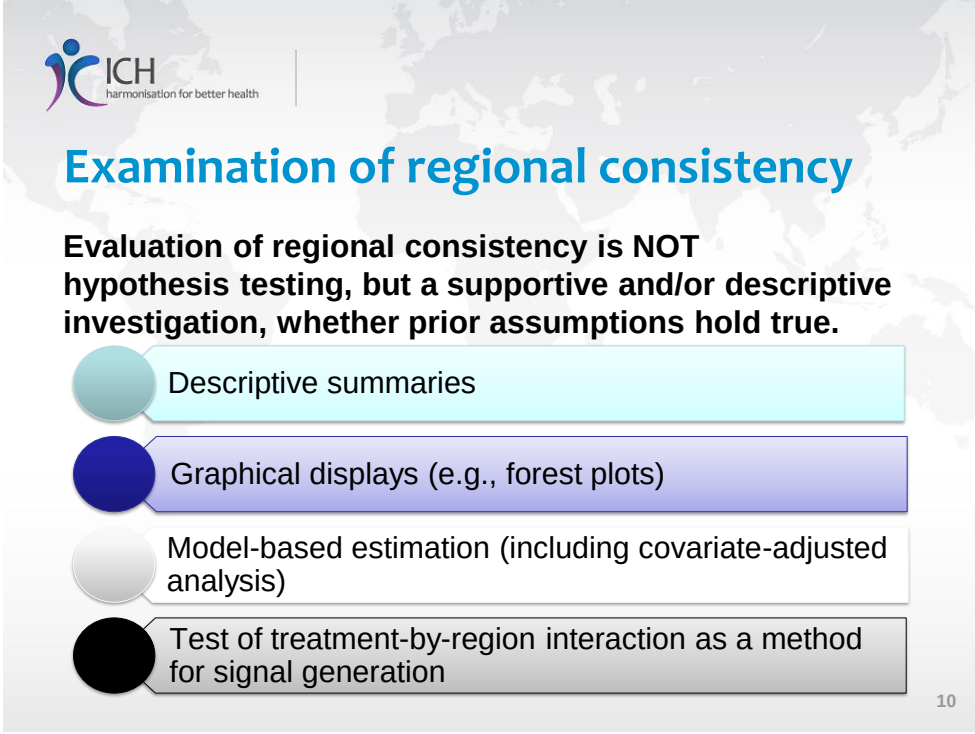
    - At the design stage, how to evaluate inconsistent findings in the end should always be considered. This involves consideration about the statistical likelihood of inconsistent findings due to chance. Proper sample size allocation and pooling across regions can reduce the impact of chance.

    - Eventually, if any inconsistency among regions is observed, it's important to evaluate whether this regional difference is clinically relevant. Assessment of clinical relevance will take regional context into consideration and evaluate whether the drug has positive benefit-risk balance.

