Subgroup analyses in randomized clinical trials: Value and limitations

Review #3 on important aspects of randomized clinical trials in cardiovascular pharmacotherapy

Heinz Drexel [1,2,3,4*#], Stuart J. Pocock [5#], Basil S. Lewis [6*], Christoph H Saely [1,3,7], Juan Carlos Kaski [8*], Giuseppe M.C. Rosano [9*], Gerda Tautermann [1,3,7], Kurt Huber [10,11], Joern F. Dopheide [1,12], Arthur Mader [1,3,7], Alexander Niessner [13*], Gianluigi Savarese [14*], Thomas A. Schmidt [15], Anne Grete Semb [16*], Juan Tamargo [17*], Sven Wassmann [18,19*], Martin Clodi [20], Keld Per Kjeldsen [21*], Stefan Agewall [22,23]

[1] Vorarlberg Institute for Vascular Investigation and Treatment (VIVIT), Carinagasse 47, 6800 Feldkirch, Austria;
[2] Department of Medicine, County Hospital Bregenz, Carl-Pedenz-Str. 2, 6900 Bregenz, Austria;
[3] Private University of the Principality of Liechtenstein, Dorfstr. 24. 9495 Triesen, Liechtenstein;
[4] Drexel University College of Medicine, 2900 W Queen Lane, Philadelphia, PA 19129, USA;
[5] Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, UK;
[6] Cardiovascular Clinical Research Institute, Lady Davis Carmel Medical Center, Michal Str. 7, 34362 Haifa, Israel and the Ruth and Bruce Rappaport School of Medicine, Technion-Israel Institute of Technology, Efron Str. 1, 31096 Haifa, Israel;
[7] Department of Medicine I, Academic Teaching Hospital Feldkirch, Carinagasse 47, 6800 Feldkirch, Austria;
[8] Molecular and Clinical Sciences Res. Inst. St George's University of London, Cranmer Terrace, London SW17 0RE, UK;
[9] Department of Medical Sciences, IRCCS San Raffaele Hospital, Via delle Pisana 249, 00163 Rome, Italy;
[10] 3rd Medical Department, Cardiology and Intensive Care Medicine, Wilhelminen Hospital, Montleartstr. 37, 1160 Vienna, Austria;
[11] Cardiology, Sigmund Freud University, Medical School, Freudplatz 3, 1020 Vienna, Austria;
[12] Department of Angiology, Cantonal Hospital of Graubünden, Loestr. 170, 7000 Chur, Switzerland;
[13] Department of Internal Medicine II, Medical University of Vienna, Währinger Gürtel 18-20, 1090, Vienna, Austria;
[14] Cardiology Unit, Department of Medicine, Karolinska University Hospital D1:04; 171 76 Stockholm, Sweden;
[15] Department of Emergency Medicine, North Zealand University Hospital, Dyrehavevey, 3400 Hillerød; Denmark;
[16] Preventive Cardio-Rheuma Clinic, Division of Rheumatology and Research, Diakonhjemmet Hospital, Diakonveien 12, 0370 Oslo, Norway;
[17] Department of Pharmacology and Toxicology, School of Medicine, Universidad Complutense, CIBERCV, Plaza de Ramón s/n, 28040 Madrid, Spain;
[18] Cardiology Pasing, Institutstr. 14, 81241 Munich, Germany;
[19] Medical Faculty, Clinical Medicine, University of the Saarland, Kirrbergerstr. 100, 66421 Homburg/Saar, Germany;
[20] Department of Medicine, St. John of God Hospital Linz, Seilerstaette 2, 4021 Linz, Austria;
[21] Department of Cardiology, Copenhagen University Hospital (Amager-Hvidovre), Italiensvej 1, 2300 Copenhagen, Denmark, and Department of Health Science and Technology, Faculty of Medicine, Aalborg University, Fredrik Bajers Vej 7D2, 9220 Aalborg, Denmark;
[22] Department of Cardiology, Ullevål, Oslo University Hospital, Kirkeveien 166, 0450 Oslo, Norway;
[23] Institute of Clinical Sciences, Søsterhjemmet, University of Oslo, Kirkeveien 166, 0450 Oslo, Norway;

* Nucleus members of the ESC Working Group on Cardiovascular Pharmacotherapy
# equal contribution

Corresponding author:

Heinz Drexel, MD, Professor of Medicine, FESC, FAHA, FRCP (Ed.)
VIVIT Institute, Academic Teaching Hospital Feldkirch
Carinagasse 47
6800 Feldkirch, Austria
Tel: 0043 5522 303 6900
Email: heinz.drexel@vivit.at; vivit@lkhf.at

Table of Contents

Introduction

Two review articles previously published from our working group were dedicated to the selection of endpoints as well as to reasons for premature stopping of randomized clinical trials (RCTs).[1,2] We there first discussed the importance of mortality and morbidity endpoints versus softer endpoints like revascularization rates, and the issue of endpoint adjudication. Second, we have shed light on the statistical methods and requirements to stop RCTs prematurely due to safety, futility, or overwhelming efficacy (versus the control arm).

The main objective of this article is now to provide the clinical cardiologist with information how to judge and interpret published subgroup analyses. This section will summarize the situation regarding subgroup analysis and put the current article in context.

Aim of this review article

Modern randomized clinical trials (RCTs) typically look at outcomes of large populations. Based on the selection of the primary endpoint(s), of size and duration of trials, their paramount prerequisite to yield reasonable data is the sample size calculation. Here, also the choices of pre-specified secondary endpoint(s) as well as of pre-specified subgroup analyses are crucial.

A look at subgroups of the study cohort is tempting both for scientists and for clinicians. However, subgroup analyses have some inherent problems. Experts, particularly cardiologists, should be aware of their value but also of their limitations. Scientists/trialists can derive new ideas and hypotheses, and clinicians also are interested in clinical characteristics of subgroups with above-average outcomes.

This review article looks at advantages and disadvantages of subgroup analysis in trials of cardiovascular pharmacotherapy with a focus on antithrombotic and metabolic interventions.

Statistical methods

Basic considerations for clinicians: Some principles

Over and above all statistical considerations it should be kept in mind that the size of a trial is based on a sample size calculation. Logically, this cannot be equally valid for any subgroup analysis (with a smaller sample). As discussed above, it would be important to clarify that subgroup analyses are only valid for the specific endpoint.

The appropriate use of statistical methods is mandatory for subgroup analysis. An important aspect is how to interpret data. A simple use of a $p < 0.05$ difference in a subgroup and not in the rest does not tell one if there is a true subgroup difference. Rather, interaction (heterogeneity) tests of treatment effects between subgroups should be applied. One paramount statistical point therefore is the need for tests of interaction (heterogeneity) of treatment effect between subgroups. P-values between subgroups are misleading and should not be used.

