Subgroup analyses with special reference to the effect of antiplatelet agents in acute coronary syndromes

Doron Aronson[1,2]

[1]Department of Cardiology, Rambam Medical Center, Haifa, Israel; [2]Rappaport Faculty of Medicine and Research Institute, Technion, Israel Institute of Technology, Haifa, Israel

Summary

Controlled trials estimate treatment effects averaged over the reference population of subjects. However, physicians are interested in whether the treatment effect varies across subgroups (effect heterogeneity) in order to target specific subgroups to maximise the benefit of treatment and minimise harm. Therefore, large clinical trials of antiplatelet agents include subgroup analyses that examine whether treatment effects differ between subgroups of subjects identified by baseline characteristics. Reporting subgroup is pervasive and often accompanied by claims of difference of treatment effects between subgroups with potential important implications for clinical practice. However, subgroup-specific analyses of clinical trial data have inherent limitations that reduce their reliability. These include reduced statistical power, failure to specify the subgroups of interest a priori, failure to account for examining large numbers of subgroups, lack of strong rationale for biological response modification, and performing analyses based on variables measured post randomisation or in trials showing no overall difference between treatments. Rules for interpretation of subgroup findings in subgroups have been suggested but are frequently not applied. In this article we draw attention to the pitfalls of subgroup analyses in the context of recent trials of antiplatelet agents.

Keywords

Clinical trials, effect modification, interaction, prasugrel, statistics, subgroup analysis, ticagrelor

Correspondence to:
Doron Aronson, MD
Department of Cardiology
Rambam Medical Center
Bat Galim
POB 9602, Haifa 31096, Israel
Tel.: +972 48 542790, Fax: +972 48 542176
E-mail: daronson@tx.technion.ac.il

Received: September 30, 2013
Accepted after major revision: January 29, 2014
Epub ahead of print: March 6, 2014

http://dx.doi.org/10.1160/TH13-09-0801
Thromb Haemost 2014; 112: 16–25

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

"Only one thing is worse than doing subgroup analyses --- believing the results."

Richard Peto

"It is right for each physician to want to know about the behaviour to be expected from an intervention or therapy when applied to his particular individual patient … It is not right, however, for a physician to expect to know this …"

John Wilder Tukey

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 participant 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).

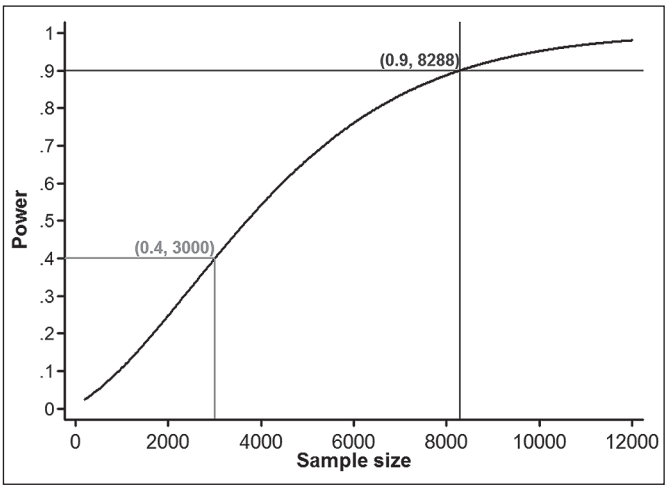

Figure 1: Relationship between sample size and power in the whole study population and in a subgroup.

