Introduction

In a clinical trial, a successful treatment for the overall population often produces different results in different patients. The participating patients have distinct characteristics leading to unavoidable heterogeneity in the response to treatment. The variability in the clinical response therapy is called “heterogeneity of treatment effect” (1-3). Many physicians are interested to know whether treatment effects vary across subgroups that are based on patient characteristics with the expectation that such knowledge will help to provide the best care for the individual patient.

Subgroup analyses within randomised controlled trials (RCTs) are conducted to demonstrate consistency of the overall treatment effect over different subgroups and to address safety concerns in specific patient subgroups (4, 5). As such, these analyses of trial data fall on fertile soil as physicians are seeking deeper knowledge of a new therapeutic agent.

After pivotal large clinical trials, clinicians are often faced with multiple publications of subgroup analyses. Intuitively, the attempt to identify subgroups for which an intervention is especially beneficial or harmful is clinically appealing. However, the interpretation of subgroup analysis findings and their implementation in treatment decision processes in clinical cardiology can be challenging (6-13). Indeed, it is increasingly recognised that subgroup analyses in clinical trials may sometimes receive undue emphasis and often raise concerns regarding the validity of the conclusions (8, 10, 14-16).

Statistical power for subgroup effects

Subgroup analyses almost always lack the power to detect subgroup effects (5, 8, 10, 17). Because subgroup-specific analysis...
requires the data to be subdivided into smaller data sets, each group has reduced power to detect a similar treatment effect. For example, we evaluate the effect of a new antiplatelet agent on the composite endpoint of cardiovascular mortality, myocardial infarction (MI) and stroke in a trial with significance level of 0.01 and power of 90%. We calculate that 8,288 patients divided equally between treatment and control groups are needed to detect a 25% reduction in the composite primary endpoint from 10% to 7.5% (▶ Figure 1). We subsequently perform an analysis to determine whether the positive trial results can be reproduced in the subgroup of patients with diabetes mellitus (DM). If patients with DM constitute 30% of the study population, the power for detecting the same 25% decrease in the primary endpoint would be only about 0.40 (▶ Figure 1). Because the number of trial participants is calculated detecting an overall treatment effect, it cannot be expected to detect similar effects within even relatively large subgroups (8, 10).

Because trials lack the power to detect subgroup effects, the best estimate of the true treatment effect within a specific subgroup is commonly the overall study result rather than the subgroup analysis (8, 18-22). The following example illustrates this point. In a report on the subgroup of patients with STEMI from Therapeutic Outcomes by Optimizing Platelet Inhibition with Prasugrel–Thrombolysis in Myocardial Infarction (TRITON–TIMI 38) trial, the authors report that the bleeding risk (TIMI major bleeding unrelated to CABG surgery) was similar with prasugrel and clopidogrel at both 30 days (hazard ratio [HR] 0.74, 95% confidence interval [CI] 0.39–1.38; p = 0.34) and at 15 months (HR 1.11, 95% CI 0.70–1.77, p = 0.65) (23). However, the STEMI subgroup (3,534 of the 13,608 study participants) may be underpowered to detect a meaningful difference in this safety outcome. Thus, the bleeding risk estimate calculated from the entire study population (with significantly higher bleeding risk with prasugrel) is more appropriate (8, 22).

### Presentation of subgroup analyses and interpretation of the forest plot

The conventional method to present treatment heterogeneity is to divide the study participants based on potential clinically relevant characteristics, with the use of forest plots to graphically display treatment effects across subgroups (1, 24-26). CIs in subgroups are wider than those for the main effect because the group size is smaller. When the CI for a subgroup crosses the no effect point, a common misinterpretation is that there is no significant treatment effect in the subgroup even if the overall study results are positive. This conclusion is incorrect because it uses within subgroup comparison to draw inferences about between subgroup differences (14, 15, 27).

