

**Statistical Appraisal of Recent Clinical Cardiology Trials:
ATTR-ACT, AUGUSTUS, CABANA, COAPT, DECLARE, and REDUCE-IT**

Brief title:

Stuart J. Pocock, PHD,^a Tim J. Collier, MSc^a

From the: ^aDepartment of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom.

Address for correspondence:

Prof. Stuart J. Pocock
London School of Hygiene and Tropical Medicine
Keppel Street
London, WC1E 7HT
United Kingdom
Telephone: +44 (0)20 7927 2413
Fax: 44-207 637 2853
E-mail: stuart.pocock@lshtm.ac.uk
Twitter: @pocock_stuart

Disclosures: Dr. Pocock has reported serving on steering committees or data monitoring committees for trials sponsored by Abbott Vascular, Amiring, AstraZeneca, Bayer, Biosensors, Boehringer Ingelheim, Boston Scientific, Bristol Meyers Squibb, Edwards, Idorsia, Medtronic, Novartis and Vifor. Dr. Collier has reported serving on data monitoring committees sponsored by AstraZeneca, Boston Scientific, Daiichi SanKyo, Devax, Infraredx, Medtronic, Pfizer, and Zoll.

Acknowledgments: The authors are grateful to Pfizer for permission to include previously unpublished numbers for the ATTR-ACT trial in Table 1 and Figure 2.

ABSTRACT

In the past 12 months, many important new clinical trials in cardiology have had their first conference presentation and publication. This paper presents a constructive critical appraisal of 6 key studies. In time order of first presentation they are CABANA, ATTR-ACT, COAPT, DECLARE, REDUCE-IT, and AUGUSTUS. For each study, our aim is to document and interpret the main findings, paying attention to new findings, their research context, and study limitations. These topical examples also provide methodological insights pertinent to future clinical trials research.

CONDENSED ABSTRACT

In the past 12 months, many important new clinical trials in cardiology have had their first conference presentation and publication. This paper presents a constructive critical appraisal of 6 key studies. For each study, our aim is to document and interpret the main findings, paying attention to new findings, their research context, and study limitations. These topical examples also provide methodological insights pertinent to future clinical trials research.

Key words: Clinical Trials, Intention to Treat Analysis, Patient Selection, Statistics, Treatment Outcome

Abbreviations and Acronyms

CI = confidence interval

CRNM = clinically relevant nonmajor

CV = cardiovascular

CVH = cardiovascular-related hospitalizations

EPA = eicosapentaic acid

HFH = heart failure hospitalization

HR = hazard ratio

ITT = intention to treat

MACE = major adverse cardiovascular events

MI = myocardial infarction

PCI = percutaneous coronary intervention

SGLT2i = sodium-glucose cotransporter-2 inhibitors

VKA = vitamin K agonist

Introduction

A year ago, we published a critical appraisal of some late-breaking trials at the 2018 American College of Cardiology (ACC) Scientific Sessions (1). Rather than just repeat that for ACC 2019, we consider it more useful to take a broader perspective of key trials first presented and published in the past 12 months. Thus, in addition to ACC 2019, we cover trials from the annual 2018 Scientific Sessions of the Heart Rhythm Society, the European Society of Cardiology, the Transcatheter Cardiovascular Therapeutics, and the American Heart Association.

We have selected 6 key recent trials (**Central Illustration**) for our constructive critical appraisal. These trials were chosen for a mix of reasons. They: 1) are of major clinical importance; 2) are within our sphere of expertise; 3) raise challenges in their interpretation; and 4) contain valuable insights for the methodology of clinical trials research.

CABANA trial

The CABANA (Catheter Ablation vs. ANtiarrhythmic Drug Therapy for Atrial Fibrillation: NCT00911508) trial (2) randomized 2,204 patients with atrial fibrillation to either catheter ablation or antiarrhythmic drug therapy. The primary composite endpoint was death, disabling stroke, serious bleeding, or cardiac arrest. Key secondary outcomes were all-cause death and the composite of death and cardiovascular hospitalization (CVH). Over a median 4 years follow-up, findings for these outcomes using analysis by intention to treat (ITT) are summarized in the top part of **Figure 1**. Neither the primary endpoint nor all-cause death showed a significant treatment difference, though there were numerically fewer primary endpoints and deaths in the ablation arm. For the composite of death and CVH, ITT analysis did show a significant reduction in the ablation arm ($p = 0.002$).

The main problem in interpreting CABANA is the high rates of treatment crossovers: 9.2% of patients randomized to ablation did not receive it, whereas 27.5% of patients randomized to drug therapy did in fact receive ablation. This means that ITT analysis will

inevitably dilute away any true benefits of ablation. The widely accepted argument in favor of ITT analysis is that it achieves an unbiased comparison of the 2 therapeutic strategies as actually carried out. But the key flaw in the conduct of CABANA is the apparent lack of clinical equipoise: over one-quarter of investigators and their patients in the drug therapy arm received ablation, even though the whole purpose of the trial is that there is insufficient prior evidence that ablation is effective. In this context, the CABANA investigators have presented some alternative analyses (**Figure 1**). The precise methods undertaken are not fully explained, but our interpretation is as follows. **Analysis by treatment received** compares patients who actually received ablation with those who did not, regardless of which group they were randomized to. This approach shows highly significant differences in favor of ablation for the primary endpoint (hazard ratio [HR]: 0.67; $p = 0.006$) and all-cause death (HR: 0.60; $p = 0.005$). **Per protocol analysis** keeps patients in their randomized groups, but it removes from analysis those who crossed over. Only results for the primary endpoint are presented, with HR = 0.73; $p = 0.046$.

Of course, all these alternative analyses may carry a substantial bias. The patients who crossover are not a random sample of all the patients in each randomized group, and such crossovers may carry a higher (or lower) risk of getting fatal and nonfatal events not directly related to the actual intervention received. Thus, such exploratory analyses, although supportive of the hypothesis that ablation is an effective treatment, cannot be considered definitive evidence. Overall, we believe that the evidence from CABANA is inconclusive.

The underlying problem is that CABANA may well have been done too late in the development of catheter ablation for atrial fibrillation. In general, for any new procedure, there is first a learning curve, while the technical details and the interventionists' (or surgeons') skills are perfected, when it is too soon to undertake the pivotal randomized trial. There then follows a "window of opportunity" before the procedure gets widely accepted (3). In this case, that opportunity for unbiased randomized evidence appears to have been missed.

Perhaps this is an example of Buxton's law: "it is always too early for rigorous evaluation, until unfortunately it's suddenly too late."

In general, there remains controversy regarding the value of such "as treated" and "per protocol" analyses (4,5). To dismiss them as irrelevant is too harsh a stance, whereas to accept them as pivotal evidence is inappropriate. The key is to define exactly what one wishes to estimate (called the estimand) (6) and then to define the appropriate analysis strategy and any assumptions that it relies on.

Given the CABANA trial's problems, what additional evidence can be drawn on to clarify the merit (or not) of ablation therapy in atrial fibrillation? Of note is a large observational study using the OPTUM administrative database (7). They identified 135,688 patients who would have been eligible for CABANA, of whom 6,907 patients underwent ablation. Mean follow-up after ablation was 1.5 years. Using a propensity score weighted analysis to adjust for potential confounders, they found patients receiving ablation had a lower risk of all-cause death (HR: 0.60; $p < 0.001$) and the composite of death, stroke, major bleed, and cardiac arrest (HR: 0.70; $p < 0.001$), the latter being the primary endpoint in CABANA (**Figure 1**). As in all such observational studies, one cannot rule out a selection bias whereby there are unmeasured confounders influencing who received ablation. Therefore, it is open to debate whether these apparent striking benefits of ablation are believable.

CASTLE-AF (Catheter Ablation vs. Standard Conventional Treatment in Patients With Left Ventricular Dysfunction and Atrial Fibrillation: NCT00643188) (8) is another trial of catheter ablation versus medical therapy in 363 patients who had both atrial fibrillation and heart failure (see **Figure 1**). Over a median 3.15 years follow-up, treatments were compared using a modified ITT analysis; this excluded events during the run-in period and also nonfatal events during a 12-week "blinking period" after the baseline visit. The ablation group had fewer deaths (HR: 0.53; $p = 0.009$) and fewer primary endpoints, a composite of death and

heart failure hospitalization (HFH; HR: 0.62; $p = 0.009$). Again, the extent of treatment crossovers was substantial: 15.6% and 9.8%, respectively, in the ablation and medical therapy groups. But CASTLE-AF studied a different population (the subset of atrial fibrillation patients who also had heart failure), and therefore is not directly relevant to the question tackled by CABANA.

The dilemma we face is that although no study provides definitive evidence that catheter ablation is an effective intervention in patients with atrial fibrillation, does the totality of evidence across a range of somewhat imperfect studies carry sufficient conviction? Recent guidelines (9) are cautious, providing catheter ablation with their weakest possible recommendation. The hope is that ongoing and future trials will resolve the issue. But will they, too, suffer from high crossover rates?

ATTR-ACT trial

The ATTR-ACT (Safety and Efficacy of Tafamidis in Patients With Transthyretin Cardiomyopathy: NCT01994889) trial (10) is a double-blind trial that randomized 441 patients with transthyretin amyloid cardiomyopathy in a 2:1:2 ratio to 80 mg tafamidis, 20 mg tafamidis, or matching placebo for 30 months. In the primary analysis, they hierarchically assessed all-cause mortality, followed by frequency of CVH using Finkelstein-Schoenfeld (11) and win ratio (12) methods. Before we interpret this relatively novel methodology and its results, let us describe the essential results for the key components of the primary analysis (**Table 1**). Throughout, the results compare both tafamidis doses combined (264 patients) with placebo (177 patients).

For all-cause death over 30 months follow-up, the HR is 0.70 (95% confidence interval [CI]: 0.51 to 0.96; $p = 0.026$). For CVH, the investigators concentrate on comparing their overall frequency including repeat CVH within the same patient, with rates per year of 0.48 and 0.70 in the tafamidis and placebo groups, respectively. One needs to allow for the nonindependence of repeat events (i.e., a few patients may contribute a lot of events), so

negative binomial analysis is the most informative (13), with a rate ratio 0.71 (95% CI: 0.54 to 0.93; $p = 0.0068$). The alternative Poisson regression analysis, though pre-specified, falsely assumes independence of events, which makes the consequent CI narrower and the p value smaller than is appropriate. Had the investigators chosen to only analyze the time to first event (ignoring repeats) the results become less impressive: HR: 0.80 (95% CI: 0.62 to 1.03; $p = 0.07$).