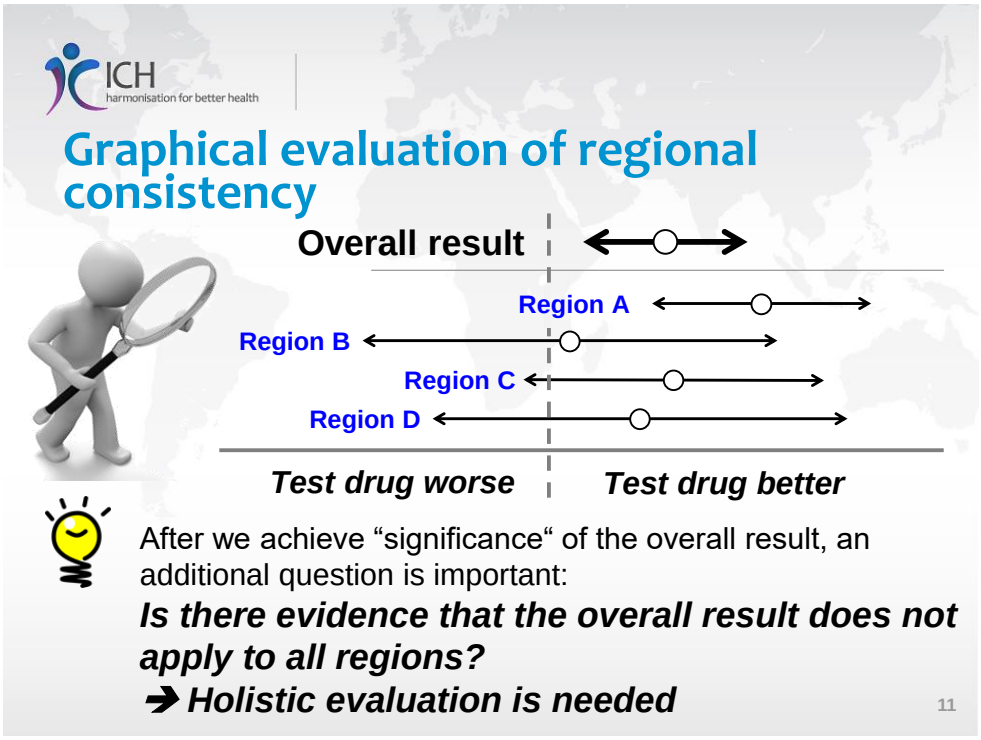


As part of the holistic evaluation of consistency, biological plausibility and replication (e.g., internal and external consistency) are key elements to examine

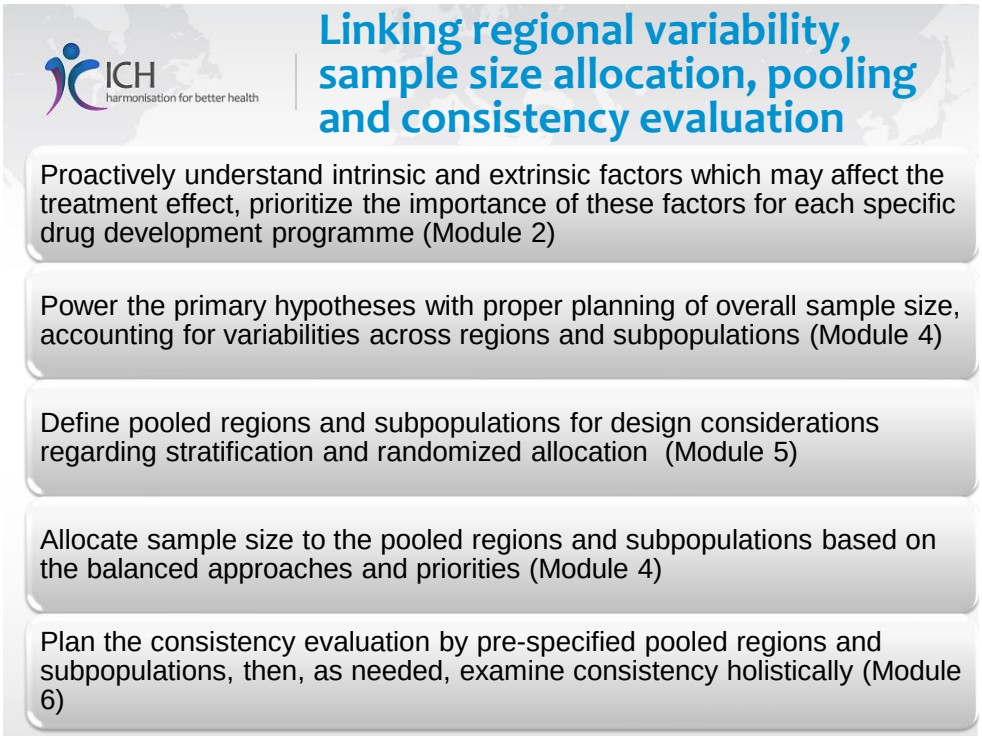
- **Biological plausibility** is a concept describing the extent to which a particular effect (in this case differential effects of treatment across regions or across subgroups) might be predicted, or might have been expected, based on clinical, pharmacological, and mechanistic considerations in association with intrinsic and extrinsic factors. Plausibility is primarily a clinical and pharmacological judgement and, unless already considered at the planning stage, is usually not a directly quantifiable or measurable concept
- **Replication links to internal consistency and external consistency.** In particular,
  - internal consistency refers to whether an impact on treatment effect by a particular factor is seen across multiple settings (e.g., endpoints, subgroups) in the MRCT trial
  - external consistency refers to whether an impact on treatment effect by a particular factor is seen in multiple data sources: specifically, whether an inconsistency in one confirmatory clinical trial is also seen in independent clinical trial data (another phase III trial, phase II exploratory trial, or a trial from outside the development programme, but with similar experimental conditions)