▶ Figure 2 shows the results of two subgroup analyses from the Study of Platelet Inhibition and Patient Outcomes (PLATO) trial (28) based on the presence or absence of ST-segment elevation myocardial infarction (STEMI) and of DM. Assessing the effects of ticagrelor versus clopidogrel on a single subgroup by asking whether the 95% CI for subgroup meets the threshold for statistical significance, would lead to the false claim of relative lack of benefit of ticagrelor compared with clopidogrel in the subgroups of STEMI patients and in patients with DM. Cuzick pointed out that this presentation focuses attention on whether or not the CI for an individual subgroup touches the no effect point (e.g. a HR of 1.0) (24).

The degree to which subgroup-specific CIs intervals overlap in ▶ Figure 2 suggests that the treatment effects do not differ. The effect size between patients with and without STEMI (relative risk 0.87 with and 0.83 without) and with and without DM (0.88 and 0.83, respectively) indicates that the efficacy of ticagrelor is similar in the subgroups. The interactions between STEMI and DM with ticagrelor were not statistically significant (p=0.68 and p=0.49, respectively) (28).

### Heterogeneity of treatment effects and tests for interaction

Treatment heterogeneity is said to occur when a difference between the size of the therapeutic effects varies across subsets (8, 29, 30). A test for interaction determines whether the treatment effect noted in the whole cohort is homogeneous across subgroups or shows a variation too large to be consistent with chance alone (8, 19, 30-34). In our example (▶ Figure 2) we test the inequality of the two HRs (i.e. that the HR for ticagrelor relative to clopidogrel differs for patients with or without STEMI or with and without DM). The tests for interaction for these two subgroup analyses are clearly not significant, indicating that no evidence of heterogeneity exists for the efficacy of ticagrelor as compared with clopidogrel with regard to baseline electrocardiography (ECG) findings or DM status.

An interaction is termed quantitative (non-crossover) when the direction of the new treatment effects in the different subgroups is the same (e.g. superior for both subgroups) but the magnitude of effect varies among subgroups. Quantitative interactions occur...
more commonly, and are generally of lesser clinical importance because differences inherent among patients may influence the degree of response to treatments. The sicker patients tend to have a greater benefit from treatment than those at lower absolute risk (8, 35). Furthermore, the trial results can be generalised because if the overall treatment effect is positive, patients should not be denied potentially beneficial therapy because they belong to a subgroup that appears to derive smaller benefit (27, 36, 37).

A qualitative (crossover or “effect reversal”) interaction is one in which treatment is beneficial in some subgroups but harmful in others (8, 15, 29, 38). The identification of qualitative interactions is obviously of greater clinical importance. Qualitative interactions are less common in clinical trials, because trial participants with the same diagnosis share many similarities that make such an interaction unlikely (8, 35, 39) but can also occur by chance alone (40). A notable example is the TRITON–TIMI 38 trial. In this trial there was a significant qualitative interaction between a history of cerebrovascular events and the degree of net clinical benefit of prasugrel as compared with clopidogrel, indicating a significant harm from prasugrel among patients with a history of cerebrovascular

![Figure 2: Forest plot of subgroup analysis.](image)

Hazard ratios (black squares), 95% CIs (horizontal lines), p-values are for the interaction between the treatment effect and subgroup variables. Bold line shows overall treatment effect and the dotted line shows the no effect point. Data was extracted from the PLATO trial (28).

![Figure 3: Forest plot of subgroup analysis with heterogeneity of treatment effect.](image)

Hazard ratios (black squares), 95% CIs (horizontal lines), p-values are for the interaction between the treatment effect and subgroup variables. Bold line shows overall treatment effect and the dotted line shows the no effect point. Estimates in patients with and without the characteristic of interest on opposite sides of the dotted line suggest heterogeneity of treatment effect. Data was extracted from the TRITON–TIMI 38 trial (41).
events, as compared with a significant benefit from prasugrel among patients without such a history (▶Figure 3) (41).

Detection of an interaction of the same magnitude as the overall treatment effect requires a four-fold greater sample size, and sometimes much more (e.g. when the subgroups are not equal in size) (10, 42). Therefore, a true differential treatment effect across subgroups may be missed (8, 43).

