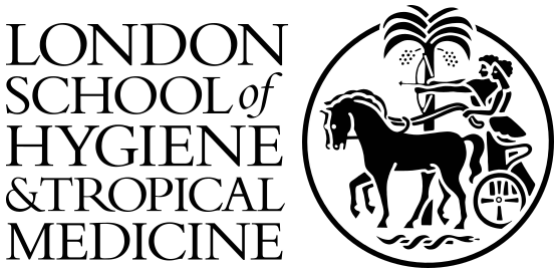


A Critical Appraisal of Late-Breaking Trials at the ACC Scientific Sessions, March 2018

Brief Title: A critical appraisal of late-breaking trials at the ACC 2018

Stuart Pocock PhD
Tim Collier MSc

Disclosure statement:
Stuart Pocock is on Steering Committees or Data Monitoring Committees for trials sponsored by Astra Zeneca, Bayer, Boehringer Ingleheim, Boston Scientific, Idorsia, Janssen, Medtronic, Novartis, Novo Nordisk, and Vifor. He also receives grant funding from Astra Zeneca and Merck.
Tim Collier is on Data Monitoring Committees sponsored by Daiichi Sankyo and Zoll,

Department of Medical Statistics, London School of Hygiene & Tropical Medicine
Corresponding Author details:
Stuart Pocock
Department of Medical Statistics
London School of Hygiene & Tropical Medicine
London WC1E 7HT
Telephone: +44 (0) 20 7927 2413
Fax: +44 (0) 20 7637 2853
E-mail: stuart.pocock@LSHTM.ac.uk

WORD COUNT: 6173

ABSTRACT
The Late-Breaking Clinical Trials presentations at the American College of Cardiology Scientific Sessions March 2018 are an important contribution to the field of cardiology. This article presents a constructive critical appraisal of seven key such studies: ODYSSEY, VEST, SECURE-PCI, TREAT, POISE, SMART-DATE and CVD REAL 2. For each one our aim is to document and interpret the main findings, noting particularly when "positive spin" appears to occur. Our aim is to provide a balanced account of each study, paying attention to both constructive new findings and study limitations. These topical examples also provide useful general insights on what to look for when critiquing clinical trial presentations and publications.

Key Words: Clinical Trials, Conference Presentations, Critical Appraisal

Condensed Abstract:
The Late-Breaking Clinical Trials presentations at the American College of Cardiology Scientific Sessions March 2018 are an important contribution to the field of cardiology. This article presents a constructive critical appraisal of seven key such studies: ODYSSEY, VEST, SECURE-PCI, TREAT, POISE, SMART-DATE and CVD REAL 2. For each our aim is to document and interpret the main findings, noting particularly when "positive spin" appears to occur. Our aim is to provide a balanced account of each study, including both new findings and study limitations. These examples provide useful general insights on how to critique clinical trial presentations and publications.

Abbreviations:

CHD	Coronary heart disease
DAPT	Dual antiplatelet therapy
MACE	Major adverse cardiac events
MI	Myocardial Infarction
NSTEMI	non-ST-Elevation Myocardial Infarction
oGLD	Other glucose lowering drugs
PCI	Percutaneous coronary intervention
SGLT2i	Sodium-glucose Cotransporter-2 Inhibitors
STEMI	ST-Elevation Myocardial Infarction
WCD	Wearable cardioverter defibrillator

Study Acronyms

ODYSSEY OUTCOMES: Evaluation of Cardiovascular Outcomes After an Acute Coronary Syndrome During Treatment With Alirocumab

VEST: Vest Prevention of Early Sudden Death Trial and VEST Registry

SECURE-PCI: Statins Evaluation in Coronary Procedures and Revascularization

TREAT: Ticagrelor in Patients with ST-Elevation Myocardial Infarction treated with Pharmacological Thrombolysis trial

POISE: PeriOperative ISchemic Evaluation trial

SMART-DATE: 6-month versus 12-month or longer dual antiplatelet therapy after percutaneous coronary intervention in patients with acute coronary syndrome: a randomised, open-label, non-inferiority trial

CVD-REAL 2: Comparative Effectiveness of Cardiovascular Outcomes in New Users of SGLT-2 Inhibitors

Introduction

Each year the American College of Cardiology (ACC) Scientific Sessions are a major forum for presentations of original findings across a broad spectrum of research activities in cardiology. Of particular interest are the Late-Breaking Clinical Trials sessions since they provide the latest pivotal evidence on both new and established treatment practices in cardiology.

This year on March 10 to 12 there were 8 such sessions in which 37 such studies were presented. To review all these studies would be an immense task, and hence we have selected 7 key presentations for us to provide a constructive critical appraisal. These were chosen as being 1) of major clinical importance and 2) within our sphere of expertise. They are listed in the Central Illustration.

For each study our aim is to place it in context, summarize the design, present the main findings and then provide a critical interpretation. We pay particular attention to the multiplicity of data available for presentation, and the consequent problems that arise e.g. in having multiple secondary endpoints or multiple subgroup analyses. Potential biases, e.g. in the one non-randomized study we review, are assessed.

There is a natural desire for trialists to wish to emphasize the more positive aspects of their study findings. This "positive spin" carries the risk that presentations may not provide a balanced account of the totality of evidence (1) . We point out instances when this appears to occur.

Overall we hope this article provides a meaningful commentary on some of the most topical (and sometimes controversial) presentations at ACC 2018.

ODYSSEY OUTCOMES Trial (2)

Alirocumab in acute coronary syndrome

This trial recruited 18,924 patients who 1) had an ACS event 1 to 12 months ago, 2) were on high-intensity statin therapy and 3) had inadequate control of lipids e.g. LDL cholesterol ≥70 mg/dl. Patients were randomized to alirocumab (a PCSK9 inhibitor) or placebo. The primary composite efficacy endpoint was coronary heart disease (CHD) death, non-fatal MI, fatal or non-fatal ischaemic stroke or unstable angina requiring hospitalization. As is common practice, let's call this MACE (Major Adverse Cardiovascular Events). Median follow-up was 2.8 years. As expected, patients on alirocumab had a marked reduction in LDL-cholesterol compared to placebo: -62.7% at 4 months which attenuated slightly to -54.7% at 4 years.

Results for the primary efficacy endpoint and its components are shown in the top half of Table 1. MACE had a highly significant 15% relative reduction (hazard ratio 0.85) with 95% confidence interval (CI) from 7% to 22% reduction, P=0.0003. All four components of MACE had fewer events on alirocumab compared with placebo, though for CHD death this was not significant.

It is relevant to express this primary result on an absolute scale. There were 149 fewer patients with a MACE event on alirocumab, out of 9462 patients per arm followed for a median 2.8 years. This translates into a reduction of 5.62 first MACE events per 1000 years of treatment with 95% CI 2.35 to 8.89 per 1000 patient years. This can be converted to a number needed to treat (NNT): to prevent one MACE event, one needs to treat 63 patients for a median of 2.8 years with 95% CI 41 to 141. This is helpful in elucidating whether an overall strategy of prescribing alirocumab to all eligible patients is sufficiently effective and in turn cost-effective.

There are several important considerations here:

1) We are confined to the trial's inevitably limited follow up so cannot generalize to the effects of longer-term treatment.

2) The plot of cumulative MACE events over time by treatment group (**Figure 1**) reveals no separation of the curves out to 1 year. This significant treatment-time interaction (P=0.03) means all the benefit appears to kick in after one year's treatment. This departure from proportional hazards calls into question whether a hazard ratio is the best overall summary of the treatment effect.
3) This absolute benefit will vary from patient to patient: that is, higher risk patients are liable to have a higher absolute benefit. For instance, the 27% of patients who were over 65 had a MACE rate around 55% higher than the rest. We would encourage the authors to undertake appropriate multivariable analysis so patients can be stratified according to their risk status (3). This will help refine which patients benefit the most from alirocumab treatment.

Now we turn to the main secondary endpoints (bottom half of **Table 1**), which are listed in a predefined order for hierarchical statistical testing (4). This is in order to keep the overall type-1 error at 0.05. The first four on the list were all highly significant, but CHD death and CV death were not (P=0.38 and P=0.15 respectively).

For all-cause death there is an observed 15% relative risk reduction (hazard ratio 0.85) with 95% CI from a 2% to a 27% reduction, P=0.026. But since this sits lower in the hierarchy of statistical testing it does not fit in the formal list of claims for treatment efficacy, within the bounds of strict type 1 error control. A counter-argument is that overall survival is clearly the most important matter for patients and hence merits special attention beyond statistical formalities. A weakness in this statement is that the all-cause death finding rests on combining non-significant reductions in both CV deaths and non CV deaths (31 and 27 fewer deaths respectively) and the latter has no obvious rationale.