- The consistency evaluation can include descriptive summary, graphical display, model-based estimation and test of treatment-by-region interaction (i.e., p-value)
    - The graphical methods (e.g., Forest plot, funnel plot) can be effective tools
    - The model-based methods, including covariate-adjusted analyses and Bayesian-based Shrinkage estimates, may be helpful for a more comprehensive evaluation.
- See the case study on the PLATO trial for specific examples of these evaluation methods



- As an example, consistency of the MRCT results can be evaluated visually using a Forest plot
- Different results may be seen among regions. In the case of heterogenous results across regions, a holistic evaluation of the results is needed (see Slide 9).



Module 2 (Pre-considerations of regional variability), Module 4 (Sample size allocation), Module 5 (Pooling strategies) and Module 6 (Evaluation of consistency) are inter-related

Consistency evaluation in relation to regional variability, sample size allocation, and pooling

Consistency evaluation and Regional variability (Module 2)

- Identify and prioritize intrinsic and extrinsic factors, which may affect the treatment effect. This enables a structured evaluation of consistency
- The regional difference of treatment effect may be explained by difference in regional distribution of intrinsic and extrinsic factors

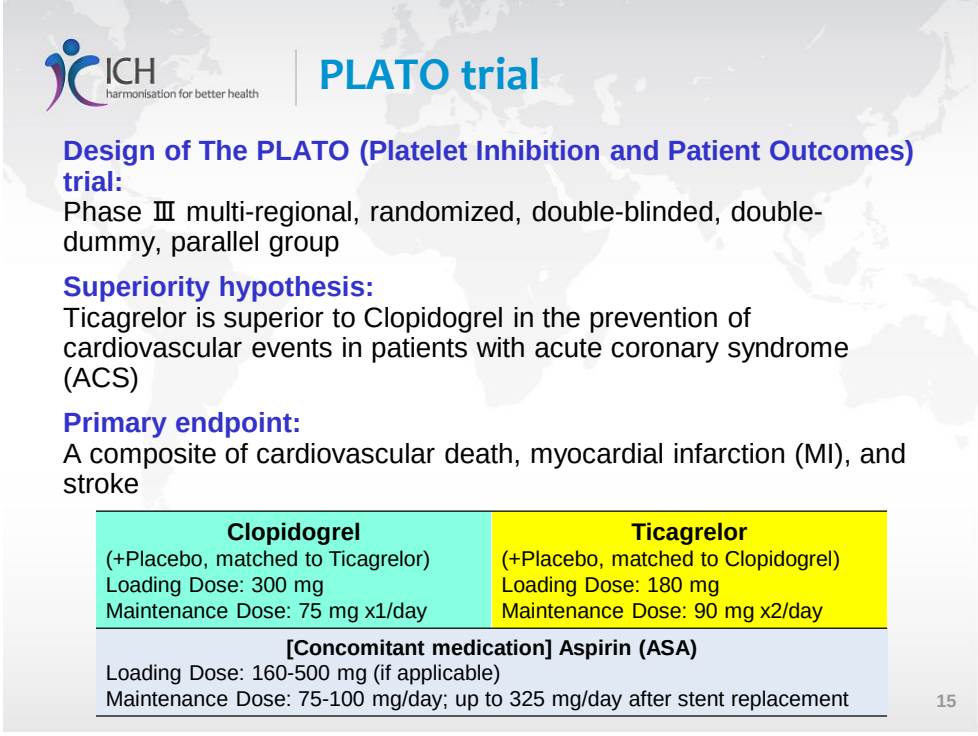
Consistency evaluation and Sample size allocation (Module 4)