The Kaplan-Meier plots of mortality and first CVH (not shown) both show the effect of treatment, with their respective plots showing a divergence after around 18 months and 9 months, respectively. Thus, an extended follow-up out to 30 months and inclusion of repeat CVHs in the primary analysis were crucial factors in demonstrating efficacy.

In addition, the tafamidis group showed highly significant steady improvements over time for both functional capacity (6-min walk test) and quality of life (Kansas City Cardiomyopathy Questionnaire overall score).

Now let us interpret the primary analysis that hierarchically assessed all-cause mortality, followed by frequency of CVH using the Finkelstein-Schoenfeld test (11) and the win ratio method (12). The win ratio is 1.70 (95% CI; 1.26 to 2.29; $p = 0.0006$). First, note that this p value is smaller than those separately obtained for mortality and CVHs. That is, combining the reductions in risk of death and CVH produces very strong evidence of a treatment effect and hence reinforces the value of this type of hierarchical analysis. But why not instead use a conventional composite endpoint of time to first CVH or death? Because this would have made inadequate use of the available data by: 1) ignoring repeat CVHs after the first one; and 2) ignoring any deaths that happen after a CVH.

The win ratio method is relatively new, so many people do not understand it. Hence, **Figure 2** is our attempt to explain how it works in ATTR-ACT. The win ratio analysis (14) is according to 2 stratification factors: transthyretin status (variant or wild type) and baseline New York Heart Association (NYHA) class (I/II or III). Thus, there are 4 patient strata:

variant type class I/II, variant type class III, wild type class I/II, and wild type class III comprising 58, 48, 242, and 93 patients, respectively. Then, within each stratum, every patient in the tafamidis group is compared with every patient in the placebo group: that is, there are $34 \times 24 + 29 \times 19 + 152 \times 90 + 49 \times 44 = 17,203$ pairs of patients. For each pair, we determine whether the tafamidis patient “won” or “lost” compared to the placebo patient. There are 2 ways to decide this: 1) who died first (they are the “loser”); and then, 2) if neither died, who had the most CVHs (again they are the “loser”), both being assessed over their shared follow-up time (**Figure 2**). Adding up across strata, we get a total of 8,595 winners and 5,071 losers. Hence the win ratio is $8,595/5,071 = 1.70$. The consequent 95% CI is 1.26 to 2.29 with $p = 0.0006$ using the Finkelstein-Schoenfeld test.

In conclusion, there is overwhelming evidence that patients on tafamidis had a better outcome than those on placebo. One outstanding issue is that the evidence is derived from 2 different daily doses, 80 mg and 20 mg, studied in a 2:1 randomization ratio. At present, no results are given on how the 2 doses compare. We are informed that such results will be available soon, which is crucial because they will have an important bearing on future use of tafamidis.

COAPT trial

The COAPT (Cardiovascular Outcomes Assessment of the MitraClip Percutaneous Therapy for Heart Failure Patients With Functional Mitral Regurgitation: NCT01626079) trial (15) randomized 614 patients with heart failure and moderate-to-severe mitral regurgitation to either mitral valve repair plus medical therapy (device group) or medical therapy alone (control group). Compliance was good: only 3% of the device group (N = 302 patients) had no procedure attempted and 1% of the control group (N = 312 patients) underwent transcatheter mitral valve repair. The primary effectiveness endpoint was all HFHs over 24 months follow-up, and all-cause death was a key secondary endpoint (**Figure 3**).

The COAPT trial is one of the first published trials in cardiology to use a repeat events analysis for its primary endpoint: that is all HFHs (including repeats) contribute to the treatment comparison, not just the conventional time to first event (HFH or CV death). This approach has certain advantages: the total burden of relevant disease is better captured and also the statistical power of the trial is likely to be enhanced (16). However, the consequent statistical techniques are more complex and as yet there is no overall expert consensus regarding a universally accepted optimal method.

But first let us look at the descriptive statistics. In the device and control groups, respectively, 92 and 151 patients had at least 1 HFH and the total number of HFHs (including repeats) were 160 and 283, respectively. These cumulative numbers of HFHs over time are shown in **Figure 3A**. The numbers of deaths over 2 years follow-up were 80 and 121 in the device and control groups, respectively (**Figure 3B**). These striking numerical differences require formal statistical analysis. For mortality that is straightforward, with a HR of 0.62 (95% CI: 0.46 to 0.82; log-rank $p < 0.001$). For the primary endpoint, all HFHs, the issue is made more complicated by the fact that death is an informative censoring or competing risk. That is, patients who are more prone to have 1 or more HFHs are also at a higher risk of dying: the jargon term for this is *joint frailty*. Another issue is that some patients have multiple HFHs, and this lack of independence of such repeat events needs correcting for in the analysis. The precise methodology required (17) is too technical to describe here, but the end result is a joint-frailty adjusted HR for all HFHs of 0.53 (95% CI: 0.40 to 0.70; $p < 0.001$).

As a simpler plausibility check on such “black box” complexity, one can convert the total HFHs in the device and control groups, 160 and 283, respectively, into annualized rates of 35.8 and 67.9 per 100 patient-years. The consequent rate ratio is 0.53, a reassuringly identical result. It is useful to convert such estimates of relative treatment effect into absolute risk reductions. Accordingly, the authors report the number needed to treat to prevent 1 HFH:

3.1 (95% CI: 1.9 to 7.9), and also to prevent a death: 5.9 (95% CI: 3.9 to 11.7)—both over 2 years of follow-up.

The MITRA-FR (Multicentre Study of Percutaneous Mitral Valve Repair MitraClip Device in Patients With Severe Secondary Mitral Regurgitation: NCT01920698) trial (18) also compared mitral valve repair with medical therapy alone and came to very different conclusions. They randomized patients with severe secondary regurgitation to undergo percutaneous mitral valve repair plus medical therapy (device group, N = 152) or to receive medical therapy alone (control group, N = 152). The primary efficacy endpoint was a composite of death or unplanned HFH, (i.e., a conventional time-to-first-event analysis and follow-up is 12 months). The key results for both this primary efficacy endpoint and for all-cause death are shown in **Figure 4**, with the corresponding results for COAPT shown alongside.

MITRA-FR showed no significant difference between the device and control groups for the primary endpoint and for death ($p = 0.53$ and 0.66 , respectively). Furthermore, the estimated HRs are numerically in favor of control (HR: 1.16 and 1.11, respectively). This is in sharp contrast to COAPT's highly significant findings in favor of the device (HR: 0.57 and 0.62, respectively; both $p < 0.001$). Interaction tests for between-trial heterogeneity of treatment effect for the primary endpoint and for all-cause death are both statistically significant ($p = 0.007$ and 0.039 , respectively).

There is no obvious single answer as to why the findings of these 2 trials differ so markedly (19,20). Points to consider are:

- 1) Follow-up was only 1 year in MITRA-FR but 2 years in COAPT. COAPT showed a marked difference in HFH incidence at 1 year, whereas its difference in mortality at 1 year, 19.3% versus 24.7%, is less striking, with a greater divergence in mortality out to 2 years, 29.1% versus 46.1%. Perhaps once MITRA-FR also achieves 2 years follow-up, the discrepancy between the trials may become less marked.

- 2) Did the trials differ in their use of medical therapy in both arms, both before patient entry and during follow-up?
- 3) Were there differences in procedural performance? In all such trials of novel devices, there is a learning curve among interventionists, and was the level of operator experience adequately achieved in both trials?
- 4) There are between-trial differences in patient eligibility criteria and also in the actual patterns of patient recruitment. Specifically, patients in MITRA-FR tended to have marked left ventricle dilation but a more modest degree of mitral regurgitation.

The evident superiority of outcomes in the device arm of the COAPT trial is impressive. But like all such strategy trials of device versus control, the proportion of patients with the relevant condition (heart failure and secondary mitral regurgitation) who actually got randomized is relatively low. Thus, the key question for future therapeutic practice is can we successfully define an identifiable patient population for whom the use of transcatheter mitral valve repair is an optimal strategy?

DECLARE trial

The DECLARE-TIMI58 (Multicenter Trial to Evaluate the Effect of Dapagliflozin on the Incidence of Cardiovascular Events: NCT01730534) trial (21) randomized 17,160 patients with type 2 diabetes to dapagliflozin or placebo. Of these patients, 41% had established atherosclerotic CV disease, with the remainder having multiple risk factors. The coprimary endpoints were major adverse cardiovascular events (MACE: CV death, myocardial infarction [MI], stroke) and the composite of CV death or HFH. The pre-specified hierarchy of secondary endpoints was a renal composite and then all-cause death. MACE was also the primary safety outcome to comply with the Food and Drug Administration (FDA) requirement to undertake noninferiority testing of all new antidiabetic agents (22).

Table 2 shows results for the primary, secondary, and other key outcomes over a median 4.2 years follow-up. In order to preserve type I error and to accommodate potential interim

stopping of the trial, it was pre-declared that each coprimary endpoint had to achieve $p < 0.0231$ to claim statistical significance. Only if both passed this hurdle was the sequence of 2 secondary endpoints allowed to enter formal significance testing.

In fact, of the 2 coprimaries, CV death and HFH revealed strong evidence of efficacy (HR: 0.83; 95% CI: 0.73 to 0.95; $p = 0.005$), whereas MACE did not ($p = 0.17$). Thus, no formal statistical testing was then performed on secondary endpoints. But the renal composite (the first secondary endpoint) had HR: 0.76 (95% CI: 0.67 to 0.87; $p = 0.00004$), an impressive result that is declared an exploratory finding, which formally does not constitute a claim of efficacy for renal benefits of dapagliflozin.

We think this interpretation lacks common sense, and it could characterize a more general failing in the way secondary endpoints are handled in major trials.