Because interaction tests are underpowered, especially for categorical variables (44), the possibility of heterogeneity may be considered when the interaction p-value is between 0.1 and 0.01 (45, 46). ▶Table 1 demonstrates that most of the interactions found in clinical trials are in this range. However, interactions are much more credible when p-values reach 0.001 or less (39, 47).

In addition to p-values, it is preferable to report the interaction HR (i.e. the ratio of the HRs for benefit of treatment versus control in the subgroup of interest compared with those not in the subgroup (48). The magnitude of the treatment-effect interaction HR

<table>
<thead>
<tr>
<th>Study</th>
<th>Patients</th>
<th>Antplatelet therapy</th>
<th>Overall efficacy</th>
<th>No. interactions/No. subgroup analyses</th>
<th>Significant or borderline subgroup-by treatment interactions</th>
<th>P int</th>
<th>Direction of Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>CURRENT–OASIS 7 (82)</td>
<td>25,086 patients with ACS referred for an invasive strategy</td>
<td>Double-dose vs standard dose clopidogrel</td>
<td>Neutral</td>
<td>3 of 14</td>
<td>1) PCI after randomisation</td>
<td>0.03</td>
<td>(QN) favours double-dose</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2) GRACE score &lt; 140</td>
<td>0.09</td>
<td>(QN) favours double-dose</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3) Clinically significant coronary artery disease</td>
<td>0.06</td>
<td>(QN) favours double-dose</td>
</tr>
<tr>
<td>TRITON–TIMI 38 (41)</td>
<td>13,608 patients with ACS</td>
<td>Prasugrel or clopidogrel</td>
<td>Prasugrel better</td>
<td>2 of 8</td>
<td>1) Diabetes mellitus</td>
<td>0.09</td>
<td>(QN) favours prasugrel</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2) History of stroke or TIA</td>
<td>0.02</td>
<td>(QN) favours clopidogrel</td>
</tr>
<tr>
<td>PLATO (28, 83)</td>
<td>18,624 patients with ACS</td>
<td>Ticagrelor or clopidogrel</td>
<td>Ticagrelor better</td>
<td>3 of 33</td>
<td>1) Weighing &gt; median</td>
<td>0.04</td>
<td>(QN) favours ticagrelor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2) Lipid-lowering drugs at randomisation</td>
<td>0.04</td>
<td>(QN) favours ticagrelor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3) Enrolled in North America</td>
<td>0.05</td>
<td>(QN) favours ticagrelor</td>
</tr>
<tr>
<td>Trilogy-ACS (72)</td>
<td>7,243 patients with NSTEMI-ACS treated medically</td>
<td>Prasugrel or clopidogrel</td>
<td>Neutral</td>
<td>3 of 17</td>
<td>1) Current or recent smoker</td>
<td>&lt;0.0001</td>
<td>(QN) favours prasugrel</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2) Angiography before randomisation</td>
<td>0.08</td>
<td>(QN) favours prasugrel</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3) PPI at randomisation</td>
<td>0.02</td>
<td>(QN) favours prasugrel</td>
</tr>
<tr>
<td>CHAMPION PHOENIX (84)</td>
<td>11,145 patients undergoing Urgent or elective PCI</td>
<td>IV cangrelor or clopidogrel</td>
<td>Cangrelor better</td>
<td>1 of 20</td>
<td>1) History of PAD</td>
<td>0.003</td>
<td>(QN) favours cangrelor</td>
</tr>
<tr>
<td>TRA 2P–TIMI 50 (85)</td>
<td>26,449 patients with history of myocardial infarction, ischaemic stroke, or peripheral arterial disease</td>
<td>Vorapaxar or placebo</td>
<td>Neutral</td>
<td>3 of 11</td>
<td>1) Body weight ≥60 Kg</td>
<td>0.03</td>
<td>(QN) favours vorapaxar</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2) Qualifying Atherosclerosis (MI)</td>
<td>0.058</td>
<td>(QN) favours vorapaxar</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3) Aspirin at Enrollment</td>
<td>0.092</td>
<td>(QN) favours vorapaxar</td>
</tr>
<tr>
<td>TRACER (86)</td>
<td>12,944 patients with NSTEMI-ACS</td>
<td>Vorapaxar or placebo</td>
<td>Neutral</td>
<td>1 of 22</td>
<td>1) No intent to use Direct thrombin inhibitors</td>
<td>0.044</td>
<td>(QN) favours vorapaxar</td>
</tr>
<tr>
<td>ACCCOAST (87)</td>
<td>4,033 patients with NSTEMI-ACS scheduled to undergo coronary angiography</td>
<td>Prasugrel pre-treatment vs placebo</td>
<td>Neutral</td>
<td>1 of 15</td>
<td>1) Time from symptoms onset to loading dose &lt;median</td>
<td>0.004</td>
<td>(QL) favours pre-loading</td>
</tr>
</tbody>
</table>