The next concern is over the interpretation of subgroup analyses for the primary MACE outcome. For the five main pre-specified subgroups, there were no statistically significant interactions with treatment. This would normally be the end of the matter: insufficient evidence that there are any identifiable effect-modifiers. But in this case, the idea is pursued that alirocumab may be more effective in the 30% of patients who had baseline LDL- cholesterol greater than or equal to 100 mg/dl: the observed relative risk reduction becomes 24% (95% CI 13% to 35%), but it is questionable whether post hoc emphasis on this finding is justifiable (5,6)

.

Even more doubtful is the claim that all-cause mortality is reduced by 29% (with 95% CI 10% to 44%) in patients with LDL-cholesterol greater than or equal to 100 mg/dl. Such data dredging amongst subgroup analyses for a secondary endpoint has little merit.

Lastly, it is interesting to compare the main ODYSSEY findings with those of the FOURIER trial of evolocumab (7), another PCSK9 inhibitor. The two study populations are different: FOURIER focused on patients with a history of myocardial infarction, non-hemorrhagic stroke or peripheral artery disease. Nevertheless, some consistency of findings emerge. Both trials show that a PCSK9 inhibitor reduces the risk of myocardial infarction and ischemic stroke. Also, neither trial shows an effect on cardiovascular death. Inconsistencies are that ODYSSEY shows apparent reductions in both unstable angina and in all cause death whereas FOURIER does not. This weakens yet further the claim that alirocumab reduces mortality in ACS patients.

VEST EARLY PREVENTION OF SUDDEN DEATH trial (VEST) (8)

Wearable Cardioverter Defibrillator (WCD) in post-MI patients

The hypothesis posed for this trial is: can a WCD reduce the risk of sudden death in the immediate post-MI period (up to 90 days) in patients with reduced ejection fraction (EF). The trial recruited 2309 patients within 7 days of hospital discharge after acute MI who had EF ≤35%. They were randomized in a 2:1 ratio to WCD + guideline treatment (N=1,524) versus guideline treatment only (N=778) and then followed for 90 days.

Results for the primary outcome (sudden death) and several pre-defined fatal and non-fatal secondary outcomes are shown in **Table 2**. There is not a significant reduction in sudden death (P=0.18) and hence some have called this a "negative" trial. This we find too dismissive since the observed difference in incidence of sudden death (1.6% versus 2.4%) is in favour of WCD: a 32.8% relative reduction but with a wide 95% CI ranging from a 21.2% increase to a 62.8% decrease. A better term is to call the trial "inconclusive".

The problem is that the trial only has good statistical power to detect very marked treatment differences. For instance, had the total of 44 sudden deaths split 22 (1.4%) on WCD and 22 (2.8%) on control, then this hypothetical 50% risk reduction would have been significant with P=0.02.

Even if the trial had been twice as big (N=4604 patients) the observed 32.8% reduction would still only have P=0.06. It would require three times as many patients (N=6906) for such a risk reduction to achieve P=0.02.

This is the dilemma we face when undertaking trials of an intervention strategy (9), such as wearing a WCD in the VEST trial. Patient recruitment is much harder than in drug trials (in VEST it took almost 10 years to recruit 2302 patients) so that definitive evidence of efficacy is much harder to achieve.

A further issue is how well did patients comply with wearing the WCD; this averaged around 18 hours per day initially and declined to around 12 hours per day by 90 days (including non-users). Such reduced compliance over time must inevitably compromise the ability to prevent sudden deaths.

Among the pre-defined secondary outcomes (**Table 2**) the one that really matters is all-cause death with 90-day incidence of 3.1% on WCD versus 4.9% on control. This is a 35.5% relative risk reduction with 95% CI from 2.2% to 57.5% reduction, P=0.04.

It is a natural instinct to now label VEST as a "positive" trial. After all, surely a significant result for all-cause death justifies such a claim?! But a more cautious interpretation is warranted. First, the result is statistically fragile: if there had been just one less death in the control arm, P becomes >0.05. Second, all-cause death is not the primary outcome. Third, it seems illogical that the WCD is equally effective in preventing both sudden- and non-sudden deaths. Thus, while it plausible that a WCD really does reduce mortality, the VEST trial's evidence is not sufficiently convincing by itself.

SECURE-PCI Trial (10)

Loading dose of atorvastatin prior to planned PCI

This double-blind trial randomized 4191 patients with ACS who had an angiogram with the intention of planned PCI to either 280 mg loading doses of atorvastatin or matching placebo. All patients subsequently received 40 mg atorvastatin daily for 30 days. The primary MACE outcome was a composite of all-cause death, myocardial infarction, stroke and unplanned coronary revascularization through 30 days.

Table 3 shows the results for the primary outcome and its components. While there were numerically fewer primary events in the atorvastatin arm 130/2087 (6.2%) versus 149/2104

(7.1%) in the placebo arm, this did not achieve statistical significance: hazard ratio 0.88 with 95% CI 0.69 to 1.11, P=0.27.

Similarly, none of the components of the primary outcome showed a significant treatment effect.

However, this apparently "negative" primary result should not be interpreted as proof that loading doses of atorvastatin have no effect (11). The confidence interval extends out to a 31% risk reduction, and such uncertainty means that a type 2 error (a false negative) is possible. In addition, the fact that patients in both arms got 40 mg atorvastatin for 30 days may have diluted any effect of the loading doses per se.

Amongst several exploratory subgroup analyses there was no evidence of statistical interactions, except when comparing those who did or not undergo PCI (interaction P=0.02), see bottom of **Table 3** and **Figure 2.** For those 65% of patients who actually underwent PCI the hazard ratio is 0.72 (95% CI 0.54 to 0.96) subgroup P=0.02 whereas for the rest the apparent treatment effect is in the opposite direction: hazard ratio 1.36 (95% CI 0.89 to 2.09) subgroup P=0.15. The latter is a curious finding since it is counter-intuitive (perhaps due to chance): it is hard to imagine how loading doses of atorvastatin could increase cardiovascular risk in patients not undergoing PCI.

The subgroup claim that loading doses of atorvastatin substantially reduce the risk of MACE in patients actually undergoing PCI requires a cautious interpretation for the following reasons:

1) In any trial, post hoc emphasis on the most positive of several subgroup analyses tends to lead to an exaggeration of the true treatment effect.
2) The significant interaction (P=0.02) is reached because the observed effect for the non-PCI subgroup is in the opposite direction. Such qualitative interactions are inherently implausible.

3) The subgroup of patients undergoing PCI is an improper subgroup (12,13) in the sense that it was not known at the time of randomization. Patient factors determining who got PCI (or who did not) may have affected the outcome: in extreme a few patients may have had a very early MACE event before PCI could begin.
4) This subgroup finding is of little practical value, since for future patients the decision whether to give loading doses of atorvastatin needs to be taken before one knows whether the patient will actually undergo PCI. In the trial 35% of patients did not undergo PCI.

The perils of subgroup analyses are further illustrated when one does separate PCI/no PCI comparisons for STEMI and NSTEMI patients, see bottom part of **Figure 2**. For STEMI patients there is a significant interaction (P=0.04) whereas for NSTEMI patients there is not. At face value, this implies that the benefits of loading doses of atorvastatin are primarily confined to STEMI patients undergoing PCI. But commonsense suggests this arises from data dredging across multiple possible sub-group analyses and should not be taken seriously.

TREAT study (14)

Ticagrelor vs Clopidogrel after Fibrinolytic Therapy

Worldwide many STEMI patients receive fibrinolytic therapy rather than primary PCI, and hence there is a need to evaluate the relative safety and efficacy of different antiplatelet regimens in this context. In this open-label trial 3799 such patients were randomized to either ticagrelor (180 mg loading dose, 90 mg twice daily thereafter) or clopidogrel (300 mg or 600 mg loading dose, 75 mg daily thereafter). Median time from fibrinolysis to randomization was 11.4 hours, and 90% were pre-treaded with clopidogrel.