Here are a few suggestions:

- 1) If a secondary endpoint exhibits overwhelming evidence of treatment superiority (as in this case), then this needs to be cautiously accepted in a trial's conclusions. The FDA approval of empagliflozin for reducing CV death in the EMPA-REG OUTCOME (BI 10773 [Empagliflozin] Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients: NCT01131676) trial (23) is an example.
- 2) The formal use of p values to aid decision-making is most important to regulators (e.g., FDA and European Medicines Agency) whose responsibility is to approve a drug (or not) for a specific indication. The medical journals have a different responsibility, the presentation of scientific evidence. Although type I error control is an important principle, it should not become an obsession suppressing important secondary findings.
- 3) We do not agree with the current *New England Journal of Medicine* practice of not presenting p values for secondary endpoints if the primary endpoint(s) are not statistically significant. For instance, in **Table 1**, only the first 2 p values are reported in the paper by Maurer et al. (10). This feeds into the misinterpretation of $p < 0.05$ as an “accept or

reject” negative or positive philosophy of trial statistics, an issue one of us has critically appraised in 2 *New England Journal of Medicine* papers (24,25).

- 4) Another widespread but suspect practice is to pre-define secondary endpoints in a pre-defined hierarchy of statistical testing. Thus, trialists struggle as to what order to choose for their secondary endpoints, in effect having to “gaze into the crystal ball” as to which outcomes have the best chance of success while also bearing in mind their relative clinical importance.
- 5) We recommend that this practice should stop. Instead we propose that in future trials we consider all pre-defined secondary endpoints on an equal footing while implementing a statistical correction for multiple testing (e.g., Bonferroni or Hochberg procedure). Thus, if one has 5 secondary endpoints and the primary endpoint is formally significant, then if one of them achieves $p < 0.01$, it should also be considered as a formally significant finding.

We now turn our attention to how the DECLARE trial findings should be integrated into the overall evidence regarding the CV benefits of sodium-glucose cotransporter-2 inhibitors (SGLT2i). One way of accomplishing this is in a meta-analysis (26) combining findings from EMPA-REG (27), CANVAS (CANagliflozin cardioVascular Assessment Study: NCT01032629) (28), and DECLARE. We summarize and also extend their findings in

Figure 5.

For MACE (CV death, MI, and stroke) (**Figure 5A**), we see a consistency of benefit across trials in patients with pre-existing atherosclerotic disease: the overall HR is 0.86 (95% CI: 0.80 to 0.93; $p = 0.0002$). Even though this fell short of conventional significance in DECLARE, there is no evidence of heterogeneity across trials (interaction $p = 0.63$). However, in the 2 trials that did include patients with multiple risk factors only (CANVAS and DECLARE), there is no such signal of a benefit on MACE for such patients: the overall HR is 1.00 (95% CI: 0.87 to 1.16; $p = 0.98$). Note that the heterogeneity test comparing these

2 overall HRs is of marginal significance (interaction $p = 0.07$), so we need to exercise some restraint before presuming the whole MACE benefit story is solely confined to patients with pre-existing disease.

For HFH and CV death, we see strong evidence of treatment efficacy across trials (**Figure 5B**), which applies equally well in patients with and without a history of heart failure: the HRs are 0.71 and 0.79, respectively, both $p < 0.0001$. This is the most consistently positive finding from these SGLT2i trials. Whether these findings for patients with heart failure also extrapolate to such patients without diabetes is currently being investigated in the EMPEROR-HF (EMPagliflozin outcome tRial in Patients With chrOnic heaRt Failure: NCT03057951, NCT03057977) and Dapa-HF (Study to Evaluate the Effect of Dapagliflozin on the Incidence of Worsening Heart Failure or Cardiovascular Death in Patients With Chronic Heart Failure: NCT03036124) trials.

For CV death and all-cause death (**Figure 5C and 5D**), the striking reductions due to empagliflozin in the EMPA-REG trial have not been found in CANVAS and DECLARE. The tests for heterogeneity of effects across trials are significant (interaction $p = 0.007$ and $p = 0.018$, respectively, for CV death and all-cause death). As far as we understand, there is no obvious biological explanation for these differing effects on mortality. Faced with no back-up evidence, one is entitled to speculate that the surprisingly large magnitudes of observed effects with empagliflozin (HR: 0.62 and 0.68 for CV death and all-cause death, respectively) may be by chance an exaggeration of the truth, but it is hard to dispute the effects exists given their extreme significance (both $p < 0.0001$).

REDUCE-IT trial

The REDUCE-IT (Evaluation of the Effect of AMR101 on Cardiovascular Health and Mortality in Hypertriglyceridemic Patients With Cardiovascular Disease or at High Risk for Cardiovascular Disease: NCT01492361) trial (29) evaluated whether 2 g twice daily of icosapent ethyl, a purified ethyl ester of eicosapentaic acid (EPA) could reduce the risk of CV

events. A total of 8,179 patients, 71% with established CV disease and 29% with diabetes and other risk factors, who had been on a statin were randomized to EPA or placebo and followed for a median 4.9 years. The primary composite endpoint comprised CV death, MI, stroke, coronary revascularization, or unstable angina.

Table 3 summarizes the key trial findings. There was a markedly reduced incidence of the primary endpoint (HR: 0.75; 95% CI: 0.68 to 0.83; $p < 0.0001$). Every component of the primary endpoint showed evidence of a significant effect, this being most marked for MI, coronary revascularization, and unstable angina. There were significantly fewer CV deaths on EPA ($p = 0.03$), but because non-CV deaths were almost identical in the 2 groups, the finding is more tenuous for all-cause death ($p = 0.09$). It is also worth noting that from Kaplan-Meier plots for the primary endpoint (and also for the key secondary endpoints of CV death, MI, and stroke), there is no evidence of a treatment effect in the first year of follow-up.

These positive results in REDUCE-IT need to be set against the previous disappointing findings from other trials of omega-3 fatty acids. For instance, a meta-analysis of 10 such trials in 77,917 patients (30) gave for any major vascular event a HR of 0.97 (95% CI: 0.95 to 1.01).

One statistical option when making sense of one positive trial in an otherwise negative field is to undertake Bayesian analyses. Thus, one expresses for the trial's primary endpoint a plausible distribution of prior belief for the true HR and then uses Bayes' theorem to combine such a prior belief distribution with evidence from the trial to produce a posterior belief distribution (31). Because no one prior suits everyone, we present 3 scenarios:

1. **The pessimist's prior.** The pessimist does not perceive the specific omega-3 fatty acid used in REDUCE-IT as being any different from the rest. Hence, they use the meta-analysis estimate as the prior distribution and integrate it with the new data, as shown in **Figure 6A**. Because the prior data and pessimist's belief distribution are so strongly negative, the evidence from REDUCE-IT only slightly shifts the belief, with

posterior belief having a median HR of 0.95 with a 95% credibility interval from 0.92 to 0.98. But one should note that the prior meta-analysis and REDUCE-IT are in sharp disagreement: an interaction test comparing their respective HRs, 0.97 and 0.75, has $p < 0.001$. Thus, one might argue that the pessimist's prior reflects an overwhelming prejudice against all omega-3 fatty acids, which is clearly overruled by REDUCE-IT.

2. **The realist's prior.** The realist accepts the principle that the EPA in REDUCE-IT may be onto something better than previously studied omega-3 fatty acids, but believes equally that it might conceivably do harm. Thus, he chooses a prior distribution centered on no true effect (HR: 1) with a wider spread of uncertainty (**Figure 6B**). To be precise, we use a normal distribution for the prior of the log (HR), which has a positive tail out to 20% reduction in hazard with prior probability 0.025. In this scenario, the realist's caution leads to a posterior belief having median HR of 0.79 with a 95% credibility interval from 0.72 to 0.86. That is, the realist now accepts the positive finding in REDUCE-IT but is inclined to believe that the true effect of EPA is somewhat less than that observed. Note that if REDUCE-IT had been a much smaller trial, then the interpretation would be more cautious. For instance, if REDUCE-IT had been only one-quarter as big with the same observed HR of 0.75, then the posterior median belief would have been 0.85 with a 95% credibility interval of 0.75 to 0.99. That is, one needs a very large trial to convince people when previous evidence has been so negative.
3. **The optimist's prior.** The optimist asserts that the EPA and its high dose used in REDUCE-IT bears no relation to previous trials of other omega-3 fatty acids. One therefore has no prior insight into what the treatment effect might truly be. In such circumstances, one chooses a vague prior, in which case the evidence is its own posterior distribution automatically. Such optimism is implicit in any conventional (frequentist) presentation of results. That is, the published estimate, 95% CI, and p

value are compatible with a vague prior. This optimistic stance is encouraged by a Japanese trial of 1.8 g daily of icosapent ethyl (32), which showed a significant 19% reduction in coronary events, though that trial was not blinded.

By its very nature, the use of Bayesian techniques to see whether a trial's evidence is strong enough to overcome prior skepticism has no single right answer. But Bayesian methods are a quantitative tool that may help clarify the arguments.

Another way of interpreting such a highly significant result in a unique trial is to argue how many replicate trials would have to achieve one-sided $p < 0.025$ to have the same strength of evidence as the observed significance for the primary endpoint, which for REDUCE-IT is $p = 0.00000001$. The answer is it would take 4 replicate trials, each with one-sided $p = 0.025$, to approach the same level of proof beyond reasonable doubt that efficacy is established.

Further insight into the efficacy of EPA is documented by the large reduction in total ischemic events in the REDUCE-IT trial (33). In addition to the 1,606 patients having a first primary event, there were a further 1,303 repeat events. There is an estimated 30% risk reduction in this total burden of ischemic events (95% CI: 22% to 38% reduction).

But we do need to recognize the adverse event profile of EPA in the REDUCE-IT trial (**Table 3**). There are statistically significant increases in incidence of constipation (+1.8%), peripheral edema (+1.5%), and atrial fibrillation (+1.4%), the latter being of particular concern. These absolute risks need setting against the observed absolute benefits, which are -4.8% for the primary endpoint and -3.6% for the composite of CV death, MI, and stroke. Furthermore, the increase in peripheral edema is not accompanied by an increase in heart failure, and the increase in atrial fibrillation is offset by a significantly lower incidence of stroke.