* Interactions are for the primary endpoint of the studies. Because of the low power of interaction tests a p < 0.1 was taken as potentially significant. Note that interactions also occur in neutral trials. QN = quantitative interaction; QL = qualitative interaction.
Aronson: Subgroup analyses of antiplatelet agents

indicates the magnitude of estimated difference in the intervention effect between the subgroup.

Adjusting for multiplicity

Large clinical trials that perform many subgroup analyses generate multiple comparisons that increase the likelihood of false-positive results (type 1 errors) (5, 8, 10, 13, 16, 32, 49). For example, the probability of finding at least one “statistically significant” false positive interaction test when 10 independent interaction tests are undertaken is 40%. If 20 independent comparisons are made, the chances increase to 64% (15, 50-52).

Table 1 depicts the number and direction of subgroup-by-treatment interactions in several recent trials of new antiplatelet agents. In these trials, the clinical setting, inclusion criteria and primary endpoint may differ somewhat, but the rational (that a more potent antiplatelet agent reduces events) is the same. Therefore, subgroup analyses of one trial might be expected to be supported by results from similar trials. However, there is no consistent pattern of effect modification across studies of potent antiplatelet agents. Evidence for treatment heterogeneity in one study is almost always absent in other studies, with an overall pattern consistent with random variation (12.8% of the interactions are positive at the <0.1 level; Table 1). There is no straightforward way to determine which of the observed subgroup-by-treatment interactions were due to chance (resulting from multiple testing) or represent a true finding (53).

Reducing the number of comparisons in a trial is probably the best solution for multiplicity. There are several methods to correct for the inflated false positive rate when multiple subgroup analyses are conducted (15, 50), but these are applied infrequently even when the clinical implications are considerable (45).

Prespecified subgroup analyses

The credibility of subgroup analyses is improved if confined to a few predefined subgroups to minimise false positive results (1, 7, 10). Prespecified subgroup analyses should be explicitly labeled in the study protocol, with exact definitions of the subgroup variables and a specified direction of the subgroup effect (8, 39).

In large clinical trials multiple subgroup analyses are prespecified in the study protocol (e.g. 20 or more) (21, 28, 54). When subgroup analyses are pre-specified, the total number of the analyses is known, and adjustment for multiplicity can be made. However, not all preplanned analyses are motivated by specific hypothesis that are based on biologically plausible rationale or prior evidence as being predictors of response to intervention. Furthermore, there is no emphasis on a smaller set of analyses for which a logical rationale exists. Therefore, even prespecified subgroup analyses are essentially exploratory in nature, and should only affect the conclusions drawn from the trial under exceptional circumstances (42).

Post-hoc subgroup analyses

Whereas some of the analyses are specified a priori, others are developed on a post-hoc basis. The number of potential candidate post-hoc hypotheses is large, thus inflating the false positive rates. Other concerns relate to hypotheses conceived after the data has been revealed (8, 15, 16, 39, 48, 55).

Another problem relates to groups of patients characterised by a variable measured after randomisation (also called improper subgroups) because the tested interventions might create a prognostic imbalance between groups (8). An example of such analysis is a recent report of patients in TRITON–TIMI 38 who underwent coronary artery bypass grafting (CABG) (post-randomisation variable) at any time during the study. All-cause mortality (which was not an endpoint of the study) was 2.31% in the prasugrel cohort compared with 8.67% in the clopidogrel cohort, with an adjusted odds ratio (OR) for mortality of 0.26 [95% CI 0.08 to 0.85; p = 0.025] (56). The large reduction in mortality is in stark contrast to the whole study results and occurred despite increased bleeding and transfusions with prasugrel.