The EMA defines the terms "subgroup" and sub-population".[3] The term ‘subgroup’ there is used to refer to a subset of the clinical trial population defined by one or more intrinsic and extrinsic factors of the patients under investigation, usually measured at baseline, and the term ‘sub-population’ will be used to refer to a subset of the patient population described in the targeted therapeutic indication.

The number of pre-specified subgroup analyses should be limited in order to avoid the possibility that few of them may be positive by chance. On the other hand, when adjusting for multiple comparisons it is important that all pre-defined subgroup analyses are presented. Otherwise many subgroup analyses without a relevant finding may be omitted.

Only pre-specification of subgroup analyses avoids that non-significant results are not published. On the other hand, when adjusting for multiple comparisons it is important that all subgroups should be specified a priori to avoid spurious conclusions, particularly

because the role of bias and variability is often under-estimated when subgroup effects are interpreted a posteriori.[4]

Statistical significance is not always identical to medical importance. In this context, a very conservative way to look at subgroup data often adopted by clinicians is to look at the 95% of confidence intervals of the two (or more) subgroups as they are useful in indicating the range of uncertainty around the estimated treatment effect. The less they overlap, the more the subgroup finding is considered clinically meaningful. This requires to examine the treatment effect on both a relative scale (e.g., by calculation of the relative risk or the hazard ratio) and an absolute scale (e.g., by calculation of the differences in the rates of events during follow-up and in the number needed to treat).[5]

Deep dive into statistical issues

Subgroup analyses are standard in current trial reporting. From a statistical point of view, however, they are problematic; if not interpreted properly, they may infer grossly false conclusions.

Analyzing subgroups takes into consideration the potential impact of a covariate, i.e. a baseline variable that is expected to influence the primary variable to be analyzed, which in the case of clinical trials is the impact of an intervention on the study outcome.
Here, multiplicity is a problem to recognize, i.e. how many subgroup analyses were done.
A good general rule how to interpret subgroup data is the following: Interaction tests have the advantage that they directly assess if a treatment effect varies by subgroups, whereas subgroup P values can be misleading since they only tackle the issue indirectly.
Three types of subgroup analyses are typically used: i) exploratory analyses from trials that failed to established efficacy in the intended population overall, ii) supportive analyses that aim at showing consistency of the intervention effect across subgroups when the intervention has been efficacious in the studied population overall, and iii)

inferential analyses that aim at establishing the efficacy of the intervention in a pre-defined targeted subgroup.

Exploratory subgroup analyses are hypothesis generating at best, regardless of the level of statistical significance they are reported with. In supportive subgroup analyses, homogeneity across strata is assumed, when the interaction subgroup by intervention (which is reflected by an interaction term subgroup x intervention in statistical analysis) is not significant. However, also these analyses must be interpreted cautiously: Clinical trials due to the inclusion of few patients usually lack statistical power to show differences between subgroups, which makes type 2 errors likely. Indeed, the size of a trial typically is based on a sample size calculation addressing the primary outcome in the overall study population. On the other hand, differences between groups also with supportive subgroup analyses can show up merely by chance, a problem aggravated by the multiplicity of comparisons frequently performed. Clinical plausibility and confirmation of such subgroup findings in subsequent or related trials therefore are important to support their credibility. As already mentioned, statistical significance is not identical to medical importance. In this context, a very critical and conservative way to look at subgroup data should be adopted by clinicians.

Inferential subgroup analyses have a different objective. They directly aim at establishing efficacy in the population pre-defined by the subgroup. To allow for inferential subgroup analysis, the trial a priori must be designed to establish efficacy in the addressed subgroup - with regard to statistical power for this subgroup (which includes the consideration of subgroup size and intervention effect in the subgroup), the method chosen for the analysis and adjustments for multiple comparisons. In this case, statistically significant interactions are meaningful only when the magnitude of interaction is similar to the magnitude of the overall treatment effect. Confidence in the overall “positivity” of a trial increases when prespecified secondary outcomes also show a treatment benefit, while when secondary outcomes show no benefit, the credibility of the results will decrease.

Finally, when multiple subgroup analyses are performed, the probability of a false positive finding can be substantial.[6]

If one detects an interaction, three types of true interaction can be distinguished (Table 1).

Types of subgroups

The selection of subgroups depends on the type of the trial, essentially on the disease category of the study population as well as on the characteristics and mechanism of the intervention. Therefore, subgroup analyses vary from trial to trial, but some are consistent: age and sex, baseline characteristics of the population (e.g. diabetes mellitus Y/N, LDL cholesterol at screening, smoking habits, and pre-treatment). Most but not all of these separations are dichotomous. Examples of less consistent subgroup separations are race and geographic region; blood pressure, glomerular filtration rate, and albuminuria categories, see Table 2.

Inclusion criteria of trials often allow for different disease entities to be included. In atherosclerosis (e.g. lipid-lowering) trials, but also some others, the different vascular bed involvement (coronary, peripheral, carotid) may be a criterion for recruitment. A subgroup analysis according to the affected vascular site is certainly interesting.

As to the endpoints, a distinction between continuous and categorized variables is necessary.

Validity of subgroup data

Only predefined subgroup stratifications are scientifically sound. A statement in the trial outline must have been published. Post hoc subgroup analysis is like cherry picking and results cannot be accepted as scientifically sound.

Credibility of subgroup data

In general, subgroup data are considerably less credible than the results of the total study cohort. If a subgroup performs better than the grand total, the logical consequence is that the counterpart performs worse than average. A different conclusion can arise if one subgroup is clearly positive and the other is just neutral as exemplified by the

IMPROVE-IT trial where diabetic patients but not non-diabetic individuals appeared to clearly benefit from ezetimibe plus simvastatin versus simvastatin alone.[7] Here, the concept prevails that the result is valid for the total cohort.

Another example related with diabetes is that the diabetic state appears to affect the efficacy of ticagrelor and prasugrel in patients with ACS. In patients with DM, the efficacy of ticagrelor was comparable with that of prasugrel.[8,9]

On the other hand, subgroup analysis is important to corroborate the central finding of a trial. For example, in the 4S trial, subgroup analyses supported the beneficial effect of simvastatin versus placebo in each subgroup, no subpopulation emerged where the statin was not superior to placebo. Thus, subgroup data are helpful to confirm the inner consistency of the overall results of a trial.

