Pre-specified vs. post-hoc subgroup analyses: are we wiser before or after a trial has been performed?

Helmut Schülen*

Vivantes Auguste-Viktoria-Klinikum, Klinik für Innere Medizin – Kardiologie, Diabetologie und konservative Intensivmedizin, Rubensstr. 125, D-12157 Berlin, Germany

Online publish-ahead-of-print 20 June 2014

This editorial refers to ‘Ticagrelor vs. clopidogrel in patients with non-ST-elevation acute coronary syndrome with or without revascularization: results from the PLATO trial†, by D. Lindholm et al., on page 2083.

Pivotal trials on antithrombotic regimens in coronary artery disease have been performed with huge patient cohorts; in CURE, 12,562 patients were randomized,1 TRITON-TIMI 38 included 13,608 patients,2 and PLATO 18,624 patients.3 Such databases provide an extraordinary opportunity to study a treatment effect in several subgroups of patients which—although subsets—still consist of thousands of patients. Such a subanalysis of the non-ST-segment elevation myocardial infarction (NSTEMI) population in PLATO (a subgroup by itself) has now been published, differentiating patients with or without revascularization.4

What is the interest of the scientific community in performing subgroup analyses?

First of all, subgroup analyses may demonstrate consistent results over various complementary subpopulations, e.g. male and female, young and elderly patients. This would indicate stability of a treatment effect over a broad study population. Subgroup analysis could also identify patient subsets with a particular treatment effect, either positive or negative. This might be of interest if high rates of side effects call for selective use of a new therapy. Finally, in trials with an overall negative result, subgroup analyses might identify patient subsets with a significant treatment effect. While the first two approaches are performed frequently, the third is considered inappropriate by the scientific community.

What do statisticians think about debates on subgroup data?

One major concern is that few trials have sufficient statistical power to estimate a treatment effect reliably in multiple subgroups. So, statisticians would caution us that these analyses are hypothesis generating at best, and cannot be regarded as evidence. Is that just the perfect line never to be missed in the limitations section of a manuscript, or should we be more cautious beyond this? Wasn’t Peter Sleight a bit too harsh warning us that subgroup analyses are ‘fun to look at, but don’t believe them!’?5 Let us look at the elderly. Data on the elderly are of common interest. Treatment effects could be quite different and side effects more frequent in older patients. However, the question is what is elderly? Is it all patients over 65 years of age? Or is it 75, 80, or perhaps 85 years? A trial database is easily analysed for all these different cut-off values. Such data sets may be very consistent. However, they could also be quite divergent, with significant differences at one particular cut-off, perhaps non-significant trends at others, or results contradicting the main results. Sometimes these results would fit current knowledge or a hypothesis, but sometimes it may be quite difficult to understand differences that emerge. They could indicate a differential treatment effect based on age, but could also be play of chance. Which data will end up in a publication? Most probably those that fit the hypothesis, or provide the most evidence. That leads us back to Peter Sleight. He once commented on subgroup analyses of ISIS-2, which had indicated a beneficial effect of aspirin after myocardial infarction in patients born under all astrological signs except for Gemini and Libra.6 ‘Of course most physicians (but not all!) laughed when they were presented with these results. However, when presented with other less ridiculous subgroup analyses they are likely to believe the results, ... particularly if the result can be justified by some pet theory.’5

These data from ISIS-2 have been the favourite example to warn us of dividing populations into multiple subgroups, yielding data which may be the result of statistical play of chance. There are further issues, relating to adjustments for covariates, for multiple comparisons, the decrease of statistical power, etc. So, how can we identify data which are the result of play of chance, but do not make us burst out laughing? How can we be certain that the data presented were not taken out of the context of an array of multiple analyses performed, chosen in the end because they best fit a hypothesis?

The opinions expressed in this article are not necessarily those of the Editors of the European Heart Journal or of the European Society of Cardiology.


* Corresponding author. Tel: +49 30 130 20 2104, Fax: +49 30 130 20 2142, Email: helmut.schuelen@vivantes.de

Published on behalf of the European Society of Cardiology. All rights reserved. © The Author 2014. For permissions please email: journals.permissions@oup.com.
Table 1  Ten criteria to assess credibility of subgroup analyses

<table>
<thead>
<tr>
<th>Design</th>
<th>Analysis</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Was the subgroup variable a baseline characteristic?</td>
<td>• Was the test of interaction significant (interaction $P &lt; 0.05$)?</td>
<td>• Was the direction of the subgroup effect correctly pre-specified?</td>
</tr>
<tr>
<td>• Was the subgroup variable a stratification factor at randomization?</td>
<td>• Was the significant interaction effect independent, if there were multiple significant interactions?</td>
<td>• Was the subgroup effect consistent with evidence from previous related studies?</td>
</tr>
<tr>
<td>• Was the subgroup hypothesis specified a priori?</td>
<td></td>
<td>• Was the subgroup effect consistent across related outcomes?</td>
</tr>
<tr>
<td>• Was the subgroup analysis one of a small number of subgroup hypotheses tested ($≤5$)?</td>
<td></td>
<td>• Was there any indirect evidence to support the apparent subgroup effect—for example, biological rationale, laboratory tests, animal studies?</td>
</tr>
</tbody>
</table>

Adapted from James et al. 10

How can we identify more reliable subgroup analyses?