A better approach for handling post randomisation variables (such as revascularisation) was employed in PLATO. The treating physician designated patients as planned for initial invasive management or initial conservative management at the time of randomisation. These designations were non-binding (57, 58) but the pre-hoc assignment of patient to the non-invasive or invasive group prevented the bias associated with the decision to perform percutaneous coronary intervention (PCI) after randomisation. The reductions of ischaemic events with ticagrelor over clopidogrel in patients managed non-invasively were consistent with those from the overall PLATO results.

Replication of subgroup treatment effects

Subgroup findings should ideally be expected on the basis of prior findings or independently confirmed in similar clinical scenarios (16, 19). Although confirmation is highly desirable, future replication remains uncommon.

The TRITON–TIMI 38, the benefit of prasugrel tended to be greater among the 3146 patients with DM [17.0% of whom had the primary end point (death/nonfatal MI/stroke) in the clopidogrel group, vs 12.2% in the prasugrel group; HR, 0.70; 95% CI, 0.58 to 0.85; p<0.001] than among the 10,462 patients without diabetes (10.6% of whom had the primary end point in the clopidogrel group, vs 9.2% in the prasugrel group; HR, 0.86; 95% CI, 0.76 to 0.89; p = 0.02). There was no significant interaction between treatment effect and diabetes status (p = 0.09) (41).

In a subsequent analysis the study investigators reported a quantitative differences such that when both ischaemic events and bleeding were integrated into a net clinical benefit composite end point (the composite of all-cause mortality, non-fatal MI, nonfatal stroke, or non-fatal TIMI major bleeding not related to CABG), a significantly greater relative improvement was observed with prasugrel in subjects with DM than without DM (45). Specifically,
among subjects without DM, a non-significant 8% reduction in the composite of all-cause mortality, non-fatal MI, non-fatal stroke, or nonfatal major bleeding (p=0.16) was observed, whereas a statistically greater 26% reduction in this composite outcome was seen for patients with DM. An unadjusted test for interaction indicated borderline significance (p=0.001; for interaction = 0.05) (45). Thus, a nonconservative interpretation might indicate greater benefit of prasugrel in the DM subgroup.

Given high platelet reactivity levels in patients with DM (59), the interaction between diabetes and the more potent antiplatelet effect of prasugrel makes biological sense. However, even a significant interaction test should still be interpreted with caution, as there are numerous examples from randomised trials in which an apparently important differential response to therapy introduced by a subgroup analysis generated a hypothesis that was subsequently refuted in a trial designed to test that hypothesis (12, 60–62).

For example, in a meta-analysis of six randomised, double-blind, placebo-controlled trials of intravenous platelet glycoprotein (GP) IIb/IIIa antagonists in non-ST-elevation acute coronary syndrome (ACS), involving a total of 6,458 diabetic and 23,072 non-diabetic patients, platelet GP IIb-IIIa inhibition as compared with placebo was associated with a significant reduction of 30-day mortality in diabetic patients (p for interaction = 0.036) (63). These results led to a strong recommendation for the use of platelet GP IIb/IIIa inhibitors in diabetic patients with ACS. However, in a subsequent dedicated randomised trial of patients with diabetes, the potent antiplatelet GP IIb/IIIa inhibitor abciximab did not reduce the cumulative risk of death or MI in patients undergoing elective PCI (64). Similar results were obtained in patients with NSTE-ACS (65).

There is no prospective confirmation of the interaction between antiplatelet therapy with prasugrel and DM. Therefore, it is helpful to place the TRITON–TIMI 38 data in the broader context of intensified antiplatelet therapy in DM. Other studies comparing older antiplatelet agents with newer, more potent antiplatelet agents showed no heterogeneity with regard to their effect in patients with DM in the setting of ACS, stable coronary disease or post PCI (Table 2).