- Proportional allocation to regions according to disease prevalence enables faster recruitment, while equal allocation optimizes the likelihood of detecting inconsistency; a balanced approach is needed
- Sample size allocation to a specific region based on preservation of effect or local significance may inflate overall sample size, and is not practical

Consistency evaluation and Pooling strategies (Module 5)

- The chance of inconsistent findings may increase with number of regions
- Pooling across regions based on intrinsic and extrinsic factors known to potentially affect the treatment effect may reduce the chance of such findings, but also reduces the chance of detecting true inconsistent findings
- It is important to balance these considerations

The PLATO case study illustrates how regional variations in observed treatment effects (benefit or risk) are handled, as well as how identification of intrinsic and extrinsic factors can potentially explain regional differences. The case also illustrates how extensive post-hoc analyses can be performed based on those factors. The primary principles that tie to this module are 4 and 5, but it touches upon other principles as well.



- The PLATO trial is a phase III multi-regional, randomized, double-blinded, double-dummy, parallel group trial.
- The efficacy hypothesis is Ticagrelor is Superior to Clopidogrel in the prevention of vascular events in patients with ACS.
- For more details please see: Wallentin, L. et al. (2009a), Ticagrelor Versus Clopidogrel in Patients With Acute Coronary Syndromes, New England Journal of Medicine, 361, 1045–1057.

Efficacy results, overall

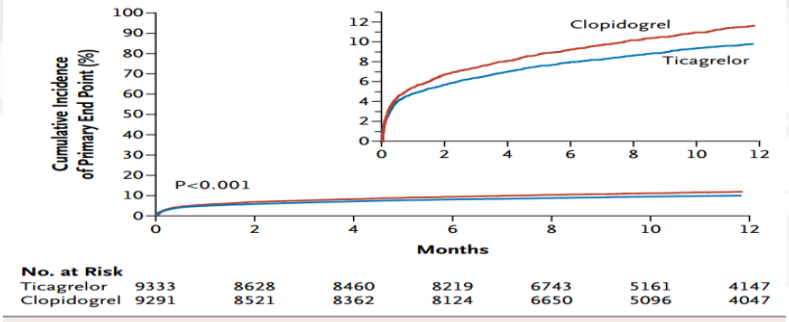


Figure 1. Cumulative Kaplan–Meier Estimates of the Time to the First Adjudicated Occurrence of the Primary Efficacy End Point.

		Cumulative Incidence of Endpoint, Ticagrelor vs Clopidogrel (%)
Primary Endpoint Met?	Yes	9.8% vs 11.7% (H.R. 0.85; 95% CI 0.77-0.92; p<0.001)
Main Secondary Endpoint Met? (Patients with planned surgery)	Yes	8.9% vs 10.6% (p=0.003)

Superiority of Ticagrelor to Clopidogrel was demonstrated

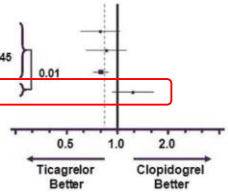
https://www.fda.gov/AdvisoryCommittees/Calendar/ucm214252.htm

16

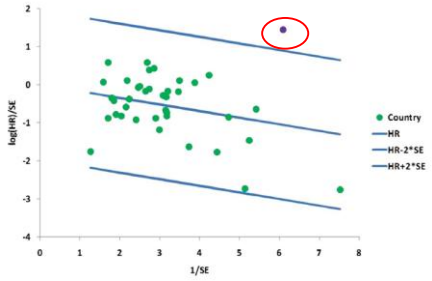
The overall result of the PLATO trial is significant in superiority.

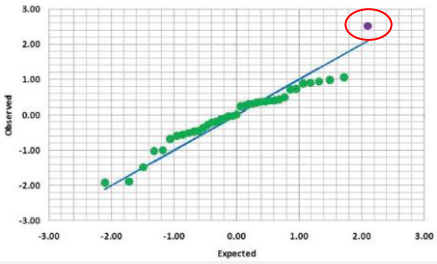
Efficacy results, by subgroup

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)	



The results of the subgroup analyses show the inconsistency across pre-defined regions





Kevin J. Carroll & Thomas R. Fleming (2013) Statistical Evaluation and Analysis of Regional Interactions: The PLATO Trial Case Study, Statistics in Biopharmaceutical Research, 5:2, 91-101.