The European Medicines Agency has recently reviewed the totality of evidence for omega-3 fatty acid (34) and concluded, "the benefit-risk balance of omega-3 ethyl esters in

secondary prevention after myocardial infarction is not favorable.” The key question is, does the REDUCE-IT trial genuinely show that a high dose of icosapent ethyl has a much more favorable benefit-risk profile?

AUGUSTUS trial

The AUGUSTUS trial (An Open-label, 2 x 2 Factorial, Randomized Controlled, Clinical Trial to Evaluate the Safety of Apixaban vs. Vitamin K Antagonist and Aspirin vs. Aspirin Placebo in Patients With Atrial Fibrillation and Acute Coronary Syndrome or Percutaneous Coronary Intervention: NCT02415400) (35) was an open-label 2 × 2 factorial trial to evaluate the safety of apixiban versus a vitamin K agonist (VKA) and aspirin versus placebo in 4,614 patients with atrial fibrillation and acute coronary syndrome and/or percutaneous coronary intervention (PCI). All patients also received a P2Y₁₂ inhibitor for 6 months, and also aspirin on the actual day of acute coronary syndrome and/or PCI. The primary outcome through 6 months was major bleed or clinically relevant nonmajor (CRNM) bleeding using the International Society on Thrombosis and Haemostasis definitions. Secondary outcomes included major bleeding, death, rehospitalization, and ischemic events (i.e., MI, stroke, stent thrombosis, urgent revascularization).

The nature of such a factorial trial (36) is that each patient is simultaneously randomized twice: to apixiban versus VKA and to aspirin versus placebo. Thus, there are 4 treatment groups: apixiban + aspirin; apixiban + placebo; VKA + aspirin; VKA + placebo. However, the pre-defined analysis plan is to concentrate on results for each of the 2 treatment factors (apixiban vs. VKA and then aspirin vs. placebo) in all 4,614 randomized patients. This is accompanied by an interaction test examining whether any difference between apixiban and VKA depends on whether the patients received aspirin or placebo. Comparison of the 4 treatment groups is considered a secondary descriptive data display without formal statistical analysis.

Results for the primary and key secondary endpoints are shown in **Table 4**. For the primary endpoint, major and CRNM bleeding, there is strong evidence that apixiban is superior to VKA (HR: 0.69; 95% CI: 0.58 to 0.81; $p < 0.0001$). Also, aspirin is clearly inferior to placebo (HR: 1.90; 95% CI: 1.60 to 2.25; $p < 0.0001$). If one confines attention to major bleeds only, the same conclusions are evident.

Results for hospitalization over 6 months reveal a significant reduction on apixiban compared with VKA ($p = 0.002$). There were also slightly more hospitalizations on aspirin compared with placebo ($p = 0.12$). There was an overall 3.3% mortality over 6 months with no evidence of treatment differences (data not shown). The composite of death and ischemic events had an overall 6.35% incidence over 6 months, again with no evidence of treatment differences (**Table 4**). For all these outcomes, the statistical tests of interaction were not statistically significant.

We now display in **Figure 7** the same outcomes broken down by the 4 randomized treatment groups: apixiban + aspirin; apixiban + placebo; VKA + aspirin; VKA + placebo. This provides reassuring evidence that the lowest incidence of both bleeding and hospitalizations is for patients on apixiban + placebo. A legitimate concern is that giving apixiban without aspirin might have increased the risk of ischemic events, but in fact, the incidence of the composite of death and ischemic events for apixiban + placebo is comparable to that for apixiban + aspirin.

Given that the primary endpoint superiorities of apixiban over VKA and placebo over aspirin were very highly significant, one might question whether the trial should have been stopped earlier (37). One of us (S.J.P.) was on the trial's Data Monitoring Committee that decided not to recommend early stopping. Like all trials of antithrombotic drugs, there is a tradeoff between treatment efficacy (prevention of ischemic events) and safety (an increased bleeding risk). Although AUGUSTUS was well powered to reach definitive findings on the

latter, the incidence of ischemic events was substantially lower. Therefore, early stopping of the trial would have inhibited the ability to reach clear conclusions on treatment efficacy.

A network meta-analysis (38) has integrated the evidence from AUGUSTUS with 3 other relevant trials. They concluded that a strategy based on non-VKA oral anticoagulants (such as apixiban) plus a P2Y12-inhibitor, but without aspirin, is the preferred regimen post-PCI for patients who also have atrial fibrillation.

Conclusion

Each of the 6 clinical trials we have critiqued has important implications for future patient care. We have aimed to provide a balanced account of their key findings, how they fit into the context of previous knowledge, and each study's strengths and weaknesses.

Along the way, we have highlighted some methodological advances and challenges pertinent to trials research in general. These include:

1. The problems of treatment crossovers and the cautious use of alternatives to ITT analysis.
2. When to do a pivotal randomized trial of a new procedure requiring investigator skills.
3. The use of new hierarchical methods (e.g., win ratio) for combining evidence from fatal and nonfatal events.
4. The use of repeat events (not just time-to-first event) when comparing treatments in a chronic disease such as heart failure.
5. Allowance for the competing risk of death when reporting the treatment effect on nonfatal events such as hospitalizations by use of joint-frailty models.
6. How to interpret the overall evidence, when 2 related trials reach contrasting conclusions.
7. The need for a better understanding of secondary endpoint findings, perhaps abandoning the current practice of hierarchical testing.

8. How to place a new trial in the context of prior related trials and appropriate use of meta-analyses.
9. How to interpret a very positive trial in an otherwise negative field and the potential value of Bayesian methods.
10. How to present and interpret the findings of a 2×2 factorial trial.

Overall, we appreciate this opportunity to review some key recent randomized trials, and hope our insights are of value, for both their clinical relevance and their implications for future clinical trials research and its methodology.

Bullet points

- This review is a constructive critical appraisal of key recent cardiology trial reports
- The trials are ATTR-ACT, CABANA, COAPT, DECLARE, REDUCE-IT, and AUGUSTUS
- For each study, we document and interpret the main findings
- We pay particular attention to new findings, their research context and study limitations
- Topical examples provide methodological insights pertinent to future clinical trials research.

References

1. Pocock SJ, Collier TJ. Critical appraisal of the 2018 ACC Scientific Sessions late-breaking trials from a statistician's perspective. *J Am Coll Cardiol* 2018;71:2957–69.
2. Packer DL, Mark DB, Robb RA, et al.; CABANA Investigators. Effect of catheter ablation vs antiarrhythmic drug therapy on mortality, stroke, bleeding, and cardiac arrest among patients with atrial fibrillation: the CABANA Randomized Clinical Trial. *JAMA* 2019 March 15 [E-pub ahead of print], <http://dx.doi.org/10.1001/jama.2019.0693>.
3. Barkun JS, Aronson JK, Feldman LS, et al. Evaluation and stages of surgical innovations. *Lancet* 2009;374:1089–96.
4. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *N Engl J Med* 2017;377:1391–8.
5. Hernán MA, Scharfstein D. Cautions as regulators move to end exclusive reliance on intention to treat. *Ann Intern Med* 2018;168:515–6.
6. Leuchs AK, Zinserling J, Brandt A, Wirtz D, Benda N. Choosing appropriate estimands in clinical trials. *Ther Innov Regul Sci* 2015;49:584–92.
7. Noseworthy PA, Gersh BJ, Kent DM, et al. Atrial fibrillation ablation in practice: assessing CABANA generalizability. *Eur Heart J* 2019 Mar 15 [E-pub ahead of print]; <http://dx.doi.org/10.1093/eurheartj/ehz085>.
8. Marrouche NF, Brachmann J, Andresen D, et al. Catheter ablation for atrial fibrillation with heart failure. *N Engl J Med* 2018;378:417–27.
9. January CT, Wann LS, Calkins H, et al. 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. *J Am Coll Cardiol* 2019 Jan 21 [E-pub ahead of print], <http://dx.doi.org/10.1016/j.jacc.2019.01.011>.
10. Maurer MS, Schwartz JH, Gundapaneni B, et al; ATTR-ACT Study Investigators. Tafamidis treatment for patients with transthyretin amyloid cardiomyopathy. *N Engl J Med* 2018;379:1007–16.
11. Finkelstein DM, Schoenfeld DA. Combining mortality and longitudinal measures in clinical trials. *Stat Med* 1999;18:1341–54.

12. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J* 2012;33:176–82.
13. Keene ON, Jones MR, Lane PW, Anderson J. Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. *Pharm Stat* 2007;6:89–97.
14. Dong G, Qiu J, Wang D, Vandemeulebroecke M. The stratified win ratio. *J Biopharm Stat* 2018;28:778–96.
15. Stone GW, Lindenfeld J, Abraham WT, et al.; COAPT Investigators. Transcatheter mitral-valve repair in patients with heart failure. *N Engl J Med* 2018;379:2307–18.
16. Claggett B, Pocock S, Wei LJ, Pfeffer MA, McMurray JJV, Solomon SD. Comparison of time-to-first event and recurrent-event methods in randomized clinical trials. *Circulation* 2018;138:570–7.
17. Rogers JK, Yaroshinsky A, Pocock SJ, Stokar D, Pogoda J. Analysis of recurrent events with an associated informative dropout time: application of the joint frailty model. *Stat Med* 2016;35:2195–205.
18. Obadia JF, Messika-Zeitoun D, Leurent G, et al.; MITRA-FR Investigators. Percutaneous repair or medical treatment for secondary mitral regurgitation. *N Engl J Med* 2018;379:2297–306.
19. Nishimura RA, Bonow RO. Percutaneous repair of secondary mitral regurgitation - a tale of two trials. *N Engl J Med* 2018;379:2374–6.
20. Packer M. When cardiovascular trials collide. *Eur Heart J* 2019;40:501–4.
21. Wiviott SD, Raz I, Bonaca MP, et al.; DECLARE-TIMI58 Investigators. Dapagliflozin and cardiovascular outcomes in type 2 diabetes. *N Engl J Med* 2019;380:347–57.
22. Food and Drug Administration, Center for Drug Evaluation and Research (CDER) Guidance for Industry: Diabetes Mellitus—Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes. 2008. Available at: <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071627.pdf>. Accessed April 2, 2019.
23. Food and Drug Administration, Center for Drug Evaluation and Research. Summary Minutes of the Endocrinologic and Metabolic Drugs Advisory Committee Meeting June 28, 2016. 2016. Available at: <https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/>

[Drugs/EndocrinologicandMetabolicDrugsAdvisoryCommittee/UCM518911.pdf](#).