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 electrogardiography (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

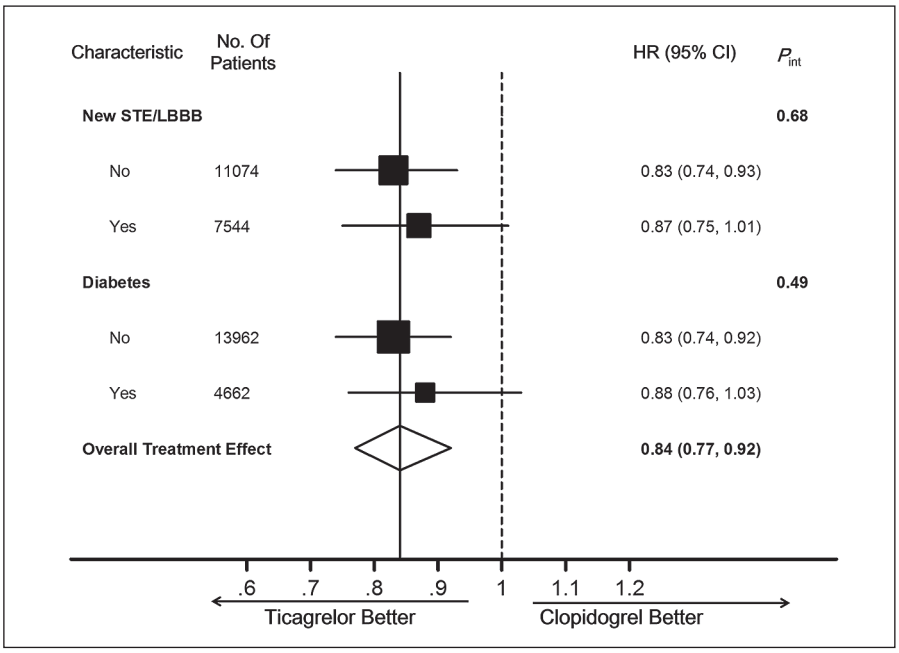

Figure 2: Forest plot of subgroup analysis. Hazard ratios (black squares), 95% CIs (horizontal lines), p-values are for the interaction between the treatment effect and subgroup variables. Bold line shows overall treatment effect and the dotted line shows the no effect point. Data was extracted from the PLATO trial (28).

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

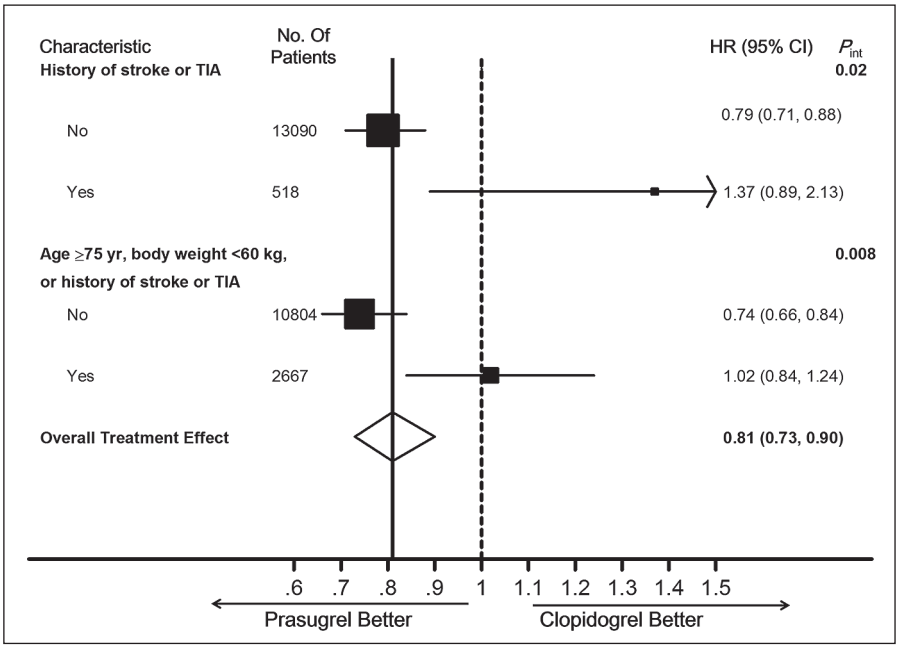

Figure 3: Forest plot of subgroup analysis with heterogeneity of treatment effect. Hazard ratios (black squares), 95% CIs (horizontal lines), p-values are for the interaction between the treatment effect and subgroup variables. Bold line shows overall treatment effect and the dotted line shows the no 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).