The primary outcome was TIMI major bleeding through 30 days. Secondary safety outcomes used other bleeding criteria. Also exploratory efficacy outcomes include the composite

of cardiovascular (CV) death, myocardial infarction (MI) and stroke. Results for the main safety and efficacy outcomes at 30 days follow-up are in **Table 4.**

The primary outcome TIMI major bleed occurred in 0.73% and 0.69% of ticagrelor and clopidogrel patients respectively, a difference of +0.04% with 95% CI from -0.49% to +0.58%. There was a pre-defined non-inferiority hypothesis with an absolute margin of +1.0%. Since the 95% CI excludes +1.0% a claim of non-inferiority of ticagrelor relative to clopidogrel can be made as regards TIMI major bleed. Similar conclusions can be made regarding the PLATO and BARC major bleeding criteria. However, for all bleeds there was a significant excess on ticagrelor compared with clopidogrel: absolute difference +1.57% with 95% CI +0.24% to +2.90%, P=0.02. The composite efficacy outcomes of CV death, MI and stroke had a similar incidence in both groups (4.0% ticagrelor, 4.3% clopidogrel P=0.57) and both groups had 49 (2.6%) deaths from all causes within 30 days.

With the proviso that ticagrelor had more minor bleeds, a conclusion that ticagrelor appears as good as clopidogrel in these patients seems appropriate. But a few outstanding issues remain:

1) This is a relatively young low-risk population of STEMI patients (e.g. age over 75 was excluded), and so the incidence of major bleeding is low.
2) The practical merit of demonstrating that ticagrelor is non-inferior to clopidogrel is open to debate, given the former is more expensive and the latter is more widely established.
3) Was the delayed timing of ticagrelor administration in this trial making the most of its potential? An alternative trial could have explored the relative safety and efficacy of ticagrelor given alongside fibrinolytic therapy.

POISE trial

Metoprolol in patients undergoing non-cardiac surgery

Whether beta-blockers are of benefit in patients undergoing non-cardiac surgery is a controversial and unresolved topic. Previous positive recommendations in both US and European guidelines were shaken when it was found that the key DECREASE studies were fundamentally flawed (15). The latest ESC/ESA guidelines (16) take a more cautious position and conclude "a high priority needs to be given to new randomised clinical trials to better identify which patients derive benefit from blocker therapy in the perioperative setting, and to determine the optimal method of beta blockade.

The POISE trial's previously published findings (17) on the short-term outcomes following extended-release metoprolol use alerted everyone to potential harms. The latest findings on one-year outcome are therefore of considerable interest.

POISE randomised 8351 patients with, or at risk of, atherosclerotic disease who were undergoing non-cardiac surgery to receive extended-release metoprolol or placebo, starting two to four hours before surgery and continuing for 30 days.

Table 5 presents results for the key outcomes for both 30 days follow up (previously reported) and 1-year follow-up (new).

The main benefit of extended metoprolol is a highly significant reduction in the incidence of MI: 63 fewer at 30 days (P=0.0017) which attenuated slightly to 52 fewer at 1 year (P=0.008). The absolute one-year reduction in incidence of MI (metoprolol vs placebo) is 1.24% with 95% confidence interval 0.26% to 2.23%.

In contrast, there is a significant excess of strokes on metoprolol: 22 more at 30 days (P=0.0053), and 26 more at 1 year (P=0.0014). The absolute one year increase in incidence of stroke is +0.62% with 95% CI 0.07% to 1.18%.

There is also a significant mortality excess on metoprolol: 32 more deaths at 30 days (P=0.032) which increases to 54 more deaths at one year (P=0.036). This is mainly driven by an excess of non-CV deaths (39 more on metoprolol, P=0.043), but there are also 15 more CV deaths on metoprolol, P=0.37. Overall, the absolute one year increase in mortality is +1.30% with 95% CI 0.06% to 2.54%.

The overall picture is that metoprolol appears to do more harm than good, the risks of death and stroke outweighing the benefits of fewer myocardial infarctions and coronary revascularizations. It has been suggested that this problem arose because the chosen dose regimen (based on 100mg oral extended-release metoprolol initially) was too high.
Whether there exists a more judicious choice of beta-blocker and dose regimen which can still be of benefit to appropriate patients undergoing non-cardiac surgery is an open question, which is only answerable by further large-scale randomised trials. As far as we know, no such trials are currently taking place.

SMART-DATE trial (17)

6 months versus 12+ months DAPT after PCI in ACS patients.

There is an extensive literature dedicated to determining the optimal duration of dual antiplatelet therapy (DAPT) after drug eluting stent (DES) implantation (18). In general a longer DAPT duration is liable to reduce the incidence of ischaemic events but is accompanied an increased risk of bleeding complications.

SMART-DATE (17) is the most recent trial to tackle this issue. A total of 2712 ACS patients undergoing PCI (99% got a DES) were randomised to either 6 months DAPT or 12 months or longer DAPT. The primary MACE endpoint was a composite of all-cause death, MI, or stroke at 18 months after PCI in the intention-to-treat population.

Results for the primary and predefined secondary endpoints are shown in **Table 6**. Let us first focus on the MACE primary endpoint which occurred in 63 (4.7%) and 56 (4.2%) patients in the 6-month DAPT group and 12-month DAPT group respectively. The trial primary hypothesis was non-inferiority of the former relative to the latter with a predefined margin of 2%. The observed difference is 0.5% with an upper one-sided 95% CI of 1.8%. Thus, formally the claim of non-inferiority is established.

However, several concerns exist regarding the interpretation of such a non-inferiority trial (19)

1) It is useful to calculate a two-sided 95% CI which in this case is from -0.05% to +2.05%. This corresponds to a one-sided Type-I error of 2.5% (rather than 5%) as is often done in non-inferiority trials. On this basis the CI includes the 2.0% margin and makes the trial inconclusive regarding non-inferiority
2) A 2% margin is very wide, and with a 4.5% MACE rate being anticipated is equivalent to a margin of 1.44 on a ratio scale i.e. ruling out such a 44% excess of events would not really be convincing.
3) Since everyone gets DAPT for the first six months, only MACE events occurring after 6 months are really relevant to the treatment comparison. This landmark analysis, which necessarily excludes patients having an event before 6 months, is shown in **Figure 3.** The MACE rate is now somewhat higher in the 6-month DAPT group: hazard ratio 1.69 (95% CI 0.97 to 2.94), P=0.07. The CI is very wide since the numbers of MACE events between 6 months and 18 months (not given) are relatively small. We are now getting close to significant inferiority of the 6-month DAPT arm, and any claim of non-inferiority is clearly ruled out.

Among the secondary endpoints in **Table 6**, prior trials and meta-analyses suggest it is wise to focus on MI, stent thrombosis and bleeding events. Here the 6-month DAPT group has 14 more MIs and 5 more stent thromboses than the 12 months DAPT group. This is counterbalanced by the former having 16 fewer BARC 2-5 bleeds and four fewer major bleeds. However, this is hard to interpret sensibly because of comment 3) above, i.e. the numbers include events before 6 months when the two groups were on identical DAPT treatment.

The meta-analysis by Giustino et al (18) based on 10 randomised trials and 32,215 patients estimated that a shorter duration significantly increased the risk of stent thrombosis (odds ratio 1.71, P=0.001) and myocardial infarction (odds ratio 1.39, P<0.001) and reduced the risk of clinically significant bleeding (odds ratio 0.63, P<0.001), and may reduce the mortality risk (odds ratio 0.87, P=0.07). Put more simply, with around 16,000 patients in each group shorter DAPT duration resulted in 111 more MIs, 63 more stent thromboses, 118 fewer clinically significant bleeds and 49 fewer deaths. Of course any meta-analysis has the potential to oversimplify findings from a heterogeneous mix of studies (e.g. the risk of stent thrombosis is lower in second generation DES), but only through combining evidence across all relevant trials can one reach robust conclusions on this issue.

While SMART-DATE appears a well conducted trial which can usefully contribute to future meta-analyses, it has too few patients and events to reach meaningful conclusions in its own right regarding the trade-off between poorer efficacy and improved safety by reducing the duration of DAPT.

CVD-REAL 2 study (20)

Comparative effectiveness of SGLT2 inhibitors for cardiovascular outcomes

This study examines whether initiation of SGLT2 inhibitors is associated with a lower risk of CV events compared with other glucose lowering drugs (oGLD) using “real world” data for over 400,000 patients with type 2 diabetes in three world regions. Before describing and interpreting the findings, it is relevant to make some general remarks on the pros and cons of this type of comparative effectiveness study.

The good news is that such studies can give access to very large numbers of patients because they use data from medical claims, primary care/hospital records and national registries. This facilitates more precise estimates of associations between treatments and patient outcomes than is possible in randomised trials which are inevitably of limited size. Secondly they reflect the "real world" practice of medicine unconstrained by the strict eligibility criteria of randomised controlled trials (RCTs). Thus, they have an aura of greater generalisability than in RCTs.