### Biological plausibility

In PLATO, although the overall treatment difference is clearly in favor of ticagrelor, there was a tendency in the opposite direction for patients in North America (p for interaction = 0.045; and p = 0.01 for the U.S./combined non-U.S. interaction). Additional extensive post-hoc analyses identified aspirin dosage as a likely explanation for the treatment-by-region interaction. Aspirin dose was strongly imbalanced between US and rest of the world patients, with 53.6% of patients in the US receiving median maintenance dose of ≥300 mg/day as compared with 1.7% in the rest of the world. The HR for the primary endpoint was 1.45 (95% CI, 1.01 to 2.09) favouring clopidogrel for maintenance aspirin dose ≥300 mg/day and 0.77 (95% CI, 0.69 to 0.86) favoring ticagrelor for a

### Table 2: Clinical trials of antiplatelet agents of patients with ACS or undergoing PCI: treatment-by-DM interaction.

<table>
<thead>
<tr>
<th>Study</th>
<th>Patients</th>
<th>Antiplatelet therapy</th>
<th>Overall efficacy* (HR, 95% CI)</th>
<th>DM subgroup analysis (HR, 95% CI)</th>
<th>P int</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRITON–TIMI 38 (41)</td>
<td>13,608 patients with ACS</td>
<td>Prasugrel or clopidogrel</td>
<td>0.81 [0.73–0.90]</td>
<td>DM: 0.70 [0.58–0.85] No DM: 0.86 [0.76–0.98]</td>
<td>0.09</td>
</tr>
<tr>
<td>CURE-PCI (88)</td>
<td>2,658 patients with NSTEMI-ACS</td>
<td>Clopidogrel or placebo</td>
<td>0.70 [0.50–0.97]</td>
<td>DM: 0.77 [0.48–1.22] No DM: 0.66 [0.50–0.87]</td>
<td>NS</td>
</tr>
<tr>
<td>PLATO (28, 83)</td>
<td>18,624 patients with ACS</td>
<td>Ticagrelor or clopidogrel</td>
<td>0.84 [0.77 to 0.92]</td>
<td>DM: 0.88 [0.76–1.03] No DM: 0.83 [0.74–0.92]</td>
<td>0.49</td>
</tr>
<tr>
<td>ISAR-SWEET (64)</td>
<td>701 patients with DM undergoing elective PCI</td>
<td>Abciximab or placebo</td>
<td>0.97 [0.58–1.62]</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>ISAR-REACT 2 (65)</td>
<td>2,022 patients with NSTEMI-ACS</td>
<td>Abciximab or placebo</td>
<td>0.75 [0.58–0.97]</td>
<td>DM: 0.90 [0.56–1.60] No DM: 0.71 [0.52–0.95]</td>
<td>0.30</td>
</tr>
<tr>
<td>Trilogy-ACS (72)</td>
<td>7243 patients with NSTEMI-ACS</td>
<td>Prasugrel or clopidogrel</td>
<td>0.91 [0.79–1.05]</td>
<td>DM: 0.90 [0.73–1.09] No DM: 0.94 [0.77–1.16]</td>
<td>0.71</td>
</tr>
<tr>
<td>CHAMPION PHOENIX (84)</td>
<td>11,145 patients undergoing Urgent or elective PCI</td>
<td>IV cangrelor or clopidogrel</td>
<td>0.78 [0.66–0.93]</td>
<td>DM: 0.92 [0.67–1.27] No DM: 0.74 [0.61–0.90]</td>
<td>0.26</td>
</tr>
<tr>
<td>TRA 2P–TIMI 50 (85)</td>
<td>26,449 patients with history of myocardial infarction, ischaemic stroke, or peripheral arterial disease</td>
<td>Vorapaxar or placebo</td>
<td>0.87 [0.80, 0.94]</td>
<td>DM: 0.89 [0.78, 1.02] No DM: 0.85 [0.77, 0.95]</td>
<td>0.61</td>
</tr>
<tr>
<td>ACCOAST (87)</td>
<td>4,033 patients with NSTE-ACS scheduled to undergo coronary angiography</td>
<td>Prasugrel pretreatment vs placebo</td>
<td>1.02 [0.84 to 1.25]</td>
<td>DM: 1.25 [0.81–1.93] No DM: 0.97 [0.78–1.21]</td>
<td>0.30</td>
</tr>
</tbody>
</table>