Geographic considerations

Treatment benefit may vary according to patient characteristics. Most phase III RCTs recruited very heterogeneous populations coming from different countries, regions, or even continents, with major differences in race, comorbidities, etiology, pathophysiology, clinical presentation, practice patterns and healthcare systems. Geographic differences in multinational trials may affect trial outcome. Trial planning should prespecify expected distribution of patient recruitment so that the database and analyses at any stage reflect a proper proportion of input from different geographic regions. Importantly, every effort should be made to ensure uniformity of interpretation and compliance with trial protocol when recruiting and treating patients in a trial. In the TOPCAT trial, whereas the country-specific and regional heterogeneity could be viewed as statistical variation in a large multinational trial, the differences in patient characteristics, lower event rates, drug adherence, lack of certain drug class-related pharmacodynamic effects, and complete lack of treatment effect in Russia and Georgia compared with the other regions strongly suggest that more than the play of chance occurred.[10] Thus, regional differences in outcome events constitute another type of subset analysis. Again, in TOPCAT, the issue remains of whether this kind of subset analysis may be considered valid in view of the

neutral results of the primary study outcome. A further example of geographical differences is the LoDoCo2 Study, investigating Colchicine in patients with chronic coronary disease were subgroup analyses showed a difference in the primary endpoint between the two investigating countries, Australia and Netherlands.[11]

Examples from lipid and diabetes trials

Before looking at subgroup results from RCTs on lipid lowering, it is worthwhile to consider some general aspects. With an intervention by cardiovascular drugs it is important to distinguish between absolute and relative risk reduction.[12] From the Cholesterol Treatment Trialists Collaboration (CTTC)[13] we have learned a crucial fact that the relative risk reduction by statins is a function of the absolute reduction of blood LDL cholesterol (LDL-C), i.e. of the difference between the pretreatment and the treated level.

If we accept that the relative reduction (e.g. of LDL-C) is a characteristic of the drug and the absolute risk reduction is additionally determined by the absolute risk of the study population, we can conclude that the success of a lipid intervention is the higher, both the higher is the efficacy of the drug regimen as well as the higher is the absolute risk of the study population. With the assumption that the given drug dose reduces LDL-C by a constant percentage, it is easy to conclude that the higher the baseline level, the larger the absolute reduction of LDL-C.

Given that information from the CTTC it can be expected that, the higher the baseline LDL-C, the larger will be the absolute difference in LDL-C which determines the outcome defined as relative reduction of clinical endpoints. This expectation was exactly confirmed by the ODDYSSEY Outcomes trial. The conclusion that the finding of a significant outcome benefits only if LDL-C is above 100 mg/dl at baseline thus simply reflects the basic epidemiologic rule from CTTC, and cannot be regarded as a subgroup result of the trial.[14]

In RCTs with a focus on atherosclerotic cardiovascular diseases, one of the larger subgroups studied are patients with peripheral artery disease (PAD). In part two of our trilogy on RCTs, we have listed several difficulties investigators might face when initiating studies solely concentrating on PAD.[1] The most common problem in trials on PAD patients are lower than anticipated recruitment rates,[15] which might be the main reason why PAD is rarely studied in a specific RCT but is rather included in the large group of cardiovascular patients. Thus, the influence of the cardiovascular pharmacotherapy tested in the RCT on PAD is mostly reflected in a subgroup analysis. A typical example in this regard is the FOURIER trial,[16] which had a large PAD subgroup (n= 3642 (13.2%)[17] and resulted in a significant influence on the present recommendations for the lipid lowering strategies in PAD patients,[18] as if the trial were designed specifically for PAD patients. The investigators found evolocumab to significantly reduce the primary end point in PAD patients and because of their higher cardiovascular risk, PAD patients had an even larger absolute risk reductions (ARR) for the primary end point (ARR 3.5%) than those without PAD (ARR 1.6%). Most importantly, however, evolocumab reduced the risk of major adverse limb events in all patients, but the number needed to treat (NNT) was impressively lower for PAD patients (NNT= 25 with PAD vs. NNT=67 without PAD).
Although a subgroup analysis is often not recommended due to low statistical power as stated above, FOURIER impressively showed that with a large enough subgroup, reliable and robust relationships between intervention and result can be achieved.

Examples of studies carried out with diabetes drugs in a cardiovascular setting:

Following the FDA regulation in 2008, it was decided that all diabetes drugs must undergo a cardiovascular safety study. The first large study in this study cycle was the TECOS study. In this study, which was published in 2015, sitagliptin was examined in 14,000 patients with diabetes, most of whom had pre-existing cardiovascular diseases. The impact of diabetes treatment on cardiovascular risk was examined using as primary endpoint MACE. In this study, it was possible to show that sitagliptin reaches the non-inferiority limit, which means that it is safe from a cardiovascular standpoint. In the

outline of the subgroups, this effect was evident across the entire spectrum of the predefined subgroup analysis.[19]

This resulted in certainty that this DPP4 inhibitor and later all others are cardiovascular safe in all subgroups.

The more recent cardiovascular outcome studies, particularly the SGLT2 inhibitor studies, show interesting positive cardiovascular results. In these studies, it is extremely interesting to look at the subgroup analyses.

The effects of SGLT2 inhibitors seem to be multidimensional or multifactorial. The subgroup analyses made it very clear here that the effect of SGLT2 inhibitors is not only due to their primary effect of glucose elimination via the kidneys. It was shown that a clearly positive effect could also be seen in the subgroup with a reduced GFR. The subgroup analyses were of crucial importance here. This has subsequently also led to studies being carried out in patients with a low GFR. Subgroup analyses and also inclusion criteria in the studies can reveal significant differences.

The subgroup analysis in the SGLT-2 inhibitor empagliflozin outcome study (EmpaReg Outcome) showed a significant p-value for interaction of 0.01 between the patients who were younger than 65 years versus those who were older, whereby the older patients were more likely to benefit. The same result was also shown in the DECLARE-TIMI-58 study, although the p-value for interaction was not significant here. Taken together, one could conclude that SGLT2 inhibitors are mainly of benefit in the group of people over 65 years of age. Although this was pre-specified in the EmpaReg study, due to the nature of the subgroup analyses, it should nevertheless be seen as hypothesis generating. Ultimately, this result could also have come about due to the age-related significantly higher absolute risk for cardiovascular events in older people.[20,21] In essence, the relative risk reduction by the specific drug together with the absolute risk level of the subpopulation determine the absolute risk reduction.