The bad news is that the so called "real world" is a "messy place" from the perspective of seeking robust, unbiased treatment comparisons (21-23). The biggest problem is the selection process that determines which treatment gets given to each patient. Without randomisation there is always a risk that patients on any given drug (or class of drug) have an underlying better average prognosis than others. One tries to correct for this by adjusting for known patient characteristics using propensity score methods, but the presence of residual confounding and hence bias in the results always remains a real possibility. Another limitation of “real world” data is that it is collected in a less reliable manner: both baseline features and patient outcomes are left to the discretion of each practicing physician e.g. for non-fatal events (e.g. myocardial infarction) and causes of death, robustness of definitions and centralised adjudication are lacking.

So now to the results of the CVD-REAL (24) and CVD-REAL2 study (20). The first study in 6 countries identified 166,033 eligible new users of an SGLT2i and 1,226,221 eligible

new users of oGLD. In every country the former tended to be younger, with less established CVD, less CKD and less frail. The second study CVD-REAL 2 includes 249,348 SGLT2i and 3,668,203 oGLD patients from a further six countries. This time the former tended to be younger, with less CKD, less frail, more on metformin and on statin and recruited more recently. To correct for these imbalances and other patient characteristics, propensity-matched pairs of patients were extracted leading to 235,064 patients for analysis in each group. In CVD-REAL 2 the SGLT2i use was 75% dapaglifozin, 9% empagliflozin and 12% others.

The main results of CVD-REAL 2 are in **Figure 4**. They show that overall, patients on SGLT2i have a lower risk of all-cause death (HR 0.51, 95% CI 0.37 to 0.70), hospitalization for heart failure (HR 0.64, 95% CI 0.50 to 0.82), myocardial infarction (HR 0.81, 95% CI 0.74 to 0.88) and stroke (HR 0.68, 95% CI 0.55 to 0.84). All four associations are highly significant, but the first three also show highly significant heterogeneity between countries, e.g. the associations were weaker in Korea, which contributed over 70% of patients.

There are two problems here. First, the random-effects meta-analysis used tends to weight countries with fewer patients more than is appropriate (25,26), e.g. the biggest apparent heart failure effect in Canada pulls the overall effect in a more positive direction. A fixed-effect meta-analysis would lead to more modest overall treatment differences for death, heart failure and stroke since it gives more weight to the largest country Korea which had much smaller treatment differences. Second, one suspects the heterogeneity across countries is more about their differing selection biases rather than genuine geographic variations in treatment effects.

These findings need to be seen in the context of key randomised trials of SGLT2is, namely EMPA-REG (27) and CANVAS (28). Both showed reductions in heart failure hospitalization compared to placebo, consistent with what CVD-REAL 2 shows. For myocardial

infarction and stroke the trials showed no significant benefit of SGLT2is which casts doubt on the causal validity of such associations in CVD-REAL 2. For all-cause death the overall claim of a 49% reduction in hazard seems liable to be an exaggeration, the 28% reduction in Korea seems plausible, given the known mortality reduction in EMPA-REG, but not in CANVAS.

Since dapagliflozin contributes most SGLT2i patients to this study, we clearly need to wait for the DECLARE trial (29,30) results before casting a final verdict on the believability of these findings in CVD-REAL 2. As always such observational studies can only study associations: whether they depict genuine beneficial treatment effects can only be established through evidence from randomised controlled trials.

DISCUSSION

In the process of reviewing these 7 late breaking clinical trials, we find some general issues worthy of comment. First, presentations of new evidence at major scientific meetings such as ACC are not peer reviewed. Hence the presenters and their collaborators have essentially a free rein to present their study findings as they see fit. Given they have often devoted years of effort in conducting a major trial there is a natural wish to present their findings "in a good light". On the whole, presenters of major studies are top-level highly-respected scientists for whom the quest for truth is paramount. Nevertheless, at the key moment of first presentation of pivotal findings it is only human nature to allow a degree of "positive spin" to creep in (1). This same temptation may also be felt by study sponsors, whether commercial or public bodies. In our **Central Illustration** the last two columns summarize how this may have occurred in these 7 specific presentations.

Second, a conference presentation is just the first step in the release and interpretation of study findings. The first peer-reviewed publication in a major medical journal is what matters in

the long run. Indeed 4 of the 7 studies we have reviewed had simultaneous publications. Given the high standards set by journal editors and reviewers and use of CONSORT guidelines (31), such publications are less prone to "positive spin": the conclusions usually focus on the predefined primary outcome in the whole trial population, with other possible claims on secondary endpoints and subgroup analyses referred to as exploratory findings. Sometimes this practice seems a little too restrictive, as if any ideas after database lock can have no bearing on what should be future clinical practice. But there is a fine line to be drawn between flexibility (let all the data speak) and selectivity (how can I make my trial more "positive"?).

Of course, journal publications are of finite size and are often written quickly to meet conference deadlines, so they in turn do not necessarily provide the "whole truth". For drug or device trials much more detailed regulatory dossier get presented to the FDA, EMA and other agencies. That is where the final detailed totality of evidence gets judged.

But the importance and excitement of a first conference presentation is hard to overstate. Thus, we greatly appreciate this opportunity to review some of the key trials of ACC 2018, and hope our insights are a stimulus to further discussion on what they each mean for future best clinical practice.

References

1. Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P. Impact of Spin in the Abstracts of Articles Reporting Results of Randomized Controlled Trials in the Field of Cancer: The SPIIN Randomized Controlled Trial. Journal of Clinical Oncology 2014;32:4120-4126.
2. Schwartz GG, Bessac L, Berdan LG, et al. Effect of alirocumab, a monoclonal antibody to PCSK9, on long-term cardiovascular outcomes following acute coronary syndromes: Rationale and design of the ODYSSEY Outcomes trial. Am Heart J;168:682-689.e1.
3. Pocock SJ, Lubsen J. More on Subgroup Analyses in Clinical Trials. N Engl J Med 2008;358:2076-2077.
4. Food and Drug Administration. Multiple Endpoints in Clinical Trials: Guidance for Industry. FDA Maryland, 2017.
5. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. The Lancet 2000;355:1064-1069.
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 2007;357:2189-2194.
7. Sabatine MS, Giugliano RP, Keech AC, et al. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. New England Journal of Medicine 2017;376:1713-1722.
8. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2000 Feb 29 - Identifier NCT01446965, Vest Prevention of Early Sudden Death Trial and VEST Registry (VEST) 2011 Oct 5 [cited 2018 Mar 23];. Available from: http://clinicaltrials.gov/ct/show/NCT01446965?order=1.

9. Pocock SJ, Clayton TC, Stone GW. Challenging Issues in Clinical Trial Design: Part 4 of a 4-Part Series on Statistics for Clinical Trials. J Am Coll Cardiol 2015;66:2886-2898.
10. Berwanger O, Santucci E, de Barros e Silva P, et al. Effect of loading dose of atorvastatin prior to planned percutaneous coronary intervention on major adverse cardiovascular events in acute coronary syndrome: The secure-pci randomized clinical trial. JAMA 2018.
11. Pocock SJ, Stone GW. The Primary Outcome Fails — What Next? New England Journal of Medicine 2016;375:861-870.
12. Hirji KF, Fagerland MW. Outcome based subgroup analysis: a neglected concern. Trials 2009;10:33.
13. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. Jama 1991;266:93-8.
14. TREAT Study Group. Ticagrelor vs clopidogrel after fibrinolytic therapy in patients with ST-elevation myocardial infarction: A randomized clinical trial. JAMA Cardiology 2018.
15. Lüscher TF, Gersh B, Landmesser U, Ruschitzka F. Is the panic about beta-blockers in perioperative care justified? Eur Heart J 2014;35:2442-2444.
16. Kristensen SD, Knuuti J, Saraste A, et al. 2014 ESC/ESA Guidelines on non-cardiac surgery: cardiovascular assessment and management. Eur Heart J 2014;35:2383-2431.
17. Hahn J-Y, Song YB, Oh J-H, et al. 6-month versus 12-month or longer dual antiplatelet therapy after percutaneous coronary intervention in patients with acute coronary syndrome (SMART-DATE): a randomised, open-label, non-inferiority trial. The Lancet.

18. Giustino G, Baber U, Sartori S, et al. Duration of Dual Antiplatelet Therapy After Drug-Eluting Stent Implantation. A Systematic Review and Meta-Analysis of Randomized Controlled Trials 2015;65:1298-1310.