Accessed April 2, 2019.

24. Pocock SJ, Stone GW. The primary outcome is positive - is that good enough? *N Engl J Med* 2016;375:971–9.
25. Pocock SJ, Stone GW. The primary outcome fails - what next? *N Engl J Med* 2016;375:861–70.
26. Zelniker TA, Wiviott SD, Raz I, et al. SGLT2 inhibitors for primary and secondary prevention of cardiovascular and renal outcomes in type 2 diabetes: a systematic review and meta-analysis of cardiovascular outcome trials. *Lancet* 2019;393:31–9.
27. Zinman B, Lachin JM, Inzucchi SE. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med* 2016;374:1094.
28. Neal B, Perkovic V, Mahaffey KW, et al.; CANVAS Program Collaborative Group. Canagliflozin and cardiovascular and renal events in type 2 diabetes. *N Engl J Med* 2017;377:644–57.
29. Bhatt DL, Steg PG, Miller M, et al.; REDUCE-IT Investigators. Cardiovascular risk reduction with icosapent ethyl for hypertriglyceridemia. *N Engl J Med* 2019;380:11–22.
30. Aung T, Halsey J, Kromhout D, et al.; Omega-3 Treatment Trialists' Collaboration. Associations of omega-3 fatty acid supplement use with cardiovascular disease risks: meta-analysis of 10 trials involving 77 917 individuals. *JAMA Cardiol* 2018;3:225–34.
31. Spiegelhalter DJ, Freedman LS, Mahesh KBP. Bayesian approaches to randomized trials. *J R Stat Soc Ser A Stat Soc* 1994;157:357–87.
32. Yokoyama M, Origasa H, Matsuzaki M, et al.; Japan EPA lipid intervention study (JELIS) Investigators. Effects of eicosapentaenoic acid on major coronary events in hypercholesterolaemic patients (JELIS): a randomised open-label, blinded endpoint analysis. *Lancet* 2007;369:1090–8.
33. Bhatt DL, Steg G, Miller M, et al. Effects of icosapent ethyl of total ischemic events: from REDUCE-IT. *J Am Coll Cardiol* 2019 Mar 18 [E-pub ahead of print], <http://dx.doi.org/10.1016/j.jacc.2019.02.032>.
34. European Medicines Agency. Omega-3 fatty acid medicines no longer considered effective in preventing heart disease. London, UK: European Medicines Agency, 2018. Available at: <https://www.ema.europa.eu/en/documents/referral/omega-3-fatty->

acid-medicines-omega-3-fatty-acid-medicines-no-longer-considered-effective-preventing_en.pdf. Accessed April 2, 2019.

35. Lopes RD, Heizer G, Aronson R, et al.; AUGUSTUS Investigators. Antithrombotic therapy after acute coronary syndrome or pci in atrial fibrillation. *N Engl J Med* 2019 Mar 17 [E-pub ahead of print], <http://dx.doi.org/10.1056/NEJMoa1817083>.
36. Montgomery AA, Peters TJ, Little P. Design, analysis and presentation of factorial randomised controlled trials. *BMC Med Res Methodol* 2003;3:26.
37. Pocock SJ. Current controversies in data monitoring for clinical trials. *Clin Trials* 2006;3:513–21.
38. Lopes R. personal communication (reference to be inserted once in press).

Figure Legends

Central Illustration: A Constructive Critical Appraisal of 6 Key Clinical Trials in

Cardiology First Presented During May 2018 to March 2019. In time order of first presentation they are CABANA, ATTR-ACT, COAPT, DECLARE, REDUCE-IT, and AUGUSTUS. For each study, our aim is to document and interpret the main findings, paying attention to new findings, their research context, and study limitations. These topical examples also provide methodological insights pertinent to future clinical trials research.

ACC = American College of Cardiology; ACS = acute coronary syndrome; AHA = American Heart Association; CV = cardiovascular; CVH = cardiovascular hospitalization; ESC = European Society for Cardiology; HFH = heart failure hospitalization; PCI = percutaneous coronary intervention; TCT = Transcatheter Cardiovascular Therapeutics; VKA = vitamin K agonist.

Figure 1: Key Findings From the CABANA Trial, and Other Studies Comparing

Catheter Ablation and Drug Therapy in Patients With Atrial Fibrillation. *The primary

endpoint for CABANA (and OPTUM) is death, disabling stroke, serious bleeding, or cardiac arrest. + In the CABANA trial, the rate of crossovers in the ablation and drug therapy arms were 9.2% and 27.5%, respectively. In the CASTLE-AF trial, the corresponding crossover rates were 15.6% and 9.8%, respectively. Key findings from the CABANA trial (plus the OPTUM database and CASTLE-AF trial) comparing catheter ablation and drug therapy in patients with atrial fibrillation. ITT analysis did not show a significant treatment difference for either the primary endpoint or all-cause death, though there were numerically fewer primary endpoints and deaths in the ablation arm. For the composite of death and CVH, ITT analysis did show a significant reduction in the ablation arm ($p = 0.001$). Analysis by treatment received shows highly significant differences in favor of ablation for the primary endpoint (HR: 0.67; $p = 0.006$) and all-cause death (HR: 0.60; $p = 0.005$). Per protocol analysis showed a borderline significant difference for the primary endpoint. CI = confidence

interval; CV = cardiovascular; CVH = cardiovascular hospitalization; HR = hazard ratio; ITT = intention to treat; NR = not reported.

Figure 2: Win Ratio Analysis of the Primary Hierarchical Endpoint of All-Cause Death and Frequency of CVH in the ATTR-ACT trial. Interpretation of win ratio = 1.7: if you compare any 2 patients in the same stratum, one on tafamidis and one on placebo, and they are not a tie (both alive with same number of CVHs) then the odds that the patient on tafamidis wins is 1.7. For people who prefer probabilities, then the probability that the patient on tafamidis wins is $1.7/(1.7 + 1) = 0.63$. Win ratio analysis of the primary hierarchical endpoint of: **(A)** all-cause death; and **(B)** frequency of CVH in the ATTR-ACT trial. Patients are in 4 strata based on transthyretin status (variant or wild type) and NYHA class (I/II or III). Within each stratum, every patient in the tafamidis group is compared with every patient in the placebo group. For each pair it is determined over their shared follow-up time, first whether the patient on tafamidis “wins” or “loses” on death and then, if “tied” on death, whether the patient on tafamidis “wins” or “loses” on the number of CVH. There were 8,595 winners and 5,071 losers, giving a win ratio of 1.70 (95% CI: 1.26 to 2.29; $p = 0.0006$). CVH = cardiovascular hospitalization; CI = confidence interval; NYHA = New York Heart Association.

Figure 3: Cumulative Incidence of HFHs (Including Repeats) and Death From Any Cause in the COAPT Trial. Cumulative incidence of **(A)** the primary effectiveness endpoint (HFHs, including repeats over 24 months) and **(B)** death from any cause in the COAPT trial. The COAPT trial randomized 614 patients with heart failure and moderate-to-severe mitral regurgitation to either mitral valve repair plus medical therapy (device group) or medical therapy alone (control group). The joint-frailty adjusted HR for the primary effectiveness endpoint is 0.53 (95% CI: 0.40 to 0.70; $p < 0.001$). The HR for all-cause mortality is 0.62 (95% CI: 0.46 to 0.82; $p < 0.001$). Reprinted, with permission, from Stone et al. (15). HFH = heart failure hospitalization. Other abbreviations as in **Figure 1**.

Figure 4: Comparison of the COAPT and MITRA-FR Trials for the Composite of Death and HFH and All-Cause Death at 1 Year. Both trials randomized patients to undergo percutaneous mitral valve repair plus medical therapy (device group) or to medical therapy alone (control group). MITRA-FR showed no significant difference between the device and control groups for the primary endpoint and for death ($p = 0.53$ and 0.66 , respectively), and furthermore, the estimated HRs are numerically in favor of control (HR: 1.16 and 1.11, respectively). This is in sharp contrast to COAPT's highly significant findings in favor of the device (HR: 0.57 and 0.62, respectively, both $p < 0.001$). Interaction tests for between-trial heterogeneity of treatment effect for the primary endpoint and for all-cause death are both statistically significant ($p = 0.007$ and 0.039 , respectively). Abbreviations as in **Figures 1 and 3**.

Figure 5: Fixed-Effect Meta-Analysis of 3 SGLT2i Trials for MACE, HFH, CV Death, and All-Cause Death. Fixed effect meta-analysis of EMPA-REG OUTCOME, CANVAS Program and DECLARE SGLT2i trials for (A) MACE, (B) HFH, (C) CV death, and (D) all-cause death. For MACE, we see a consistent benefit across trials for patients with pre-existing atherosclerotic disease (HR: 0.86; 95% CI 0.80 to 0.93; $p = 0.0002$) but no evidence of any benefit for patients with multiple risk factors only. For HFH and CV death, we see strong evidence of treatment efficacy across trials, which applies equally well in patients with and without a history of heart failure (HR: 0.71 and 0.79 respectively, both $p < 0.0001$). For CV death and all-cause death, the striking reductions in the EMPA-REG trial have not been found in CANVAS and DECLARE. The tests for heterogeneity of effects across trials are significant (interaction $p = 0.007$ and $p = 0.018$, respectively, for CV death and all-cause death). MACE = major adverse cardiovascular events. Other abbreviations as in **Figures 1 and 3**.

Figure 6: Alternative Bayesian Analyses and Interpretations of the REDUCE-IT Trial Results for its Primary Endpoint. First (A), using a pessimist's prior based on a meta-

analysis of previous trials of omega-3 fatty acids, the posterior belief remains cautious (HR: 0.95; 95% CI: 0.92 to 0.98). Second **(B)**, using a realist's prior, which cautiously accepts that some modest benefit (or harm) of eicosapentaic acid is possible, the evidence from REDUCE-IT is very influential producing a posterior belief distribution centered on a 21% risk reduction (HR: 0.79; 95% CI: 0.72 to 0.86). Third, using an optimist's vague prior, implying previous studies are irrelevant, the posterior belief is the same as the REDUCE-IT evidence, as shown by the red distribution curves in **(A)** and **(B)** (HR: 0.75; 95% CI: 0.68 to 0.83). Abbreviations as in **Figure 1**.