* Hazard ratio for the primary endpoint.
maintenance aspirin dose <300 mg/day (p for interaction = 0.0006) (66). The authors cite experimental studies and raise potential hypotheses to explain why higher aspirin doses may attenuate the treatment effect of ticagrelor via the reduction in endogenous prostacyclin production (66).

Observed treatment heterogeneities are likely due to chance when they are lacking mechanistic support by the literature (8). Therefore, authors often construct a potential mechanism as supportive evidence for an unexpected interaction. When assessing potential biologic explanations for subgroup findings, it is important to realise that in retrospect almost any subgroup finding can be justified on the basis of a theory (19, 53, 67). Ware noted that "the human imagination seems capable of developing a rationale for most findings, however unanticipated" (68). Altman shares a similar view that "doctors seem able to find a biologically plausible explanation for any finding" (69). Finally, Barrett-Connor writes that "biological plausibility is quite easy to theorise; anyone with 2 hours and a little imagination can do it" (53). Notwithstanding, most of the proposed mechanisms remain unconfirmed as is their ability to fully explain large treatment heterogeneities. In the current example, the biological mechanism must be powerful enough not only to nullify the beneficial effects of ticagrelor relative to clopidogrel, but also to worsen outcome with ticagrelor use. The acceptance of this qualitative interaction is strongly dependent on the existence of an underlying biological rationale for effect reversals regardless of the strength of the statistical analysis (27).

The authors of the PLATO treatment-by-region interaction report appropriately considered their finding with skepticism and concluded that "chance alone rather than aspirin dose must remain a potential explanation for the geographic inconsistency in the PLATO results" (66). However, these findings led the Food and Drug Administration (FDA) to issue a boxed warning 'use of ticagrelor with aspirin doses exceeding 100 mg/day decreases its effectiveness'. Interestingly, apparent geographic variation in treatment effect occurred in the ISIS trials but were not reported because of the possibility of harm caused by misinterpretation of such data (i.e. failure to use a highly effective treatment in a certain country) (18).

### Subgroup analyses following neutral trials

Subgroup analyses are commonly performed in RCTs showing no overall difference between treatments (6, 70) (Table 1). The rationale for these analyses is to identify subgroups that might be benefited from treatment when the overall results show no benefit. The temptation to perform these analyses is high because clinicians are unlikely to consider the intervention for any patients unless a benefit in a subgroup of patients is demonstrated. In this setting of neutral overall neutral trial results true positive subgroup effects are even less plausible (10, 51, 53, 70, 71), due to the increase of the type I error rate (49, 51).

The Targeted Platelet Inhibition to Clarify the Optimal Strategy to Medically Manage Acute Coronary Syndromes (TRILOGY ACS) showed no overall benefit of prasugrel in ACS patients who do not undergo revascularisation (72), and on-treatment platelet reactivity was not strongly related to ischaemic outcomes in these patients (73). However, a subgroup analysis suggested benefit

---

### Design and analysis

- Subgroups analyses should be preplanned and clearly defined in the study protocol
- Subgroups created based on variables measured after randomisation may lead to bias
- Only one outcome should be used for subgroup analyses, preferably the primary endpoint
- Investigators must report all subgroup analyses done
- Subgroups analyses should be limited to a small number of clinically important questions
- The only reliable statistical approach is to evaluate heterogeneity in treatment effect is a test for interaction

### Interpretation

- Reports of the significance of the effect of treatment in individual subgroups should be ignored
- Assess magnitude of interaction. Does interaction test suggests a low likelihood that chance explains the apparent subgroup effect?
- A large number of preplanned interaction tests increases the likelihood of spurious findings
- Is the magnitude of the difference in treatment effect clinically important?
- Is there a strong biological rational that supports the difference in treatment effect?
- Assess the results in the context of similar data from other trials. The validity of subgroup-treatment effect interactions increases if reproducible in other trials
- Consider subgroup analysis as hypothesis-generating finding to stimulate further research

*Modified from references (1, 7, 12, 32, 36, 39, 47, 48, 61, 89–92). See individual references for detailed criteria.