19. Macaya F, Ryan N, Salinas P, Pocock SJ. Challenges in the Design and Interpretation of Noninferiority Trials. Insights From Recent Stent Trials 2017;70:894-903.

20. Kosiborod M, Lam CSP, Kohsaka S, et al. Lower Cardiovascular Risk Associated with SGLT-2i in >400,000 Patients: The CVD-REAL 2 Study. J Am Coll Cardiol 2018.

21. Brown ML, Gersh BJ, Holmes DR, Bailey KR, Sundt TM, 3rd. From randomized trials to registry studies: translating data into clinical information. Nat Clin Pract Cardiovasc Med 2008;5:613-20.

22. Freemantle N, Marston L, Walters K, Wood J, Reynolds MR, Petersen I. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. BMJ : British Medical Journal 2013;347.

23. Pocock SJ, Elbourne DR. Randomized Trials or Observational Tribulations? N Engl J Med 2000;342:1907-1909.

24. Kosiborod M, Cavender MA, Fu AZ, et al. Lower Risk of Heart Failure and Death in Patients Initiated on SGLT-2 Inhibitors Versus Other Glucose-Lowering Drugs: The CVD-REAL Study. Circulation 2017.

25. M. Borenstein LVH, J.P.T. Higgins and H.R. Rothstein. Introduction to Meta-Analysis. Chichester, West Sussex, United Kingdom: Wiley, 2009.

26. Poole C, Greenland S. Random-effects meta-analyses are not always conservative. Am J Epidemiol 1999;150:469-75.

27. Zinman B, Wanner C, Lachin JM, et al. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. New England Journal of Medicine 2015;373:2117-2128.

28. Neal B, Perkovic V, Mahaffey KW, et al. Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes. N Engl J Med 2017;377:644-657.

29. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2000 Feb 29- Identifier NCT01730534, Multicenter Trial to Evaluate the Effect of Dapagliflozin on the Incidence of Cardiovascular Events (DECLARE-TIMI58) 2012 Nov 21 [cited 2018 Mar 23];. Available from: http://clinicaltrials.gov/ct/show/NCT01730534?order=1.

30. Raz I, Mosenzon O, Bonaca MP, et al. DECLARE-TIMI 58: Participants' baseline characteristics. Diabetes Obes Metab 2018.

31. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c869.

Table 1: Primary and Secondary Efficacy Endpoints in the ODYSSEY OUTCOMES Trial

Endpoint, n (%)	Alirocumab (N=9462)	Placebo (N=9462)	HR (95% CI)	Log-rank P-value
MACE	903 (9.5)	1052 (11.1)	0.85 (0.78, 0.93)	0.0003
CHD death	205 (2.2)	222 (2.3)	0.92 (0.76, 1.11)	0.38
Non-fatal MI	626 (6.6)	722 (7.6)	0.86 (0.77, 0.96)	0.006
Ischemic stroke	111 (1.2)	152 (1.6)	0.73 (0.57, 0.93)	0.01
Unstable angina	37 (0.4)	60 (0.6)	0.61 (0.41, 0.92)	0.02
CHD event	1199 (12.7)	1349 (14.3)	0.88 (0.81, 0.95)	0.001
Major CHD event	793 (8.4)	899 (9.5)	0.88 (0.80, 0.96)	0.006
CV event	1301 (13.7)	1474 (15.6)	0.87 (0.81, 0.94)	0.0003
Death, MI, ischemic stroke	973 (10.3)	1126 (11.9)	0.86 (0.79, 0.93)	0.0003
CHD death	205 (2.2)	222 (2.3)	0.92 (0.76, 1.11)	0.38
CV death	240 (2.5)	271 (2.9)	0.88 (0.74, 1.05)	0.15
All-cause death	334 (3.5)	392 (4.1)	0.85 (0.73, 0.98)	0.026*

Table 2: Results of the VEST trial over 90 days of follow-up

Clinical event type	WCD (N=1524)	Control (N=778)	P-value
FATAL EVENTS, n (%)			
Sudden Death (primary outcome)	25 (1.6%)	19 (2.4%)	0.18
Non-sudden death	21 (1.4%)	17 (2.2%)	0.15
Congestive heart failure death	10 (0.7%)	5 (0.6%)	1.00
Recurrent MI death	1 (0.1%)	1 (0.1%)	1.00
Stroke death	0 (0.0%)	4 (0.5%)	0.01
Other cardiovascular death	5 (0.3%)	3 (0.4%)	1.00
Other death	5 (0.3%)	4 (0.5%)	0.72
Indeterminate death	2 (0.1%)	2 (0.3%)	0.83
Death, any cause	48 (3.1%)	38 (4.9%)	0.04
NON-FATAL EVENTS, n (%)			
Rehospitalization, cardiovascular	334 (22%)	174 (22%)	0.81
Rehospitalization, any cause	475 (31%)	253 (33%)	0.51

Table 3: 30 Day Outcomes Overall and in Patients Undergoing and Not Undergoing PCI in the SECURE-PCI trial

	No. /Total No. (%)		Absolute Difference,	Hazard Ratio	
Outcomes	**Atorvastatin**	**Placebo**	**% (95% CI)**	**(95% CI)**	**P value**
Primary Outcome at 30 d					
MACE	130/2087 (6.2)	149/2104 (7.1)	0.85 (−0.70 to 2.41)	0.88 (0.69-1.11)	.27
Components of Primary Outcome at 30 d					
Death	67/2087 (3.2)	70/2104 (3.3)	0.12 (−1.01 to 1.24)	0.97 (0.69-1.35)	.84
Cardiovascular death	59/2087 (2.8)	61/2104 (2.9)	0.07 (−0.99 to 1.13)	0.98 (0.68-1.40)	.90
Myocardial infarction	61/2087 (2.9)	77/2104 (3.7)	0.74 (−0.39 to 1.86)	0.80 (0.57-1.11)	.18
Peri-PCI	42/2087 (2.0)	54/2104 (2.6)	0.55 (−0.40 to 1.51)	0.78 (0.52-1.17)	.23
Non–PCI-related	20/2087 (1.0)	26/2104 (1.2)	0.28 (−0.40 to 0.96)	0.77 (0.43-1.39)	.39
Coronary revascularization	11/2087 (0.5)	14/2104 (0.7)	0.14 (−0.38 to 0.65)	0.79 (0.36-1.75)	.57
Urgent or target vessel	5/2087 (0.2)	9/2104 (0.4)	0.19 (−0.21 to 0.58)	0.56 (0.19-1.67)	.30
Stroke	10/2087 (0.5)	11/2104 (0.5)	0.04 (−0.43 to 0.51)	0.92 (0.39-2.16)	.85
Stent thrombosis	7/2087 (0.3)	15/2104 (0.7)	0.38 (−0.11 to 0.86)	0.47 (0.19-1.15)	.10
Exploratory Analysis of 30 day MACE in Subgroup of Patients Undergoing and Not Undergoing PCI					
PCI	81/1351 (6.0)	112/1359 (8.2)	2.25 (0.24 to 4.25)	0.72 (0.54-0.96)	.02
No PCI	49/734 (6.7)	37/743 (5.0)	−1.70 (−4.22 to 0.83)	1.36 (0.89-2.09)	.15

Table 4: Main Safety and Efficacy Outcomes at 30 days in the TREAT study

	No. (%)				
	Ticagrelor (n = 1913)		Clopidogrel (n = 1886)		Absolute Difference, % (95% CI)
Safety Outcomes					
TIMI major bleeding	14	(0.73)	13	(0.69)	0.04 (-0.49 to 0.58)
PLATO major bleeding	23	(1.20)	26	(1.38)	-0.18 (-0.89 to 0.54)
BARC type 3-5 bleeding	23	(1.20)	26	(1.38)	-0.18 (-0.89 to 0.54)
Any bleeding	103	(5.38)	72	(3.82)	1.57 (0.24 to 2.90)
Efficacy Outcomes					
CV death, MI, or stroke	76	(3.97)	82	(4.35)	-0.38 (-1.65 to 0.90)
All-cause death	49	(2.56)	49	(2.60)	-0.04 (-1.04 to 0.97)

Table 5: Results for key fatal and non-fatal outcomes in the POISE trial for both 30 day and 1 year follow-up.

	30 days			1 year +		
	Metoprolol [N=4174]	Placebo [N=4177]		Metoprolol [N=4174]	Placebo [N=4177]	
CV death, MI or stroke*	244	290	P=0.040			
MI	176	239	P=0.0017	208	260	P=0.008
Cardiac Revascularisation	11	27	P=0.012	21	45	P=0.004
Stroke	41	19	P=0.0053	85	59	P=0.014
All-cause death	129	97	P=0.032	410	356	P=0.036
CV death	75	58	P=0.14	182	167	P=0.37
Non-CV death	54	39	P=0.12	228	189	P=0.043

* The pre-defined primary composite outcome at 30 days.