Table 1: Subgroup-by-treatment interaction with the primary endpoint in clinical trials of new antiplatelet agents*.

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

* 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.

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

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 subgroup*s) 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 0.98; 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 platelet 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.

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

* 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

Table 3: Criteria for conducting and assessing the credibility of subgroup analyses*.

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.

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 categorises 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

1. Kent DM, Rothwell PM, Ioannidis JP, et al. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. Trials 2010; 11: 85.
2. Altman DG, Matthews JN. Statistics notes. Interaction 1: Heterogeneity of effects. Br Med J 1996; 313: 486.
3. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. Milbank Q 2004; 82: 661-687.
4. Groenwold RH, Donders AR, van der Heijden GJ, et al. Confounding of subgroup analyses in randomized data. Arch Intern Med 2009; 169: 1532-1534.
5. Cui L, Hung HM, Wang SJ, et al. Issues related to subgroup analysis in clinical trials. J Biopharm Stat 2002; 12: 347-358.
6. Hernandez AV, Boersma E, Murray GD, et al. Subgroup analyses in therapeutic cardiovascular clinical trials: are most of them misleading? Am Heart J 2006; 151: 257-264.
7. Oxman AD. Subgroup analyses. Br Med J 2012; 344: e2022.
8. Yusuf S, Wittes J, Probstfield J, et al. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. J Am Med Assoc 1991; 266: 93-98.
9. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. N Engl J Med 1987; 317: 426-432.
10. Assmann SF, Pocock SJ, Enos LE, et al. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 2000; 355: 1064-1069.
11. Bulpitt CJ. Subgroup analysis. Lancet 1988; 2: 31-34.
12. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. Lancet 2005; 365: 176-186.
13. Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. Science 1977; 198: 679-684.