Figure 7: Main Findings from the AUGUSTUS Trial for the 4 Randomized Treatment

Options. Main findings from the AUGUSTUS trial, an open-label 2×2 factorial trial evaluating the safety of: i) apixiban versus vitamin K agonist (VKA); and ii) aspirin versus placebo in patients with atrial fibrillation and acute coronary syndrome (ACS) and/or percutaneous coronary intervention. The percentage with an event is shown for the primary endpoint (major or CRNM bleed) and key secondary endpoints for each of the 4 randomized treatment combinations. The number of patients on each of the 4 treatment options was 1,145 apixiban + aspirin (green), 1,143 apixiban + placebo (gold), 1,123 VKA + aspirin (red), and 1,126 VKA + placebo (blue). This provides reassuring evidence that the lowest incidence of both bleeding and hospitalizations is for patients on apixiban + placebo. A legitimate concern is that giving apixiban without aspirin might have increased the risk of ischemic events, but in fact, the incidence of the composite of death and ischemic events for apixiban + placebo is comparable to that for apixiban + aspirin. *Ischemic events are: myocardial infarction, stroke, stent thrombosis, urgent revascularization.

Table 1 Findings of the ATTR-ACT Trial for All-Cause Mortality and CVH

| | Tafamidis (N = 264) | Placebo (N = 177) | HR (95% CI) (Risk Ratio for CVH rate) | p Value |
|------------------------------------|------------------------|----------------------|---|---------|
| Deaths* | 39 (14.8%) | 38 (21.5%) | 0.70 (0.51–0.96) | 0.026 |
| Number with ≥ 1 CVH | 138 (52.3%) | 107 (60.5%) | 0.80 (0.62–1.03) | 0.078 |
| Number of CVHs (including repeats) | 267 | 231 | | |
| CVH rate per year | 0.48 | 0.70 | 0.71 [†] (0.54–0.93) | 0.0068 |

* Heart transplant and implantation of a cardiac mechanical assist device were treated as deaths for this published analysis (and also for the win ratio analysis in **Figure 2**). If they were not treated as deaths, the HR for death becomes 0.67 with $p = 0.018$.

[†] This risk ratio is from a negative binomial model that allows for the nonindependence of repeat hospitalizations within a patient

CI = confidence interval; CVH = cardiovascular hospitalization; HR = hazard ratio.

Table 2: Key Efficacy Outcomes (and Their Components) in the DECLARE trial

| Outcome | Dapagliflozin | | Placebo | | HR (95% CI) | p Value |
|-----------------------|---------------|------------|-------------|------------|------------------|----------|
| | (N = 8,582) | | (N = 8,578) | | | |
| | rate/1,000 | | rate/1000 | | | |
| | n | patient-yr | n | patient-yr | | |
| CV death or HFH* | 417 | 12.2 | 496 | 14.7 | 0.83 (0.73-0.95) | 0.005 |
| MACE* | 756 | 22.6 | 803 | 24.2 | 0.93 (0.84-1.03) | 0.17 |
| Renal composite† | 370 | 10.8 | 480 | 14.1 | 0.76 (0.67-0.87) | 0.000038 |
| Death from any cause‡ | 529 | 15.1 | 570 | 16.4 | 0.93 (0.82-1.04) | 0.23 |
| HFH | 212 | 6.2 | 286 | 8.5 | 0.73 (0.61-0.88) | 0.00076 |
| Myocardial infarction | 393 | 11.7 | 441 | 13.2 | 0.89 (0.77-1.01) | 0.092 |
| Ischemic stroke | 235 | 6.9 | 231 | 6.8 | 1.01 (0.84-1.21) | 0.91 |
| CV death | 245 | 7.0 | 249 | 7.1 | 0.98 (0.82-1.17) | 0.82 |
| Non-CV death | 211 | 6.0 | 238 | 6.8 | 0.88 (0.73-1.06) | 0.18 |

| | | | | | | | |
|------------------------------------|-----|-----|-----|-----|-------------|--|-----------|
| ≥40% decrease in eGFR | | | | | | | |
| to <60 ml/min/1.73m ² , | 127 | 3.7 | 238 | 7.0 | 0.53 (0.43- | | <0.000001 |
| ESRD or renal death | | | | | 0.66) | | |

* Coprimary endpoints. † Renal composite (the first secondary endpoint) comprises ≥40% decrease in eGFR to <60 ml/min/1.73 m², ESRD, or death from renal or CV cause. ‡ The second secondary endpoint.

CV = cardiovascular; eGFR = estimated glomerular filtration rate; ESRD = end-stage renal disease; HFH = heart failure hospitalization; MACE = major adverse cardiovascular events. Other abbreviations as in **Table 1**.

Table 3: Key Results From the REDUCE-IT Trial: Primary Endpoint, its Components, Mortality, and Adverse Events

| Outcome | Icosapent | | HR (95% CI) | p Value |
|-------------------------------------|----------------------|------------------------|------------------|---------|
| | Ethyl (N = 4,089) | Placebo (N = 4,090) | | |
| Primary composite | 17.2% | 22.0% | 0.75 (0.68–0.83) | <0.001 |
| Myocardial infarction | 6.1% | 8.7% | 0.69 (0.58–0.81) | <0.001 |
| Urgent revascularization | 5.3% | 7.8% | 0.65 (0.55–0.78) | <0.001 |
| Unstable angina | 2.6% | 3.8% | 0.68 (0.53–0.87) | 0.002 |
| Stroke | 2.4% | 3.3% | 0.72 (0.55–0.93) | 0.01 |
| Cardiovascular death | 4.3% | 5.2% | 0.80 (0.66–0.98) | 0.03 |
| All-cause death | 6.7% | 7.6% | 0.87 (0.74–1.02) | 0.09 |
| Constipation* | 5.4% | 3.6% | | <0.001 |
| Peripheral edema* | 6.5% | 5.0% | | 0.002 |
| Atrial fibrillation* | 5.3% | 3.9% | | 0.003 |
| Hospitalized atrial fibrillation | 3.1% | 2.1% | | 0.004 |

* These adverse events were not adjudicated

Abbreviations as in **Table 1**.

Table 4 Results for the AUGUSTUS Trial’s Primary and Key Secondary Endpoints Comparing Apixiban Versus VKA and Aspirin Versus Placebo Over 6 Months Follow-Up

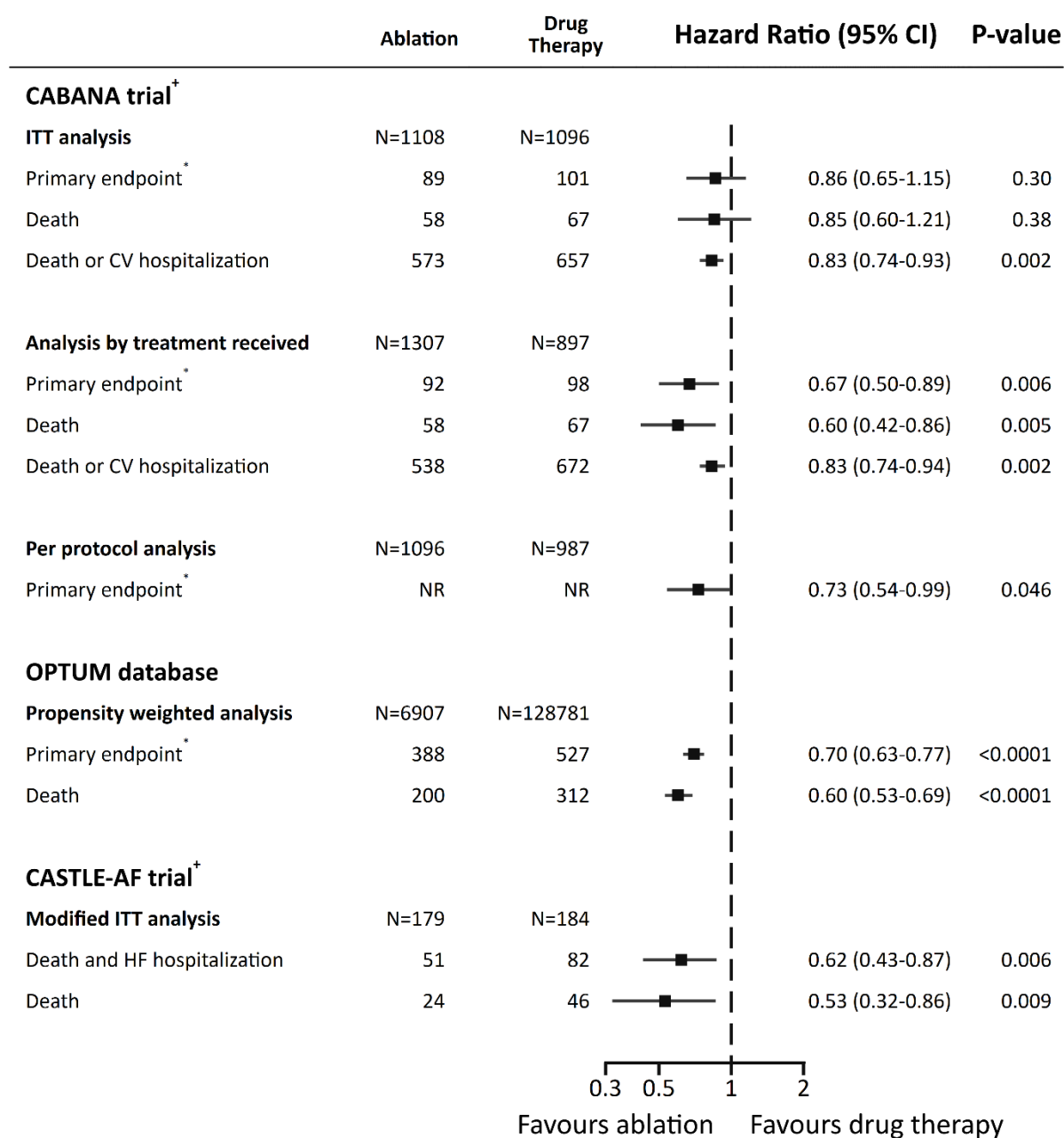
| | Apixiban (N = 2,306) | VKA (N = 2,308) | HR (95% CI) | p Value | Aspirin (N = 2,307) | Placebo (N = 2,307) | HR (95% CI) | p Value | Interaction p* |
|--|---------------------------------|----------------------------|--------------------|----------------|--------------------------------|--------------------------------|--------------------|----------------|---------------------------|
| Major + CRNM bleed (primary endpoint) | 10.5% | 14.7% | 0.69 (0.58–0.81) | <0.0001 | 16.1% | 9.0% | 1.90 (1.60–2.25) | <0.0001 | 0.716 |
| Major bleed | 3.0% | 4.6% | 0.64 (0.47–0.86) | 0.005 | 4.7% | 2.9% | 1.70 (1.25–2.32) | 0.001 | 0.197 |
| Hospitalization | 22.5% | 26.3% | 0.83 (0.74–0.93) | 0.002 | 25.4% | 23.4% | 1.10 (0.98–1.24) | 0.123 | 0.269 |
| Death or ischemic event [†] | 6.2% | 6.5% | 0.94 (0.75–1.18) | 0.592 | 5.9% | 6.8% | 0.88 (0.70–1.11) | 0.251 | 0.325 |

* The interaction test assesses whether the treatment difference between apixiban and VKA depends on whether the patient received aspirin or placebo.