+ One-year follow-up was achieved in around 90% of patients

Table 6: Clinical primary and secondary outcomes at 18 months in the SMART-DATE trial

	6-month DAPT group (n=1357)	12-month or longer DAPT group (n=1355)	HR (95% CI)	p value
MACE (primary)	63 (4·7%)	56 (4·2%)	1·13 (0·79–1·62)	0·51
Death	35 (2·6%)	39 (2·9%)	0·90 (0·57–1·42)	0·90
Myocardial infarction: total	24 (1·8%)	10 (0·8%)	2·41 (1·15–5·05)	0·02
Target vessel MI	14 (1·1%)	7 (0·5%)	2·01 (0·81–4·97)	0·13
Non-target vessel MI	10 (0·8%)	3 (0·2%)	3·35 (0·92–12·18)	0·07
Stroke	11 (0·8%)	12 (0·9%)	0·92 (0·41–2·08)	0·84
Cardiac death	18 (1·4%)	24 (1·8%)	0·75 (0·41–1·38)	0·36
Cardiac death or myocardial infarction	39 (2·9%)	32 (2·4%)	1·22 (0·77–1·95)	0·40
Stent thrombosis	15 (1·1%)	10 (0·7%)	1·50 (0·68–3·35)	0·32
BARC type 2–5 bleeding	35 (2·7%)	51 (3·9%)	0·69 (0·45–1·05)	0·09
Major bleeding	6 (0·5%)	10 (0·8%)	0·60 (0·22–1·65)	0·33
Net adverse clinical and cerebral events*	96 (7·2%)	99 (7·4%)	0·97 (0·73–1·29)	0·84

Percentages are Kaplan-Meier estimates. We defined MACE as a composite of all-cause mortality, myocardial infarction, and stroke. HR=hazard ratio. BARC=Bleeding Academic Research Consortium. *Net adverse clinical and cerebral events were defined as MACE plus BARC type 2–5 bleeding.

Central Illustration: 7 Key Studies in ACC Late-Breaking Sessions

study	treatments	patients	main finding	positive spin	caution
ODYSSEY	alirocumab vs placebo	ACS	15% reduction in MACE P=0.0003	29% mortality reduction for LDL≥100mg/dl	post hoc subgroup in secondary endpoint no significant interaction
VEST	wearable cardioverter defibrillator vs control	post-MI with EF≤35%	no significant reduction in sudden death	35.5% mortality reduction P=0.04	secondary endpoint underpowered
SECURE-PCI	atorvastatin loading dose vs placebo	ACS with planned PCI	no significant reduction in MACE	38% MACE reduction after undergoing PCI P=0.02	an improper subgroup analysis
TREAT	ticagrelor vs clopidogrel	STEMI patients after fibrinolysis	no treatment differences	ticagrelor non-inferior re major bleed	a claim of little practical value?
POISE	metoprolol vs placebo	non-cardiac surgery at high CV risk	1.3% excess 1-year mortality P=0.032	none	metoprolol dose too high?
SMART-DATE	6 months DAPT vs 12 months DAPT	after DES in ACS patients	no significant difference in MACE	6-months DAPT non-inferior	trial too small
CVD REAL 2	SGLT2i vs other antiglycaemic drugs	type 2 diabetes	lower event rates for death, HF, MI, stroke	wide ranging benefits of SGLT2i	observational study selection biases likely

Figure 1 Primary efficacy Endpoint MACE from the ODYSSEY OUTCOMES trial

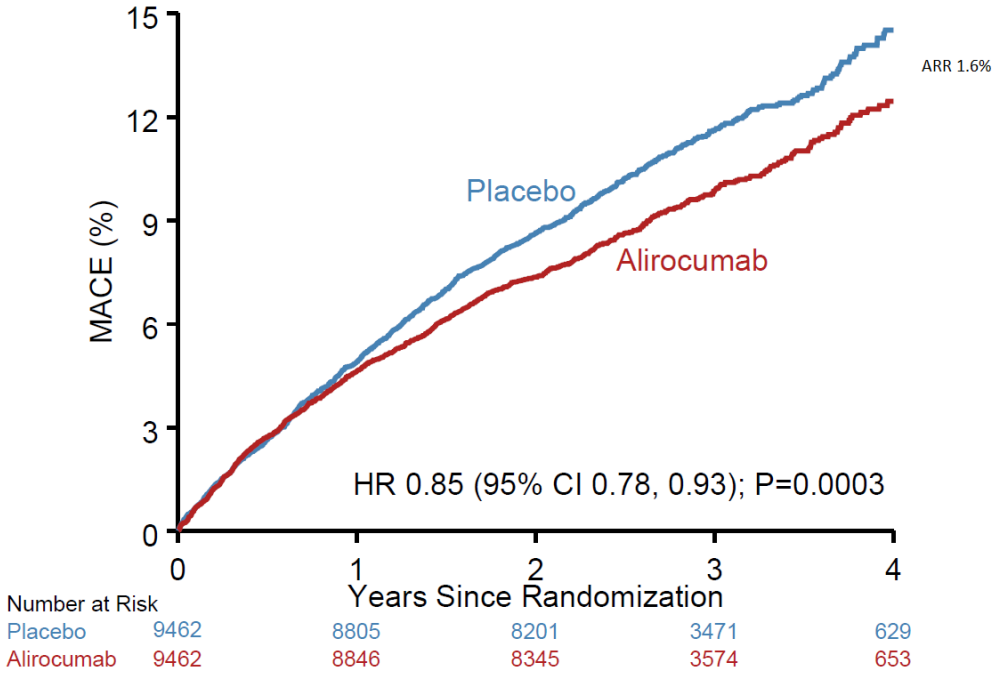


Figure 1 Primary efficacy endpoint MACE from the ODYSSEY OUTCOMES trial

Figure 1 CAPTION: Kaplan-Meier plot of cumulative MACE events (CHD death, non-fatal MI, ischemic stroke, or unstable angina requiring hospitalization) by treatment group. The figure reveals a significant time-treatment interaction whereby all the benefit appears to kick in after one year's treatment. This departure from the proportional hazards assumption calls into question whether a hazard ratio is the best overall summary of the treatment effect.

Figure 2 Hazard ratios for MACE outcome both overall and for key subgroups in the SECURE-PCI trial