14. Wang R, Lagakos SW, Ware JH, et al. Statistics in medicine--reporting of subgroup analyses in clinical trials. N Engl J Med 2007; 357: 2189-2194.
15. Lagakos SW. The challenge of subgroup analyses--reporting without distorting. N Engl J Med 2006; 354: 1667-1669.
16. Adams KF, Jr. Post hoc subgroup analysis and the truth of a clinical trial. Am Heart J 1998; 136: 753-758.
17. Pocock SJ, Assmann SE, Enos LE, et al. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med 2002; 21: 2917-2930.
18. Sleight P. Debate: Subgroup analyses in clinical trials: fun to look at - but don't believe them! Curr Control Trials Cardiovasc Med 2000; 1: 25-27.
19. DeMets DL, Califf RM. Lessons learned from recent cardiovascular clinical trials: Part I. Circulation 2002; 106: 746-751.
20. Peto R. Current misconception 3: that subgroup-specific trial mortality results often provide a good basis for individualising patient care. Br J Cancer 2011; 104: 1057-1058.
21. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Lancet 1988; 2: 349-360.
22. Wedel H, Demets D, Deedwania P, et al. Challenges of subgroup analyses in multinational clinical trials: experiences from the MERIT-HF trial. Am Heart J 2001; 142: 502-511.
23. Montalescot G, Wiviott SD, Braunwald E, et al. Prasugrel compared with clopidogrel in patients undergoing percutaneous coronary intervention for ST-elevation myocardial infarction (TRITON-TIMI 38): double-blind, randomised controlled trial. Lancet 2009; 373: 723-731.
24. Cuzick J. Forest plots and the interpretation of subgroups. Lancet 2005; 365: 1308.
25. Pocock S, Travison T, Wruck L. Figures in clinical trial reports: current practice & scope for improvement. Trials 2007; 8: 36.
26. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. Br Med J 2001; 322: 1479-1480.
27. Parker AB, Naylor CD. Subgroups, treatment effects, and baseline risks: some lessons from major cardiovascular trials. Am Heart J 2000; 139: 952-961.
28. Wallentin L, Becker RC, Budaj A, et al. Ticagrelor versus clopidogrel in patients with acute coronary syndromes. N Engl J Med 2009; 361: 1045-1057.
29. Poulson RS, Gadbury GL, Allison DB. Treatment Heterogeneity and Individual Qualitative Interaction. Am Stat 2012; 66: 16-24.
30. Greenfield S, Kravitz R, Duan N, et al. Heterogeneity of treatment effects: implications for guidelines, payment, and quality assessment. Am J Med 2007; 120: S3-9.
31. Sormani MP, Bruzzi P. Reporting of subgroup analyses from clinical trials. Lancet Neurol 2012; 11: 747; author reply -748.
32. Altman D, Schulz K, Moher D. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Inter Med 2001; 134: 663-694.
33. Fletcher J. Subgroup analyses: how to avoid being misled. Br Med J 2007; 335: 96-97.
34. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. Br Med J 2003; 326: 219.
35. Califf RM, DeMets DL. Principles from clinical trials relevant to clinical practice: Part I. Circulation 2002; 106: 1015-1021.
36. Montori V, Jaeschke R, Schunemann H. Users' guide to detecting misleading claims in clinical research reports. Br Med J 2004; 329: 1093-1096.
37. Sun X, Briel M, Busse JW, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. Br Med J 2012; 344: e1553.
38. Gail M, Simon R. Testing for qualitative interactions between treatment effects and patient subsets. Biometrics 1985; 41: 361-372.
39. Sun X, Briel M, Walter SD, et al. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. Br Med J 2010; 340: c117.
40. Senn S, Harrell F. On wisdom after the event. J Clin Epidemiol 1997; 50: 749-751.
41. Wiviott SD, Braunwald E, McCabe CH, et al. Prasugrel versus clopidogrel in patients with acute coronary syndromes. N Engl J Med 2007; 357: 2001-2015.
42. Brookes S, Whitley E, Peters T, et al. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. Health Technol Assess 2001; 5: 56.
43. Pocock S, Hughes M, Lee R. Statistical problems in the reporting of clinical trials. A survey of three medical journals. N Engl J Med 1987; 317: 426-432.