Very important aspects from the subgroup analysis come to light in the most recently published studies with dapagliflozin and empagliflozin in patients with a reduced ejection fraction. Specifically, when assessed from many other points of view, there was no difference here between the subgroups of the population that had diabetes upon inclusion and those who had no diabetes upon inclusion. The result here was almost identical. From the subgroup analysis, which was again pre-specified, one could clearly see that the drugs have a positive effect regardless of diabetes status (Figure 1).[22,23]

In the most recently published data on SGLT2 inhibitors in renal insufficiency, the subgroup analysis also showed that the effect of the SGLT2 inhibitors was independent of diabetes status. For example in the DAPA-CKD, a significantly positive result was also seen in the group of patients with IgA nephropathy (interestingly, the largest study on IgA nephropathy so far) in this subgroup.[24]

Numerous interesting subgroup analyses are also available for the drug class of GLP-1 analogues. Data from the REWIND study (dulaglutide) as well as data from the LEADER study (liraglutide) consecutively show a profound reduction of 3-point mace, independent of eGFR and albuminuria subgroups.[25,26]

In summary, it can be said that, on the one hand, the subgroup analyses in the diabetes studies provide clear data that positive (or noninferior) results are found across the broad spectrum of subgroups. On the other hand, the subgroup analyses have also shown – especially in the heart failure and kidney studies – that any benefit may be advantageous for both the one and the other subgroup. This has a significant impact on the further possible uses of the substances. Finally, as already mentioned, some subgroup analyses at best generate hypotheses.

Examples from antithrombotic and anticoagulation trials

Antithrombotic drugs form an integral part of the management of many cardiovascular diseases, including patients with atherothrombosis, patients with an acute cardiovascular or cerebrovascular syndrome, patients with evidence of cardiac or

venous thrombosis and patients with atrial fibrillation requiring anticoagulation for stroke prevention. Antithrombotic drug trials are based on an evaluation of efficacy in preventing thrombosis while safety issues revolve around bleeding risk using an accepted bleeding scales.[27] When a given treatment proved greater efficacy, it is important to analize its safety profile to confirm that safety concerns did not offset the benefits. Following Thus, following analysis of the primary outcome, it is indeed appropriate to examine the benefit-risk ratio (the “net benefit”) in predetermined subsets, since there are very clearly substantial expected differences in patients in different age groups, in patients with differing co-morbid conditions such as renal or hepatic dysfunction and in patients receiving concomitant medication that may interact with the effect of an antithrombotic drug, often to cause an increased bleeding tendency.

The ARISTOTLE trial comparing apixaban to warfarin for stroke prevention in patients with atrial fibrillation showed an overall efficacy benefit of apixaban in stroke prevention of 21% with a 31% decrease in major bleeding.[28] Although specifically with apixaban, approximately only a quarter of drug elimination depends on renal function, it was of major importance to examine bleeding risk in patients with different levels of renal function.[29] Subset analysis showed that the safety benefit of apixaban over warfarin regarding major bleeding was significantly greater in patients with baseline eGFR ≤50 ml/min (p value for interaction 0.03).[30] This information from subset analysis translates importantly into clinical practice when considering choice of anticoagulant in patients with renal disease.

A secondary analysis of the ARISTOTLE trial showed that, although bleeding rates were higher among patients with CKD, compared with warfarin, apixaban treatment reduced the rate of stroke, death, and major bleeding, regardless of renal function. Additionallly, apixaban was associated with less major bleeding events across all ranges of eGFRs, paryticularly in patients with with a CrCl ≤50 mL/min, regardless of methods to stimate the GFR (Cockcroft-Gault or CKD-EPI equations or serum cystatin C).[30]

As discussed based on subgroup analyses from ARISTOTLE the interpretation of results of subgroup analyses are essential when overall results show a significant effect of the (pharmacologic) intervention. In this case it is important to emphasize that patients with

an eGFR < 50 ml/min particularly benefit with regard to the safety endpoint but this does not imply that apixaban is not safe in the other subgroups based on the overall results.

In ELDERCARE, another trial of direct oral anticoagulants for stroke prevention including elderly Japanese patients with atrial fibrillation, a once-daily low 15-mg dose of edoxaban was superior to placebo in preventing stroke or systemic embolism in a very elderly (≥80 year old) population.[31] Although there was not a significantly higher overall incidence of major bleeding than placebo, it is noteworthy that there were substantially more gastrointestinal bleeding events in the edoxaban group whereas there was no difference in total bleeding event. The overall net benefit of edoxaban over placebo was greater in patients not receiving nonsteroidal anti-inflammatory (NSAIDs) or antiplatelet drugs.

Examples from recent studies

From the rapidly developing field of cardiovascular RCTs, we want to discuss 3 recently published studies that underline the importance of subgroup analyses (Figure 2).

Summary of key findings in the 3 recent studies
RECOVERY (Figure 2, Panel A): In patients hospitalized with Covid-19, the use of dexamethasone resulted in lower 28-day mortality among those who were receiving either invasive mechanical ventilation or oxygen alone at randomization but not among those receiving no respiratory support.[32]

PARAGON (Figure 2, Panel B): The primary composite outcome of total hospitalizations for heart failure and death from cardiovascular causes did not differ significantly between the two groups. In a multivariable model that accounted for all potential interactions and that used continuous measures when appropriate, there was suggestion of heterogeneity of treatment effect with possible benefit in patients with lower ejection fraction and in women.[33] This led to FDA approval on this basis. However, it should be kept in mind that – strictly judged - this finding (from a generally neutral study) has to be considered hypothesis generating.

THEMIS (Figure 2, Panel C): In patients with stable coronary artery disease and type 2 diabetes without a history of myocardial infarction or stroke, those who received ticagrelor plus aspirin had a lower incidence of ischemic cardiovascular events, but a higher incidence of major bleeding, including intracranial hemorrhage, than those who received placebo plus aspirin.[34]

THEMIS-PCI: THEMIS-PCI investigated a specific group of patients with stable coronary artery disease, type 2 diabetes and previous PCI. Ticagrelor was added to aspirin and reduced cardiovascular death, myocardial infarction, and stroke, although with increased major bleeding, but with a net clinical benefit when comparing irreversible harms.[35] Prior DAPT exposure probably reduced bleeding risk in the THEMIS-PCI cohort, as well, the selection of patients for the performance of PCI probably signals that the patient is not extremely frail or at high fall risk and further identifies a patient who may be less likely to bleed.[36]

Conclusions

- **individualized benefit risk trade-off is important**
- **usually keep emphasis on overall result**
- **pre-specify a few key subgroups**
- **view subgroup analyses as exploratory**
- **use interaction tests, not subgroup P-values**
- **lack power to explore subgroup issues**
- **subgroup claims usually exaggerated**
- **beware, don't overinterpret**
- **subgroups by overall patient risk are useful**

More rarely, an academic cardiologist/trialist may obtain the task to suggest subgroup analyses for an ongoing RCT. For this occasion, four important considerations before one starts subgroup analyses are depicted in Table 3.