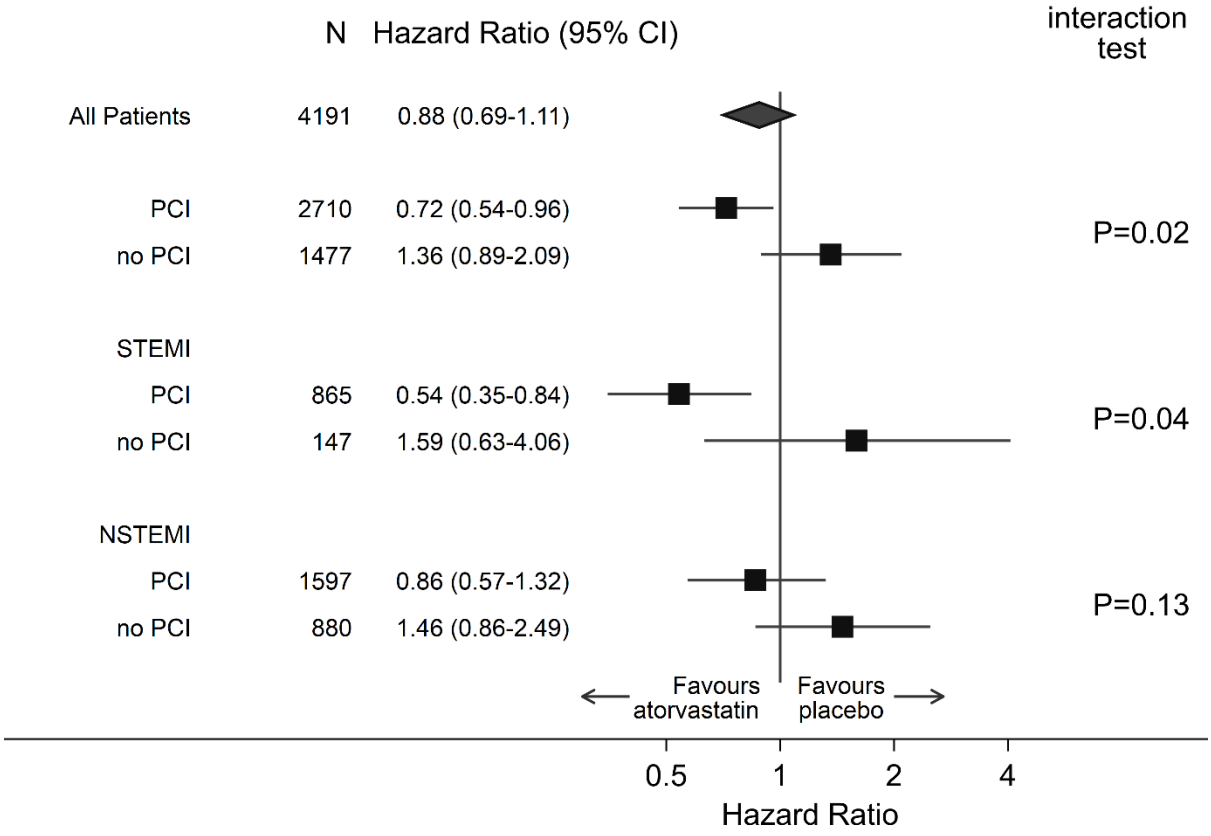


Figure 2 CAPTION: Forest plot showing the hazard ratio for MACE in all patients (HR 0.88, 95% CI 0.69-1.11) and in patients undergoing PCI (HR 0.72, 95% CI 0.54-0.96) or not (HR 1.36, 95% CI 0.89-2.09). This was the only significant interaction (P=0.02) among several exploratory subgroup analyses. But undergoing PCI is an improper subgroup being unknown at the time of randomization. Further PCI v no PCI comparisons are presented for STEMI (interaction P=0.04) and NSTEMI (interaction P=0.13) patients. Such multiple subgroup analyses require a cautious interpretation.

Figure 3: Landmark analysis of MACE at 6 months after the index procedure in the SMART-DATE trial

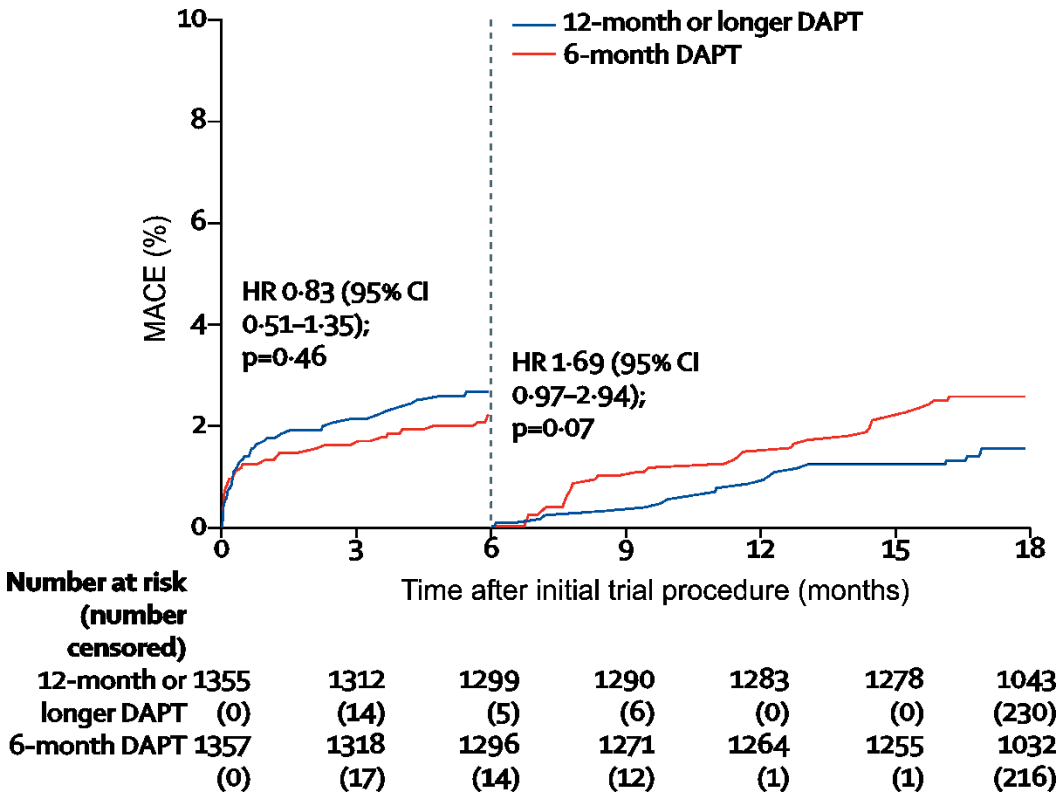
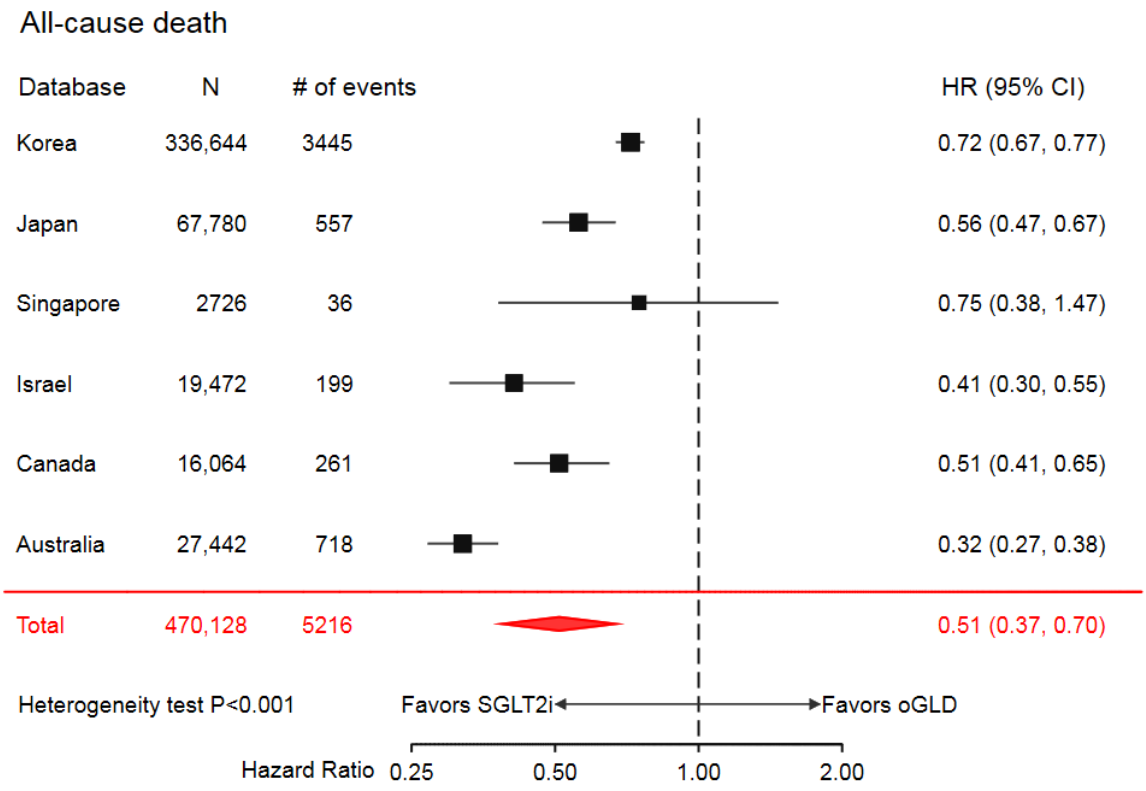
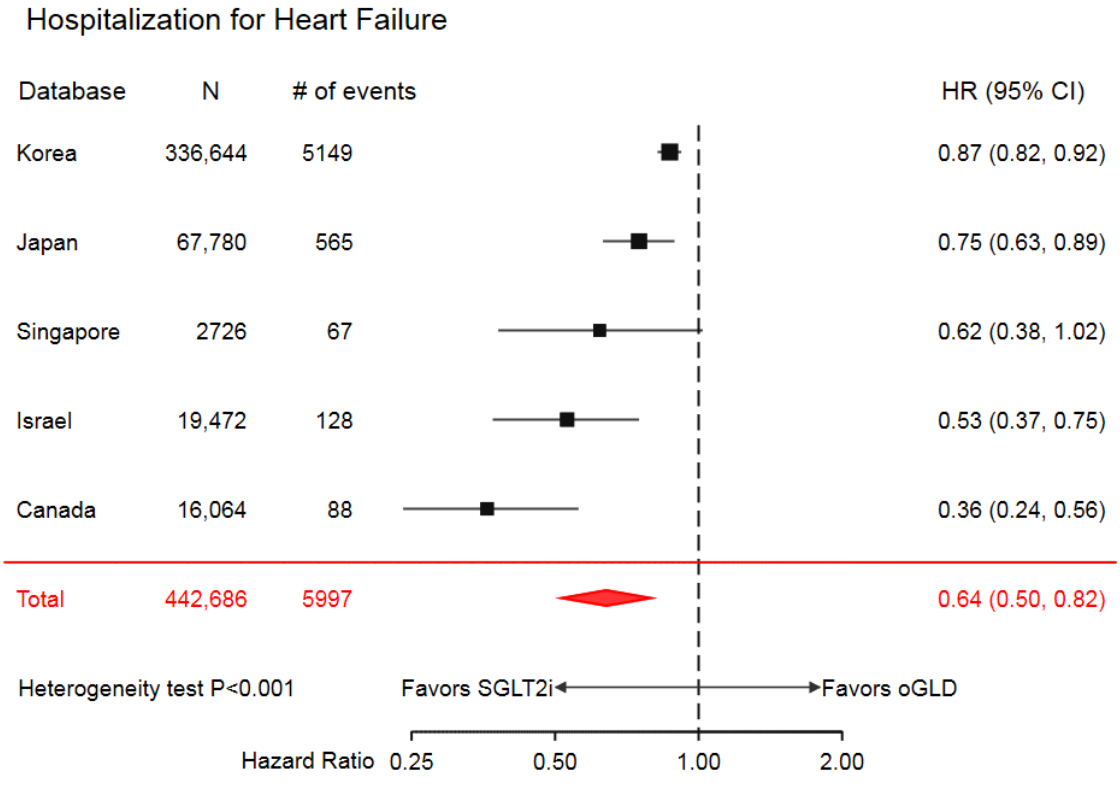