† Ischemic events are: myocardial infarction, stroke, stent thrombosis, urgent revascularization.

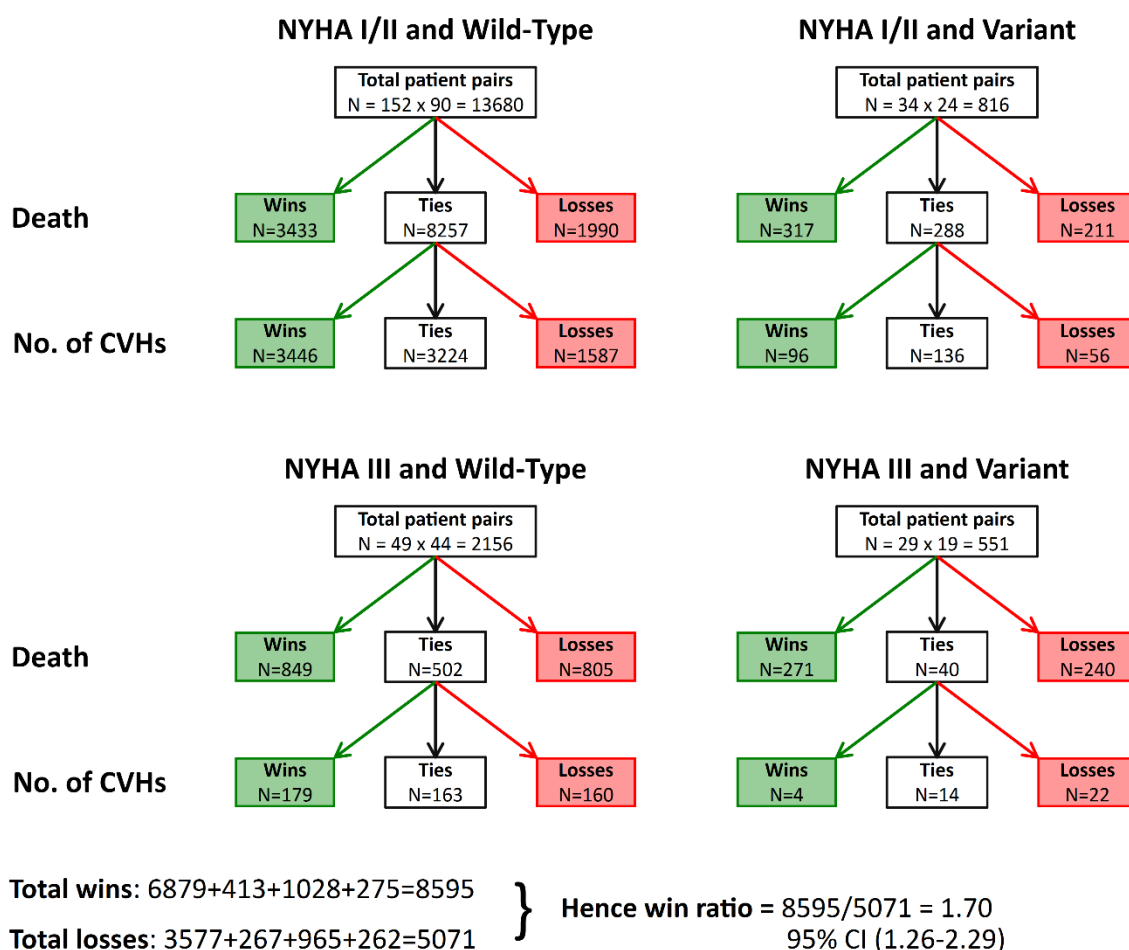
CRNM = clinically relevant nonmajor; VKA = vitamin K agonist. Other abbreviations as in **Table 1**.

Figure 1: Key findings from the CABANA trial, and other studies comparing catheter ablation and drug therapy in patients with atrial fibrillation



* The primary endpoint for CABANA (and OPTUM) is death, disabling stroke, serious bleeding or cardiac arrest. + In the CABANA trial the rate of crossovers in the ablation and drug therapy arms were 9.2% and 27.5% respectively. In the CASTLE-AF trial the corresponding crossover rates were 15.6% and 9.8% respectively. NR=not reported.

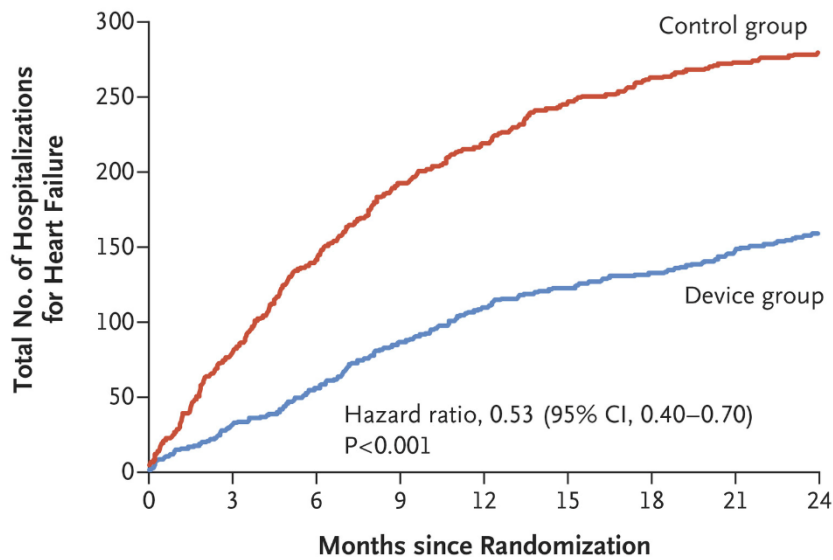
Figure 2 Win ratio analysis of the primary hierarchical endpoint of a) all-cause death and b) frequency of cardiovascular hospitalizations (CVH) in the ATTR-ACT trial. Patients are in 4 strata based on TTR status (variant or wild type) and NYHA class (I/II or III).



Interpretation of win ratio 1.7: if you compare any two patients in the same stratum, one on tafamidis and one on placebo, and they are not a tie (both alive with same number of CVHs) then the odds that the tafamidis patient wins is 1.7. For people who prefer probabilities, then the probability that the tafamidis patient wins is $1.7/(1.7+1) = 0.63$.

Figure 3: Cumulative incidence of A) heart failure hospitalizations (including repeats) and B) death from any cause in the COAPT trial

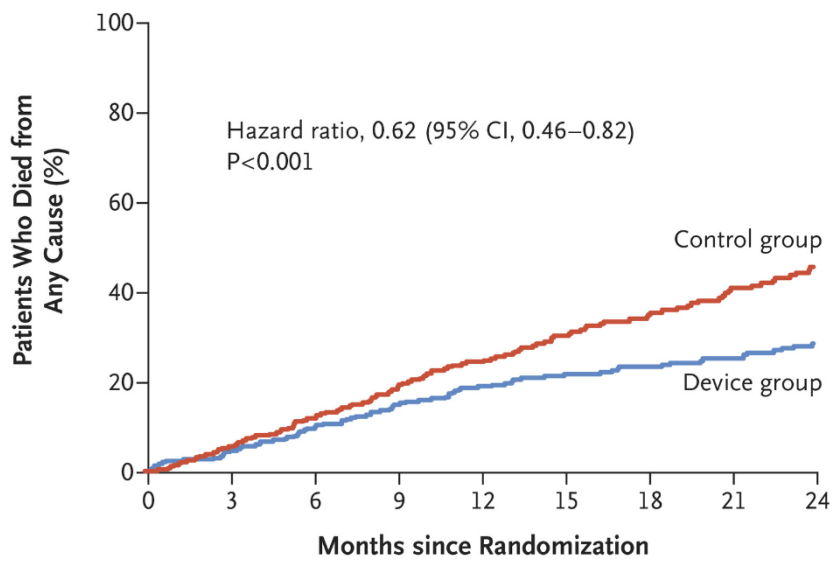
A Hospitalization for Heart Failure



No. at Risk

| | | | | | | | | | |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Control group | 312 | 294 | 271 | 245 | 219 | 176 | 145 | 121 | 88 |
| Device group | 302 | 286 | 269 | 253 | 236 | 191 | 178 | 161 | 124 |

B Death from Any Cause



No. at Risk

| | | | | | | | | | |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Control group | 312 | 294 | 271 | 245 | 219 | 176 | 145 | 121 | 88 |
| Device group | 302 | 286 | 269 | 253 | 236 | 191 | 178 | 161 | 124 |

Figure 4: A comparison of the COAPT and MITRA-FR trials for A) composite of death and heart failure hospitalization and B) all-cause death.