Conflict of Interest

Drexel H, Pocock S, Lewis B, Kaski JK, Rosano G, Tautermann G, Huber K, Dopheide JF, Mader A, Niessner A, Saely CH, Schmidt TA, Tamargo J, Kjeldsen KP, Agewall S, Wassmann S: **Nothing to disclose**

Semb AG: **Grants/contract from Eli Lilly, Speakers honorarium from AbbVie, Novartis, Sanofi, Eli Lilly**

Clodi M: **Consulting fees and lecture honorarium from AstraZeneca, Boehringer Ingelheim, Novo Nordisk**

Savarese G: **Grants/contract from Novartis, Boston Scientific, Pharmacosmos, Merck; Consulting fees from Vifor, Societa prodotti antibiotic, AstraZeneca, Roche, Servier, GENESIS, Cytokinetics, Medtronic**

1. Drexel H, Lewis BS, Rosano GMC, Saely CH, Tautermann G, Huber K, Dopheide JF, Kaski JC, Mader A, Niessner A, Savarese G, Schmidt TA, Semb A, Tamargo J, Wassmann S, Per Kjeldsen K, Agewall S, Pocock SJ. The age of randomized clinical trials: three important aspects of randomized clinical trials in cardiovascular pharmacotherapy with examples from lipid, diabetes, and antithrombotic trials. *Eur Hear J - Cardiovasc Pharmacother* 2020;
2. Drexel H, Rosano GMC, Lewis BS, Huber K, Vonbank A, Dopheide JF, Mader A, Niessner A, Savarese G, Wassmann S, Agewall S. The age of randomized clinical trials: three important aspects of randomized clinical trials in cardiovascular pharmacotherapy with examples from lipid and diabetes trials. *Eur Hear journal Cardiovasc Pharmacother* 2020;**6**:97–103.
3. EMA. Guideline on the investigation of subgroups in confirmatory clinical trials -DRAFT. *Comm Med Prod Hum Use - Euroepan Med Agency [online]* 2014;**44**:1–20.
4. Lesko CR, Henderson NC, Varadhan R. Considerations when assessing heterogeneity of treatment effect in patient-centered outcomes research. *J Clin Epidemiol* 2018;**100**:22–31.
5. Pocock SJ, Stone GW. The Primary Outcome Is Positive — Is That Good Enough? *N Engl J Med* New England Journal of Medicine (NEJM/MMS); 2016;**375**:971–979.
6. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials. *N Engl J Med* New England Journal of Medicine (NEJM/MMS); 2007;**357**:2189–2194.
7. Cannon CP, Blazing MA, Giugliano RP, McCagg A, White JA, Theroux P, Darius H, Lewis BS, Ophuis TO, Jukema JW, Ferrari GM De, Ruzyllo W, Lucca P De, Im K, Bohula EA, Reist C, Wiviott SD, Tershakovec AM, Musliner TA, Braunwald E, Califf RM. Ezetimibe Added to Statin Therapy after Acute Coronary Syndromes. *N Engl J Med* 2015;**372**:2387–2397.
8. Schüpke S, Neumann F-J, Menichelli M, Mayer K, Bernlochner I, Wöhrle J, Richardt G, Liebetrau C, Witzenbichler B, Antoniucci D, Akin I, Bott-Flügel L, Fischer M, Landmesser U, Katus HA, Sibbing D, Seyfarth M, Janisch M, Boncompagni D, Hilz R, Rottbauer W, Okrojek R, Möllmann H, Hochholzer W, Migliorini A, Cassese S, Mollo P, Xhepa E, Kufner S, Strehle A, et al. Ticagrelor or Prasugrel in Patients with Acute Coronary Syndromes. *N Engl J Med* Massachusetts Medical Society; 2019;**381**:1524–1534.
9. Ndrepepa G, Kastrati A, Menichelli M, Neumann FJ, Wöhrle J, Bernlochner I, Richardt G, Witzenbichler B, Sibbing D, Gewalt S, Angiolillo DJ, Hamm CW, Hapfelmeier A, Trenk D, Laugwitz KL, Schunkert H, Schüpke S, Mayer K. Ticagrelor or Prasugrel in Patients With Acute Coronary Syndromes and Diabetes Mellitus. *JACC Cardiovasc Interv* Elsevier Inc.;

2020;**13**:2238–2247.

10. Bristow MR, Silva Enciso J, Gersh BJ, Grady C, Rice MM, Singh S, Sopko G, Boineau R, Rosenberg Y, Greenberg BH. Detection and Management of Geographic Disparities in the TOPCAT Trial: Lessons Learned and Derivative Recommendations. JACC Basic to Transl. Sci. Elsevier Inc; 2016. p. 180–189.
11. SM N, ATL F, A M, JW E, A S, TSJ O, SHK T, XF X, MA I, T L, D L, P H, A J, P N, A W, R H, H S, J S, AFM K, MWJ van H, P S, I T, AG T, A M, C J, WA B, M D, M A, GJ H, CA B, et al. Colchicine in Patients with Chronic Coronary Disease. *N Engl J Med* 2020;**383**:1838–1847.
12. Navarese EP, Andreotti F, Raggi P, Kołodziejczak M, Buffon A, Bliden K, Tantry U, Kubica J, Sardella G, Lauten A, Agewall S, Gurbel PA, Brouwer MA. Baseline low-density lipoprotein cholesterol to predict the extent of cardiovascular benefit from lipid-lowering therapies: a review. *Eur Hear J - Cardiovasc Pharmacother* 2019;**5**:47–54.
13. CTT Collaboration. https://www.cttcollaboration.org/ (12 May 2021)
14. Schwartz GG, Steg PG, Szarek M, Bhatt DL, Bittner VA, Diaz R, Edelberg JM, Goodman SG, Hanotin C, Harrington RA, Jukema JW, Lecorps G, Mahaffey KW, Moryusef A, Pordy R, Quintero K, Roe MT, Sasiela WJ, Tamby J-F, Tricoci P, White HD, Zeiher AM. Alirocumab and Cardiovascular Outcomes after Acute Coronary Syndrome. *N Engl J Med* 2018;**379**:2097–2107.
15. Bernardez-Pereira S, Lopes RD, Carrion MJM, Santucci EV, Soares RM, Oliveira Abreu M de, Laranjeira LN, Ikeoka DT, Zazula AD, Moreira FR, Cavalcanti AB, Mesquita ET, Peterson ED, Califf RM, Berwanger O. Prevalence, characteristics, and predictors of early termination of cardiovascular clinical trials due to low recruitment: Insights from the ClinicalTrials.gov registry. *Am Heart J* 2014;**168**:213-219.e1.
16. Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, Kuder JF, Wang H, Liu T, Wasserman SM, Sever PS, Pedersen TR. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *N Engl J Med* 2017;**376**:1713–1722.
17. Bonaca MP, Nault P, Giugliano RP, Keech AC, Pineda AL, Kanevsky E, Kuder J, Murphy SA, Jukema JW, Lewis BS, Tokgozoglu L, Somaratne R, Sever PS, Pedersen TR, Sabatine MS. Low-Density Lipoprotein Cholesterol Lowering With Evolocumab and Outcomes in Patients With Peripheral Artery Disease. *Circulation* 2018;**137**:338–350.
18. Mach F, Baigent C, Catapano AL, Koskinas KC, Casula M, Badimon L, Chapman MJ, Backer GG De, Delgado V, Ference BA, Graham IM, Halliday A, Landmesser U, Mihaylova B, Pedersen TR, Riccardi G, Richter DJ, Sabatine MS, Taskinen M-R, Tokgozoglu L, Wiklund O, Mueller C, Drexel H, Aboyans V, Corsini A, Doehner W, Farnier