The CONSORT (Consolidated Standards of Reporting Trials) initiative7 has stated a strong criticism of post-hoc analyses done after looking at the data, and question their credibility at large. Subgroup analyses which have been pre-specified before data are available would eliminate data selection, but not play of chance. To increase the reliability of such data, principles for subgroup analyses have been defined starting in the 1990s,8 followed by several definitions of criteria to assess credibility (Table 1).9

Like most recent large randomized trials, the design and methods of PLATO have been described in a dedicated publication10 which specified an array of subgroup analyses of efficacy and safety variables.

This brings us back to the study of Lindhom et al.11 This analysis focuses on the NSTEMI subgroup of PLATO, comparing patients with or without revascularization. PLATO had allowed randomization of patients with acute coronary syndromes regardless of whether aiming for an invasive or non-invasive strategy, or if a revascularization procedure had been intended or actually performed. As a consequence, angiography during the index admission was performed in only 81.5%.3 Furthermore, only a subset of patients actually did undergo revascularization procedures. During the index admission, 61.0% of the overall population underwent percutaneous coronary intervention (PCI) and 4.5% underwent coronary artery bypass grafting (CABG). At the end of the trial, 64.3% had undergone PCI and 10.2% CABG. This distinguishes the design of PLATO from that of TRITON-TIMI 38, the trial on prasugrel in acute coronary syndromes,2 where only patients who had been scheduled for PCI were randomized. Differences between the two trials in design as well as results have spurred intensive discussions and debates. A subgroup analysis of PLATO patients who underwent PCI has long been awaited, but has not been presented yet.

For the study of Lindholm et al., patients with and without revascularization procedures (PCI or CABG) during the first 10 days were analysed,4 illustrating consistent trends for a positive treatment effect in both sub-subgroups, without an interaction for the higher overall bleeding risk associated with ticagrelor. But was it pre-specified? No, the design paper had announced analyses of patients with any revascularization within the initial 30 days, not differentiating the STEMI and NSTEMI population.8

So, what would be the rationale to report on the NSTEMI population only? Current guidelines do not recommend routine invasive treatment in all NSTEMI patients,11 supporting this focus on NSTEMI. However, the patients in the STEMI subgroup in PLATO were not all treated invasively: 93.5% underwent angiography, and only 72.1% had primary PCI performed within the first 12 h.12 The second difference was to analyse patients with revascularization within the first 10 days, instead of the pre-specified 30 days. The latter would be based on more patients with revascularization, but what should be the rationale beyond?

During the review process of this manuscript, the authors had been asked to provide the data based on the pre-specified definitions. As a result of this debate, these data are now illustrated in the Supplementary material online. This allows for full transparency of the two data sets, based on the pre-specified vs. the post-hoc cut-off values.

How could we all be more stringent in the future?

Pre-specified subgroup analyses are more reliable and valuable than post-hoc analyses. Therefore, pre-specification of subgroups could be made mandatory for publication. This still does not eliminate selective publication of results that fit a hypothesis. So, the scientific community might postulate that all pre-specified subgroups need to be presented consecutively, at some point after the data are available for analyses.

Conflicts of interest: H.S. has received grant support, honoraria, or fees from Astra-Zeneca, Bayer, Boehringer-Ingelheim, Bristol-Myers-Squibb, Correvio, Daiichi-Sankyo, and Eli Lilly & Company.

References

Like a dented bumper: a heart impressed by a giant left atrial appendage in a 22-year-old patient

Akram A. Yousef1, Manuel Wilbring*, Michael Laniado1, and Utz Kappert2

1Department of Cardiology, University Heart Center Dresden, Fetscherstrasse 76, 01307 Dresden, Germany; 2Department of Cardiac Surgery, University Heart Center Dresden, Dresden, Germany; and 3Department of Radiology, University Hospital Carl Gustav Carus, Dresden, Germany

* Corresponding author. Tel: +49 351 4501805, Fax: +49 351 4501802, Email: manuel.wilbring@gmail.com

Congenital aneurysm of the left atrial appendage (LAA) is a rare cardiac anomaly. Owing to the initial absence of symptoms, LAA aneurysms usually manifest in the second or third decade of life. The pathogenesis of LAA aneurysms is not known, but some authors have hypothesized dysplasia of the musculi pectinati to be the cause. Potential hazardous sequelae include arrhythmias or systemic embolization. Early diagnosis and resection are of utmost importance to prevent secondary morbidity.

Here, we report on a 22-year-old subject presenting with new-onset atrial fibrillation. Radiography of the chest demonstrated a suspicious prominent cardiac silhouette. Subsequent echocardiography (Panel A) and MRI (Panel B) revealed a giant aneurysm of the LAA (9 × 7 cm). The patient then underwent cardiac surgery.

Panels C–F demonstrate the intra-operative aspect of the giant aneurysm, significantly compressing adjacent cardiac structures (F). The left atrial appendage was resected (D). Ablation therapy was undertaken concomitantly for treatment of atrial fibrillation. Intra-operative and post-operative courses were uneventful.

Our patient demonstrated the importance of early diagnosis and therapy in this rare case of LAA aneurysm.