44. Brookes ST, Whitely E, Egger M, et al. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004; 57: 229-236.
45. Wiviott SD, Braunwald E, Angiolillo DJ, et al. Greater clinical benefit of more intensive oral antiplatelet therapy with prasugrel in patients with diabetes mellitus in the trial to assess improvement in therapeutic outcomes by optimizing platelet inhibition with prasugrel-Thrombolysis in Myocardial Infarction 38. Circulation 2008; 118: 1626-1636.
46. Wiviott SD, White HD, Ohman EM, et al. Prasugrel versus clopidogrel for patients with unstable angina or non-ST-segment elevation myocardial infarction with or without angiography: a secondary, prespecified analysis of the TRILOGY ACS trial. Lancet 2013; 382: 605-613.
47. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992; 116: 78-84.
48. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Br Med J 2010; 340: c869.
49. Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. Lancet 2005; 365: 1657-1661.
50. Schulz KF, Grimes DA. Multiplicity in randomised trials I: endpoints and treatments. Lancet 2005; 365: 1591-1595.
51. Lee KL, McNeer JF, Starmer CF, et al. Clinical judgment and statistics. Lessons from a simulated randomized trial in coronary artery disease. Circulation 1980; 61: 508-515.
52. Cook DI, Gebski VJ, Keech AC. Subgroup analysis in clinical trials. Med J Aust 2004; 180: 289-291.
53. Barrett-Connor E. Looking for the pony in the HERS data. Heart and Estrogen/progestin Replacement Study. Circulation 2002; 105: 902-903.
54. Bhatt DL, Fox KA, Hacke W, et al. Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. N Engl J Med 2006; 354: 1706-1717.
55. Stallones RA. The use and abuse of subgroup analysis in epidemiological research. Prev Med 1987; 16: 183-194.
56. Smith PK, Goodnough LT, Levy JH, et al. Mortality benefit with prasugrel in the TRITON-TIMI 38 coronary artery bypass grafting cohort: risk-adjusted retrospective data analysis. J Am Coll Cardiol 2012; 60: 388-396.
57. Cannon CP, Harrington RA, James S, et al. Comparison of ticagrelor with clopidogrel in patients with a planned invasive strategy for acute coronary syndromes (PLATO): a randomised double-blind study. Lancet 2010; 375: 283-293.
58. James SK, Roe MT, Cannon CP, et al. Ticagrelor versus clopidogrel in patients with acute coronary syndromes intended for non-invasive management: substudy from prospective randomised PLATelet inhibition and patient Outcomes (PLATO) trial. Br Med J 2011; 342: d3527.
59. Ferreiro JL, Angiolillo DJ. Diabetes and antiplatelet therapy in acute coronary syndrome. Circulation 2011; 123: 798-813.
60. Pfeffer MA, Jarcho JA. The charisma of subgroups and the subgroups of CHARISMA. N Engl J Med 2006; 354: 1744-1746.
61. Freemantle N. Interpreting the results of secondary end points and subgroup analyses in clinical trials: should we lock the crazy aunt in the attic? Br Med J 2001; 322: 989-991.
62. Guyatt G, Wyer P, Ioannidis J. When to Believe a Subgroup Analysis. User's Guide to the Medical Literature: A Manual for Evidence-Based Clinical Practice 2008: pp. 571-583.
63. Roffi M, Chew DP, Mukherjee D, et al. Platelet glycoprotein IIb/IIIa inhibitors reduce mortality in diabetic patients with non-ST-segment-elevation acute coronary syndromes. Circulation 2001; 104: 2767-2771.
64. Mehilli J, Kastrati A, Schuhlen H, et al. Randomized clinical trial of abciximab in diabetic patients undergoing elective percutaneous coronary interventions after treatment with a high loading dose of clopidogrel. Circulation 2004; 110: 3627-3635.
65. Kastrati A, Mehilli J, Neumann FJ, et al. Abciximab in patients with acute coronary syndromes undergoing percutaneous coronary intervention after clopidogrel pretreatment: the ISAR-REACT 2 randomized trial. J Am Med Assoc 2006; 295: 1531-1538.
66. Mahaffey KW, Wojdyla DM, Carroll K, et al. Ticagrelor compared with clopidogrel by geographic region in the Platelet Inhibition and Patient Outcomes (PLATO) trial. Circulation 2011; 124: 544-554.