M, Gigante B, Kayikcioglu M, Krstacic G, et al. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Eur Heart J* 2020;**41**:111–188.

19. Green JB, Bethel MA, Armstrong PW, Buse JB, Engel SS, Garg J, Josse R, Kaufman KD, Koglin J, Korn S, Lachin JM, McGuire DK, Pencina MJ, Standl E, Stein PP, Suryawanshi S, Werf F Van de, Peterson ED, Holman RR. Effect of Sitagliptin on Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med* New England Journal of Medicine (NEJM/MMS); 2015;**373**:232–242.
20. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, Mattheus M, Devins T, Johansen OE, Woerle HJ, Broedl UC, Inzucchi SE. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. *N Engl J Med* 2015;**373**:2117–2128.
21. Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A, Silverman MG, Zelniker TA, Kuder JF, Murphy SA, Bhatt DL, Leiter LA, McGuire DK, Wilding JPH, Ruff CT, Gause-Nilsson IAM, Fredriksson M, Johansson PA, Langkilde A-M, Sabatine MS. Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med* 2019;**380**:347–357.
22. McMurray JJV, Solomon SD, Inzucchi SE, Køber L, Kosiborod MN, Martinez FA, Ponikowski P, Sabatine MS, Anand IS, Bělohlávek J, Böhm M, Chiang C-E, Chopra VK, Boer RA de, Desai AS, Diez M, Drozdz J, Dukát A, Ge J, Howlett JG, Katova T, Kitakaze M, Ljungman CEA, Merkely B, Nicolau JC, O'Meara E, Petrie MC, Vinh PN, Schou M, Tereshchenko S, et al. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction. *N Engl J Med* 2019;**381**:1995–2008.
23. Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, Januzzi J, Verma S, Tsutsui H, Brueckmann M, Jamal W, Kimura K, Schnee J, Zeller C, Cotton D, Bocchi E, Böhm M, Choi D-J, Chopra V, Chuquiure E, Giannetti N, Janssens S, Zhang J, Gonzalez Juanatey JR, Kaul S, Brunner-La Rocca H-P, Merkely B, Nicholls SJ, Perrone S, Pina I, et al. Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure. *N Engl J Med* 2020;**383**:1413–1424.
24. Heerspink HJL, Stefánsson B V., Correa-Rotter R, Chertow GM, Greene T, Hou F-F, Mann JFE, McMurray JJV, Lindberg M, Rossing P, Sjöström CD, Toto RD, Langkilde A-M, Wheeler DC. Dapagliflozin in Patients with Chronic Kidney Disease. *N Engl J Med* 2020;**383**:1436–1446.
25. Gerstein HC, Colhoun HM, Dagenais GR, Diaz R, Lakshmanan M, Pais P, Probstfield J, Riesmeyer JS, Riddle MC, Rydén L, Xavier D, Atisso CM, Dyal L, Hall S, Rao-Melacini P, Wong G, Avezum A, Basile J, Chung N, Conget I, Cushman WC, Franek E, Hancu N,

Hanefeld M, Holt S, Jansky P, Keltai M, Lanas F, Leiter LA, Lopez-Jaramillo P, et al. Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomised placebo-controlled trial. *Lancet* 2019;**394**:121–130.

26. Marso SP, Daniels GH, Brown-Frandsen K, Kristensen P, Mann JFE, Nauck MA, Nissen SE, Pocock S, Poulter NR, Ravn LS, Steinberg WM, Stockner M, Zinman B, Bergenstal RM, Buse JB. Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med* 2016;**375**:311–322.
27. Caldeira D, David C, Costa J, Ferreira JJ, Pinto FJ. Non-vitamin K antagonist oral anticoagulants in patients with atrial fibrillation and valvular heart disease: systematic review and meta-analysis. *Eur Hear J - Cardiovasc Pharmacother* 2018;**4**:111–118.
28. Granger CB, Alexander JH, McMurray JJV, Lopes RD, Hylek EM, Hanna M, Al-Khalidi HR, Ansell J, Atar D, Avezum A, Bahit MC, Diaz R, Easton JD, Ezekowitz JA, Flaker G, Garcia D, Geraldes M, Gersh BJ, Golitsyn S, Goto S, Hermosillo AG, Hohnloser SH, Horowitz J, Mohan P, Jansky P, Lewis BS, Lopez-Sendon JL, Pais P, Parkhomenko A, Verheugt FWA, et al. Apixaban versus Warfarin in Patients with Atrial Fibrillation. *N Engl J Med* 2011;**365**:981–992.
29. Chang M, Yu Z, Shenker A, Wang J, Pursley J, Byon W, Boyd RA, LaCreta F, Frost CE. Effect of renal impairment on the pharmacokinetics, pharmacodynamics, and safety of apixaban. *J Clin Pharmacol* Blackwell Publishing Inc.; 2016;**56**:637–645.
30. Hohnloser SH, Hijazi Z, Thomas L, Alexander JH, Amerena J, Hanna M, Keltai M, Lanas F, Lopes RD, Lopez-Sendon J, Granger CB, Wallentin L. Efficacy of apixaban when compared with warfarin in relation to renal function in patients with atrial fibrillation: Insights from the ARISTOTLE trial. *Eur Heart J* Oxford University Press; 2012;**33**:2821–2830.
31. Okumura K, Akao M, Yoshida T, Kawata M, Okazaki O, Akashi S, Eshima K, Tanizawa K, Fukuzawa M, Hayashi T, Akishita M, Lip GYH, Yamashita T. Low-Dose Edoxaban in Very Elderly Patients with Atrial Fibrillation. *N Engl J Med* Massachusetts Medical Society; 2020;**383**:1735–1745.
32. Peter Horby; Wei Shen Lim; Jonathan R. Emberson; Richard Haynes; Martin J. Landray. Dexamethasone in Hospitalized Patients with Covid-19 — Preliminary Report. *N Engl J Med* Massachusetts Medical Society; 2020;NEJMoa2021436.
33. Solomon SD, McMurray JJV, Anand IS, Ge J, Lam CSP, Maggioni AP, Martinez F, Packer M, Pfeffer MA, Pieske B, Redfield MM, Rouleau JL, Veldhuisen DJ van, Zannad F, Zile MR, Desai AS, Claggett B, Jhund PS, Boytsov SA, Comin-Colet J, Cleland J, Düngen H-D, Goncalvesova E, Katova T, Kerr Saraiva JF, Lelonek M, Merkely B, Senni M, Shah