---

Table 3: Criteria for conducting and assessing the credibility of subgroup analyses*.
among patients who were treated after coronary angiography (p for interaction = 0.08). The beneficial effect of prasugrel in the subgroup was particularly prominent for the endpoint of stroke (46). These results are at variance with the lack of additional positive effect of prasugrel over clopidogrel with regard to the endpoint of stroke in other studies (74).

Criteria for performing subgroup analyses and assessing their credibility

Guides to help recognise methodological weaknesses of subgroup analyses are available, as well as criteria that can systematically be applied in the interpretation of subgroups effects (Table 3). The more criteria subgroup analyses fulfill the credibility of a putative subgroup effect increases. However, more lenient statistical standards may be used when investigating potential treatment-related harm (27). One example is the post-hoc exploratory analyses in the TRITON-TIMI 38 trial which identified significant harm from prasugrel among patients with a history of cerebrovascular events (41).

FDA regulations require sponsors of new drug applications to present a summary of safety and effectiveness data by demographic subgroups (age, gender, and race) (75). Regulatory authorities also frequently explore regional subgroups because medical practice patterns and the genetic pool of the population studied may produce difference in the magnitude of a treatment effect (5, 67, 76). Beyond these standard analyses, the appropriate subgroup analyses to be performed are likely dependent on the specific intervention being tested, but should be motivated by specific hypothesis based on biologically plausible rational. Testing for effect modifiers related directly to drug metabolism or action (77) are preferable over looking at the effects of simple patients characteristics. Furthermore, individual clinical characteristics in isolation almost never act directly to alter the effects of treatments. Rather, the joint effect of multiple clinical variables determines the patients’ baseline risks of adverse events and may underlie heterogeneity in treatment benefit (27, 78-80). RCTs patients can be categorised into several ordered risk subgroups based on a prespecified, externally validated prognostic model involving multiple baseline characteristics (80, 81). Differential responsiveness to therapy can be subsequently estimated in each group with tests of interaction between treatment effect and risk strata (78-81). For example, in 3 RCTs of antithrombotic treatments for ACS stratified by the TIMI risk score, the intervention was of no benefit in low-risk patients but was highly beneficial in high-risk cases (80). Several authors advocated this approach as an alternative to conventional subgroup analysis (78-81).

Conclusion

The primary interest in RCTs of new antiplatelet agents is the effects of a therapeutic intervention on the entire study population. The sponsor’s interest is to enlarge the number of candidates for treatment by including a broad range of patients with coronary disease, which may increase heterogeneity. Clinicians are often concerned that the average treatment response may not adequately represent the varying impact across subgroups classified on the basis of clinical characteristics. The expectation that the overall benefit of a new treatment will not apply equally to all patients is obvious, and as such, subgroup analyses offer the enticing possibility of refined individual patient care. However, the observed treatment effect is also expected to vary across subgroups, simply through random variation.

Large clinical trials of antiplatelet agent often report the results of multiple subgroup analyses to demonstrate the consistency of treatment effect rather than because many baseline characteristics are perceived as potentially influential on the treatment effect. Most of these subgroup analyses simply confirm the result of the overall trial in a specific subgroup. However, when an unexpected heterogeneity occurs, the results are frequently not tested is future studies and remain unproven. Because the hypotheses raised by subgroup analyses may be interesting and highly clinically relevant, clinicians must either accept or reject the findings based on clinical judgment in the absence of confirmatory data, with the risk of inappropriate uptake into clinical practice. Even when an apparent interaction and putative biologic rationale are present, a random effect is more likely than a true underlying pathophysiologic entity. For these reasons, it is important to not over interpret the results of subgroup analyses as sound basis for clinical decision-making.

Conflict of interest

None declared.

References