67. Ohman EM, Roe MT. Explaining the unexpected: insights from the PLATelet inhibition and clinical Outcomes (PLATO) trial comparing ticagrelor and clopidogrel. Editorial on Serebruany "Viewpoint: Paradoxical excess mortality in the PLATO trial should be independently verified" (Thromb Haemost 2011; 105.5). Thromb Haemost 2011; 105: 763-765.
68. Ware JH. The National Emphysema Treatment Trial--how strong is the evidence? N Engl J Med 2003; 348: 2055-2056.
69. Altman DG. Within trial variation--a false trail? J Clin Epidemiol 1998; 51: 301-303.
70. Sun X, Briel M, Busse JW, et al. The influence of study characteristics on reporting of subgroup analyses in randomised controlled trials: systematic review. Br Med J 2011; 342: d1569.
71. Counsell CE, Clarke MJ, Slattery J, et al. The miracle of DICE therapy for acute stroke: fact or fictional product of subgroup analysis? Br Med J 1994; 309: 1677-1681.
72. Roe MT, Armstrong PW, Fox KA, et al. Prasugrel versus clopidogrel for acute coronary syndromes without revascularisation. N Engl J Med 2012; 367: 1297-1309.
73. Gurbel PA, Erlinge D, Ohman EM, et al. Platelet Function During Extended Prasugrel and Clopidogrel Therapy for Patients With ACS Treated Without Revascularisation: The TRILOGY ACS Platelet Function Substudy. J Am Med Assoc 2012: 1-10.
74. Aradi D, Komocsi A, Vorobcsuk A, et al. Impact of clopidogrel and potent P2Y 12 -inhibitors on mortality and stroke in patients with acute coronary syndrome or undergoing percutaneous coronary intervention: a systematic review and meta-analysis. Thromb Haemost 2013; 109: 93-101.
75. U.S. Department of Health and Human Services. Food and Drug Administration. Guidance for Industry: Collection of Race and Ethnicity Data in Clinical Trials. 2005; Available from: www.fda.gov/downloads/Guidances/ucm126396.pdf.
76. Pocock S, Calvo G, Marrugat J, et al. International differences in treatment effect: do they really exist and why? Eur Heart J 2013; 34: 1846-1852.
77. Wallentin L, James S, Storey RF, et al. Effect of CYP2C19 and ABCB1 single nucleotide polymorphisms on outcomes of treatment with ticagrelor versus clopidogrel for acute coronary syndromes: a genetic substudy of the PLATO trial. Lancet 2010; 376: 1320-1328.
78. Pocock SJ, Lubsen J. More on subgroup analyses in clinical trials. N Engl J Med 2008; 358: 2076.
79. Kent DM, Lindenauer PK. Aggregating and disaggregating patients in clinical trials and their subgroup analyses. Ann Intern Med 2010; 153: 51-52.
80. Rothwell PM, Mehta Z, Howard SC, et al. Treating individuals 3: from subgroups to individuals: general principles and the example of carotid endarterectomy. Lancet 2005; 365: 256-265.
81. Rothwell PM. Can overall results of clinical trials be applied to all patients? Lancet 1995; 345: 1616-1619.
82. Mehta SR, Bassand JP, Chrolavicius S, et al. Dose comparisons of clopidogrel and aspirin in acute coronary syndromes. N Engl J Med 2010; 363: 930-942.
83. James S, Angiolillo DJ, Cornel JH, et al. Ticagrelor vs. clopidogrel in patients with acute coronary syndromes and diabetes: a substudy from the PLATelet inhibition and patient Outcomes (PLATO) trial. Eur Heart J 2010; 31: 3006-3016.
84. Bhatt DL, Stone GW, Mahaffey KW, et al. Effect of platelet inhibition with cangrelor during PCI on ischaemic events. N Engl J Med 2013; 368: 1303-1313.
85. Morrow DA, Braunwald E, Bonaca MP, et al. Vorapaxar in the secondary prevention of atherothrombotic events. N Engl J Med 2012; 366: 1404-1413.
86. Tricoci P, Huang Z, Held C, et al. Thrombin-receptor antagonist vorapaxar in acute coronary syndromes. N Engl J Med 2012; 366: 20-33.
87. Montalescot G, Bolognese L, Dudek D, et al. Pretreatment with prasugrel in non-ST-segment elevation acute coronary syndromes. N Engl J Med 2013; 369: 999-1010.
88. Mehta SR, Yusuf S, Peters RJ, et al. Effects of pretreatment with clopidogrel and aspirin followed by long-term therapy in patients undergoing percutaneous coronary intervention: the PCI-CURE study. Lancet 2001; 358: 527-533.
89. Yusuf S, Zhao F, Mehta SR, et al. Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. N Engl J Med 2001; 345: 494-502.
90. Hernandez A, Boersma E, Murray G. Subgroup analyses in therapeutic cardiovascular clinical trials: are most of them misleading? Am Heart J 2006; 151: 257-264.
91. Sun X, Briel M, Busse JW, et al. Subgroup Analysis of Trials Is Rarely Easy (SATIRE): a study protocol for a systematic review to characterize the analysis, reporting, and claim of subgroup effects in randomized trials. Trials 2009; 10: 101.
92. Browner WS, Hulley SB. Effect of risk status on treatment criteria. Implications of hypertension trials. Hypertension 1989; 13: I51-56.