SJ, Zhou J, et al. Angiotensin–Neprilysin Inhibition in Heart Failure with Preserved Ejection Fraction. *N Engl J Med* Massachusetts Medical Society; 2019;**381**:1609–1620.

34. Steg PG, Bhatt DL, Simon T, Fox K, Mehta SR, Harrington RA, Held C, Andersson M, Himmelmann A, Ridderstråle W, Leonsson-Zachrisson M, Liu Y, Opolski G, Zateyshchikov D, Ge J, Nicolau JC, Corbalán R, Cornel JH, Widimský P, Leiter LA. Ticagrelor in Patients with Stable Coronary Disease and Diabetes. *N Engl J Med* Massachusetts Medical Society; 2019;**381**:1309–1320.
35. Bhatt DL, Steg PG, Mehta SR, Leiter LA, Simon T, Fox K, Held C, Andersson M, Himmelmann A, Ridderstråle W, Chen J, Song Y, Diaz R, Goto S, James SK, Ray KK, Parkhomenko AN, Kosiborod MN, McGuire DK, Harrington RA, Santos V, Jain A, Lendel I, Russo M, Haught WH, Bouza M, Gogia H, Banerjee S, Kichura G, Kantaros L, et al. Ticagrelor in patients with diabetes and stable coronary artery disease with a history of previous percutaneous coronary intervention (THEMIS-PCI): a phase 3, placebo-controlled, randomised trial. *Lancet* 2019;**394**:1169–1180.
36. Bhatt DL, Steg PG. THEMIS and THEMIS-PCI. *Eur Heart J* 2019;**40**:3378–3381.

Table 1. Types of true interaction in subgroup analyses.

Types of true interaction	
Qualitative	Treatment effect in reverse directions, implausible, rare
All or nothing	Treatment only works in a subgroup, more plausible, important
Quantitative	Treatment benefits some more than other, very likely, but not crucial

Table 2 Examples of subgroup splits: Summary.

Dichotomous subgroups	**Multiple subgroups**
Age	Age
Gender	-
Smoking Y/N	Smoker, nonsmoker, ex-smoker
Pretreatment	Pretreatment dosage
Diabetes Y/N	HbA1c ranges
Race	Caucasians, Afroamericans, Asians
Geographic region	Europe, Easter countries, North-/South-America
Albuminuria Y/N	No albuminuria, microalbuminuria, macroalbuminuria
Hypertension Y/N	Blood pressure range
CKD Y/N	GFR range

Table 3 Considerations before starting a subgroup analysis.

Consideration before starting subgroup analyses	
Patients are not homegeneous	Response to treatment may well vary, legitimate to explore in subgroup analyses
Trials usually not large enough	Lack power to detect subgroup effects
Many possible subgroups	Guard against data dredging/false positive
Do not rely on subgroup P-values	Use interaction tests instead

Figure

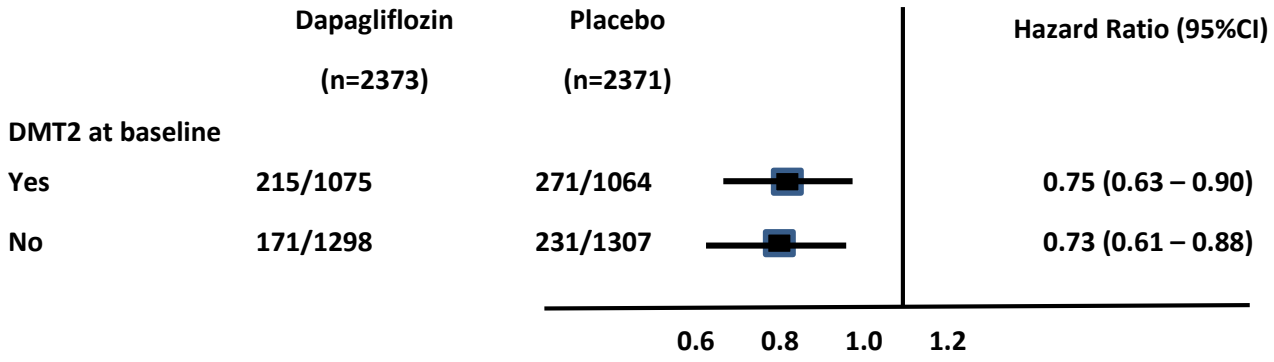


Figure 1 Diabetes mellitus type 2 at baseline in DAPA-HF (Figure adapted from Mc Murray JJV et al. NEJM 2019).[21]