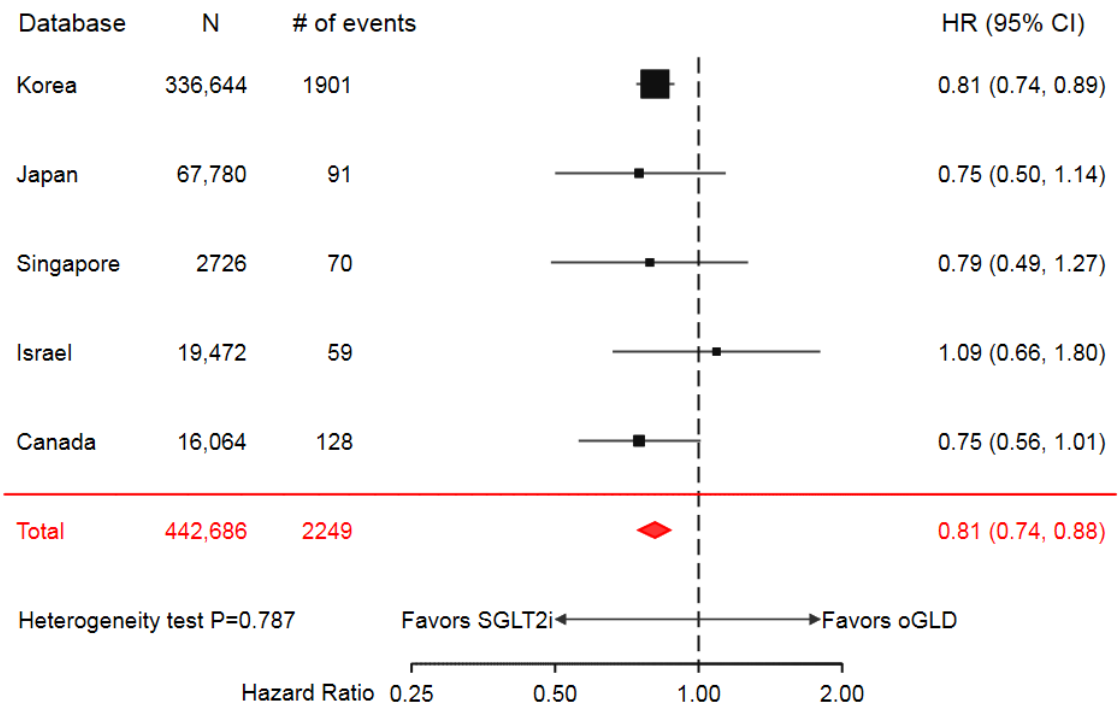
Figure 3 CAPTION: Kaplan-Meier plot of cumulative MACE events over time by treatment group with landmark analysis at 6 months in the SMART-DATE trial. Since everyone gets DAPT for the first six months, only MACE events occurring after 6 months are really relevant to the treatment comparison. The MACE rate after 6 months is somewhat higher in the 6-month DAPT group: hazard ratio 1.69 (95% CI 0.97 to 2.94), P=0.07. The CI is very wide since the numbers of MACE events between 6 months and 18 months are relatively small. We are now close to significant inferiority of the 6-month DAPT arm, and any claim of non-inferiority is clearly ruled out.

Figure 4 Forest plots from random-effects meta-analyses of outcomes in the CVD-REAL 2 study





Myocardial Infarction



Stroke

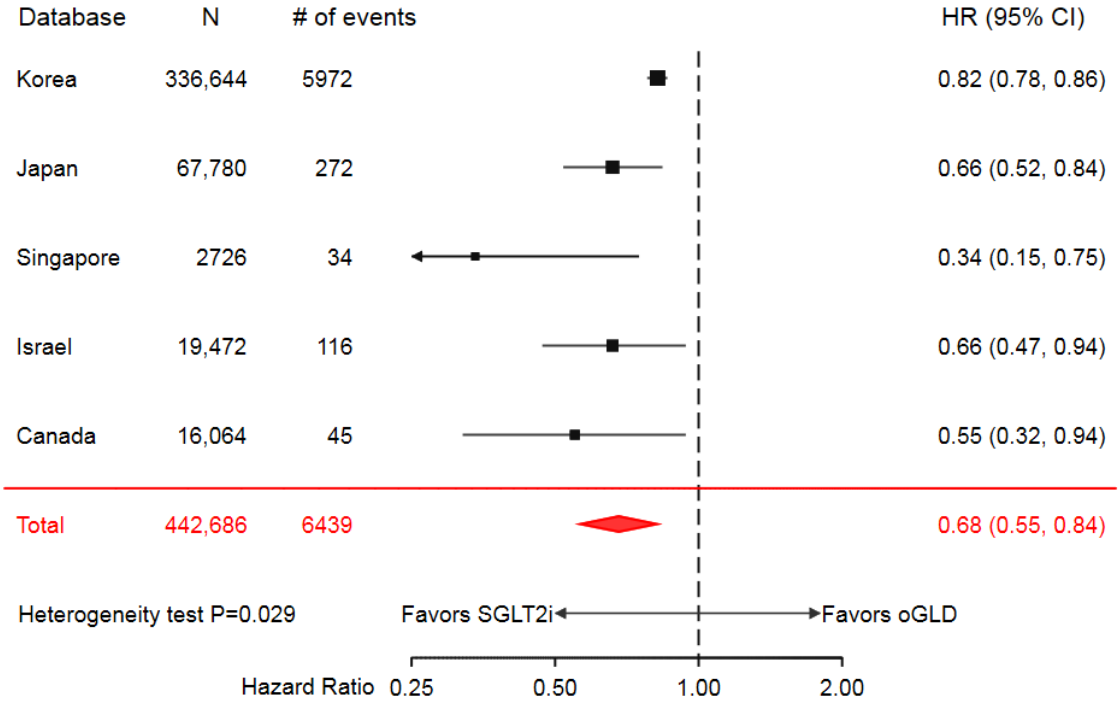


Figure 4 CAPTION: Forest plots from random-effects meta-analyses showing superiority of SGLT2i v oGLD for the outcomes of all-cause death (P<0.001), hospitalization for heart failure (P<0.001), myocardial infarction (P<0.001), and stroke (P<0.001) in the CVD-REAL 2 study. Heterogeneity tests were statistically significant for all-cause death (P<0.001), hospitalisation for heart failure (P<0.001) and stroke (P=0.029). The associations were generally weaker in Korea, which contributed 70% of the data. Such non-randomized treatment comparisons require a cautious interpretation because of the potential for selection bias.