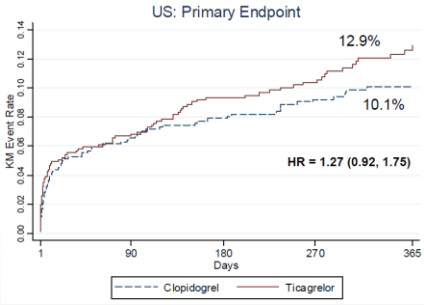
17

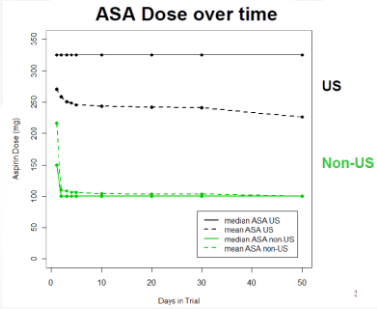
- The sponsor conducted 31 pre-specified subgroup analyses (including region) for consistency. The subgroup factors include demographics, baseline and concurrent medications
- Most of the analyses show effects consistent with the overall results, but there is a finding of heterogeneity by region.

Imbalances that might explain the US vs Non-US regional interaction

In the US, event rates in the Ticagrelor group start surpassing event rates in the Clopidogrel group by 150 days of treatment , the exact opposite of what is observed in the rest of the world.

The mean and median aspirin doses (ASA) throughout the trial were significantly higher in the US population than in the non-US populations.





https://wayback.archive-it.org/7993/20170405212359/https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/UCM221383.pdf

18

- In considering if there are imbalances that might explain the US vs Non-US regional interaction, differences in ASA maintenance dose were considered.
- In the US, 57% of patients received doses above 100 mg and 54% received doses above 300 mg. But the proportions receiving these doses in non-US population were 8% and 2%, respectively.

Internal consistency: impact of aspirin dose by regions

Region	ASA Dose (mg)	Ticagrelor N	Ticagrelor E	Clopidogrel N	Clopidogrel E	HR (95% CI)
US						
	> = 300	324	40	352	27	1.62 (0.99, 2.64)
	> 100 – < 300	22	2	16	2	
	< = 100	284	19	263	24	0.73 (0.40, 1.33)
Non–US						
	> = 300	140	28	140	23	1.23 (0.71, 2.14)
	> 100 – < 300	503	62	511	63	1.00 (0.71, 1.42)
	< = 100	7449	546	7443	699	0.78 (0.69, 0.87)

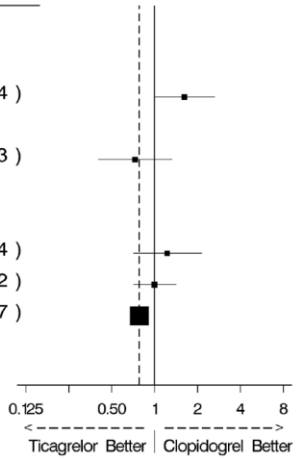


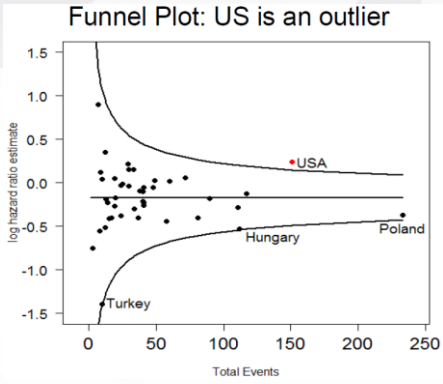
Figure 19 of the AZ briefing document

https://wayback.archive-it.org/7993/20170405212347/https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/UCM220197.pdf

19

Looking at the impact of aspirin dose by region, it is similar in the US and non-US.