Figure

Panel A

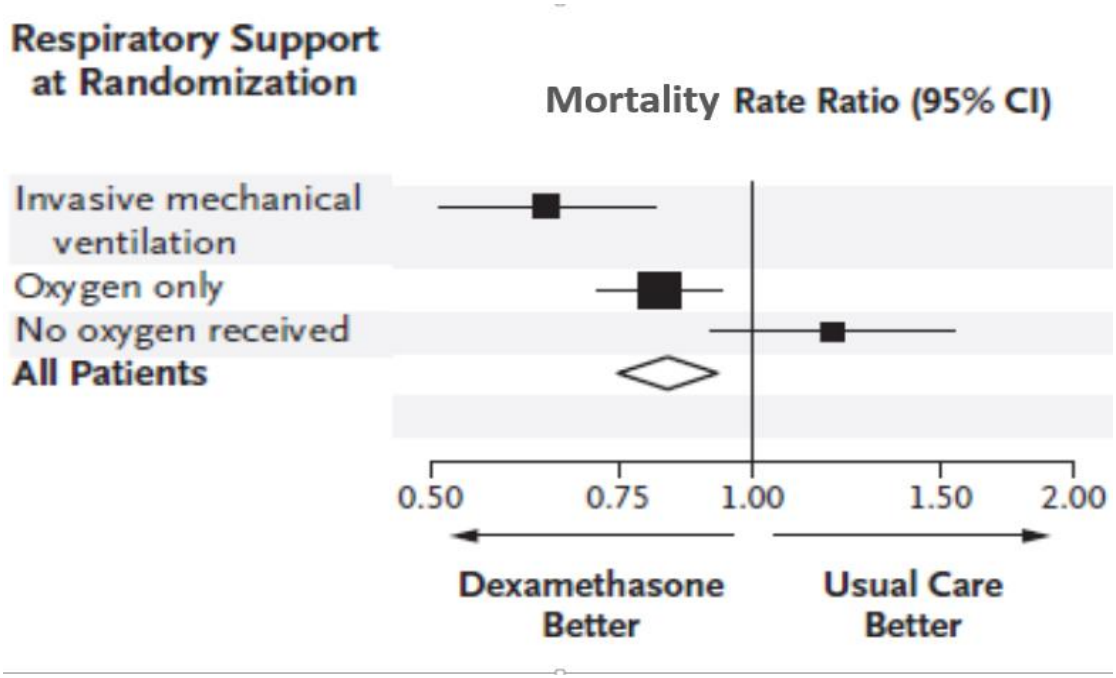
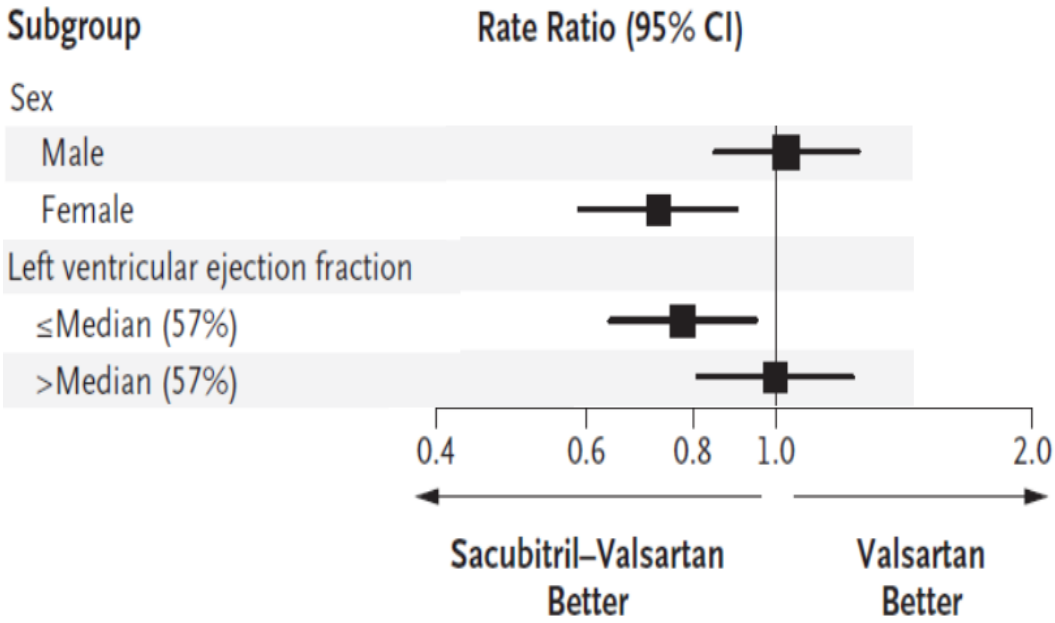


Figure 2 Examples from three recently published cardiovascular RCTs.

Legend to Panel A) RECOVERY trial in hospitalized COVID-19 patients (Figure from (Horby et al. 2020): Key subgroup finding using a forest plot, P for heterogeneity < 0.001; treatment benefit confined to patients getting respiratory support, mortality rate ratio 0.83 (95% CI 0.75 to 0.93), P<0.001, proof beyond reasonable doubt of treatment benefit.[31]

Comments to Panel A): Large simple trials are vital to achieve convincing results, too many other small trials lack clear evidence. Given a highly significant overall benefit, subgroup analysis can then be of value. A highly significant interaction can help refine who really needs a new treatment. Note: RECOVERY had 3 types of multiplicity: multiple treatments, subgroup analysis, and early stopping.

Panel B



Legend to Panel B) PARAGON: sacubitril + valsartan vs valsartan in preserved EF heart failure; 4822 patients, median 35 months follow-up; primary composite outcome: all heart failure hospitalizations and CV death; 13 pre-specified subgroup analyses, two had “significant interactions”; (Figure from McMurray et al. 2019).[32]
Comment to Panel B): Interaction tests: sex P=0.006; LVEF P=0.03 (categorical), P=0.002 (continuous); “sacubritril/valsartan may benefit patients with HF; not frankly reduced, but less than normal”.

Panel C

	Difference in Event Rate (Ticagrelor vs Placebo)		
	Prior PCI	**No Prior PCI**	Interaction test
CV death, MI, stroke	-1.3%	-0.3%	P=0.16
CV death	-0.1%	+0.3%	P=0.41
MI	-0.8%	-0.3%	P=0.42
stroke	-0.6%	-0.2%	P=0.26
TIMI major bleed	+0.9%	+1.4%	P=0.20
Irreversible Harm*	**-1.7%**	**+0.5%**	**P=0.012**

Legend to Panel C) THEMIS: 19 220 patients with stable coronary disease and diabetes; ticagrelor + aspirin vs placebo + aspirin; primary composite outcome: CV death, MI, stroke over mean 39.9 months; hazard ratio 0.90 (95% CI 0.81-0.99) P=0.04; efficacy and safety by prior PCI (58% Yes, 41% No); (Table from Steg 2019).[33]

* All cause death, MI, stroke, fatal bleed, or intracranial haemorrhage. Ticagrelor provided a favourable net clinical benefit after prior PCI. Beware: a subgroup analysis of a post hoc endpoint.

Comment to Panel C): ESC Headline: “Ticagrelor plus aspirin reduce ischaemic events in stable coronary patients with diabetes” but an excess of TIMI major bleeds. Hazard ratio 2.32 (95% CI 1.82-2.94) P<0.001. They then produced a post hoc endpoint “irreversible harm“ to claim net benefit in a PCI subgroup.

Revision 1

The age of randomized clinical trials

Subgroup analyses in randomized clinical trials: Value and limitations

Review #3 on important aspects of randomized clinical trials in cardiovascular pharmacotherapy

Word count: 4116 words (excl. Affiliations, References; incl title)