Graphical evaluation of regional consistency: statistical uncertainty



A Play of Chance?

- Total primary events: 1878 (151 in US)
- Treatment-by-US interaction is significant (p=0.0095)
- 3 countries had estimated HR ≥ 1.27
  - Australia (N=92), Taiwan (N=83) and US (N=1413)
- P(HR≥1.27 in US | true HR=0.84) < 0.006

Although US is an outlier in the funnel plot, it is still possible that this observation could be due to play of chance.

https://wayback.archive-it.org/7993/20170405212359/https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/UCM221383.pdf (accessed May 19 2019)

20

- Although US is an outlier in the funnel plot, the likelihood of the US result as a chance finding was also evaluated
- Given the overall PLATO result, the chance of observing a HR>1.27 in US is less than 0.006.

Learning from the PLATO case study

- **Design:**
  - 4 geographical regions were pre-defined, pooling from 43 countries.
  - Sample size was somewhat balanced across regions, so that exploration of consistency was possible
  - Effect of Aspirin maintenance dose was not known *a priori*. If the use of Aspirin was known to be a potential predictive factor, stratification of the trial by the Aspirin dose or restriction to low dose Aspirin may have been considered.

- 31 of the pre-specified subgroup analyses show effects consistent with the overall results, except the treatment-by-region interaction.
- In PLATO trial, the sponsor evaluated consistency across regions using multiple approaches, such as descriptive summaries, graphical plots and test of treatment-by-region interaction.
- Exploratory analyses detected the unexpected finding regarding effect of Aspirin use.

Learning from the PLATO case study

- **Unfortunately, pre-specified analyses didn't reveal plausible reasons for the regional differences.**
- **A holistic evaluation was performed to better understand the observed regional heterogeneity**
  - Study Conduct
  - Internal Consistency: Impact of ASA dose, across US and non-US regions
  - Statistical Uncertainty (i.e., Play of chance)
  - Biological Plausibility

Epilogue of the PLATO trial

- **Special warning on Aspirin dose was described in the Ticagrelor label in various regions**
- **A subsequent PEGASUS trial demonstrated a similar treatment effect in the low dose aspirin maintenance group**
  - In patients who had prior MI 1-3 years earlier, ticagrelor (90 mg or 60 mg, twice daily) in combination with Aspirin (75-150 mg daily) significantly reduced the risk of CV death, MI, or stroke

Ticagrelor 90 mg vs. Placebo; % (HR, 95% CI; P-value)	Ticagrelor 60 mg. vs. Placebo; % (HR, 95% CI; P-value)
7.85 vs 9.04 (0.85, 0.75-0.96, p=0.008)	7.77 vs 9.04 (0.84, 0.74-0.95, p=0.004)

Bonaca MP, et al. "Long-term use of ticagrelor in patients with prior myocardial infarction". *The New England Journal of Medicine*. 2015. 372(19):1791-1800.

24

- ICH E17 does not cover guidance on study result interpretation and regulatory decision making
- These regulatory approval outcomes regarding the label of Ticagrelor are provided here as reference:
  - US label: Maintenance doses of aspirin above 100 mg reduce the effectiveness of Ticagrelor and should be avoided.
  - EMA label: A special warning "co-administration of Ticagrelor and high maintenance dose ASA (>300) is not recommended
  - China NMPA label: a special mentioning of Aspirin maintenance dose 75-100mg
  - Japan PMDA label: the drug was approved based on a local trial plus PLATO result. Special mentioning of with maintenance Aspirin dose (81-100 mg/day)

Concluding remarks

- **Evaluation of regional consistency is not hypothesis testing, but a key supportive analysis**
- **Planning of a structured consistency evaluation will require careful design considerations, including stratification, sample size allocation (see Module 4), and pooling strategies (see Module 5)**
- **A holistic evaluation of expected and/or unexpected inconsistencies in an MRCT offers opportunities to understand the overall treatment effect and intrinsic and/or extrinsic factors that modify this treatment effect across regions**
- **PLATO case study brings forth several important points to consider on design and analysis of MRCTs**