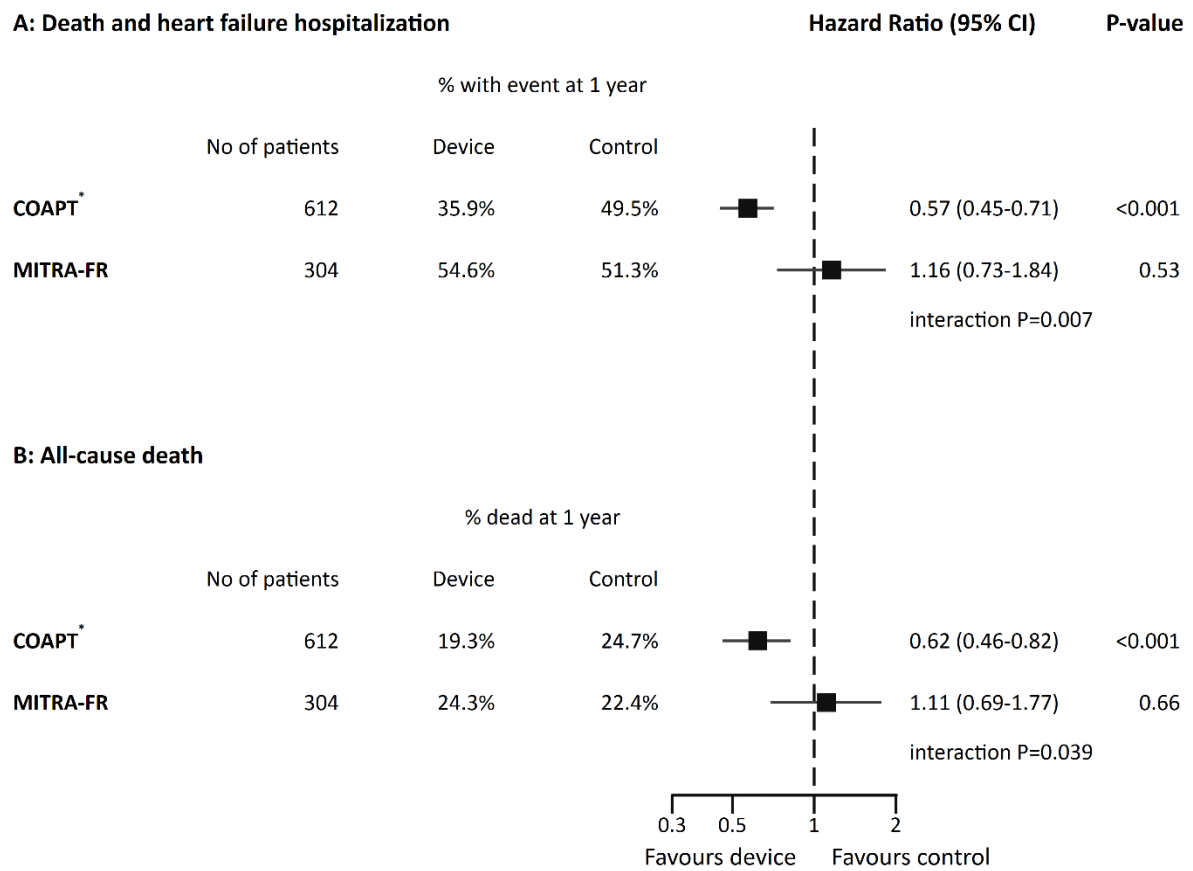
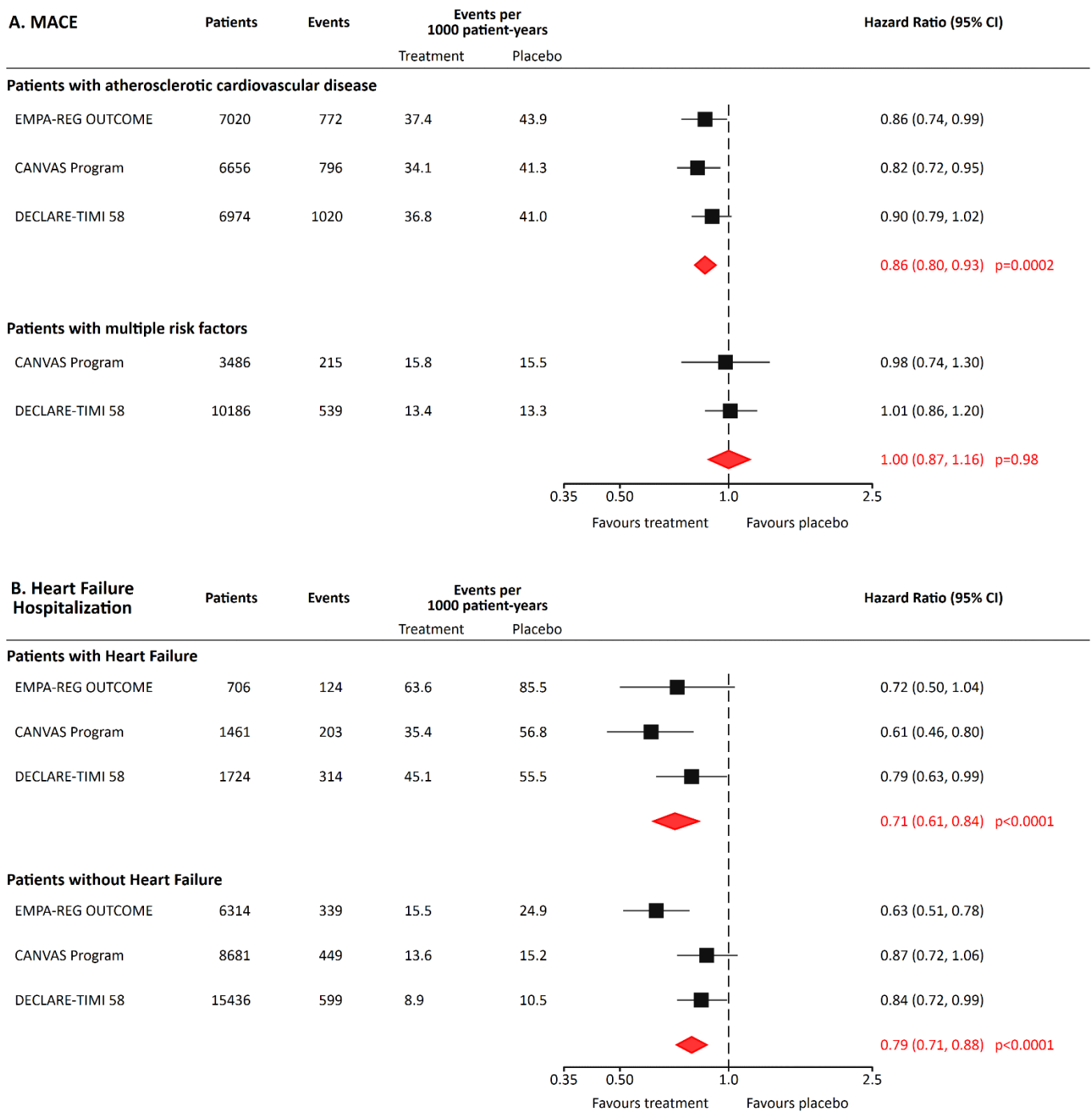


Figure 5: Fixed effect meta-analysis of three SGLT2i trials for A) MACE B) heart failure hospitalization, C) cardiovascular death and D) all cause death.



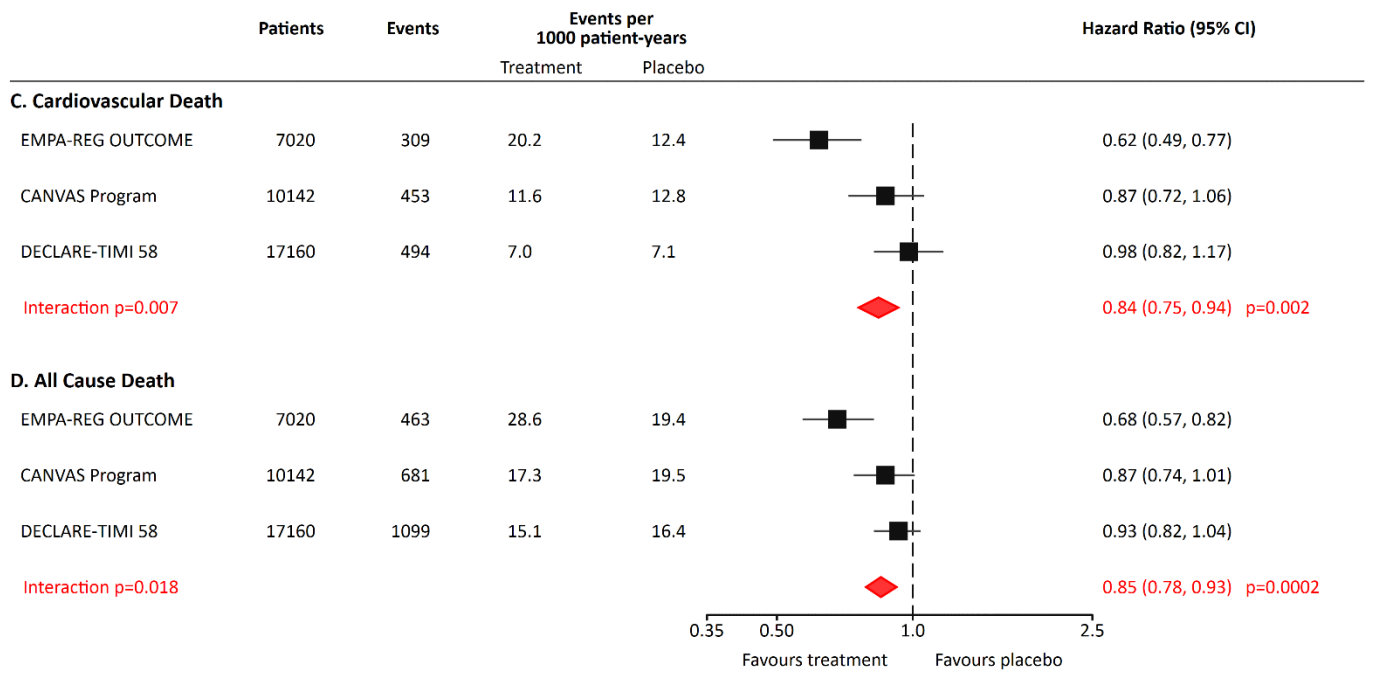
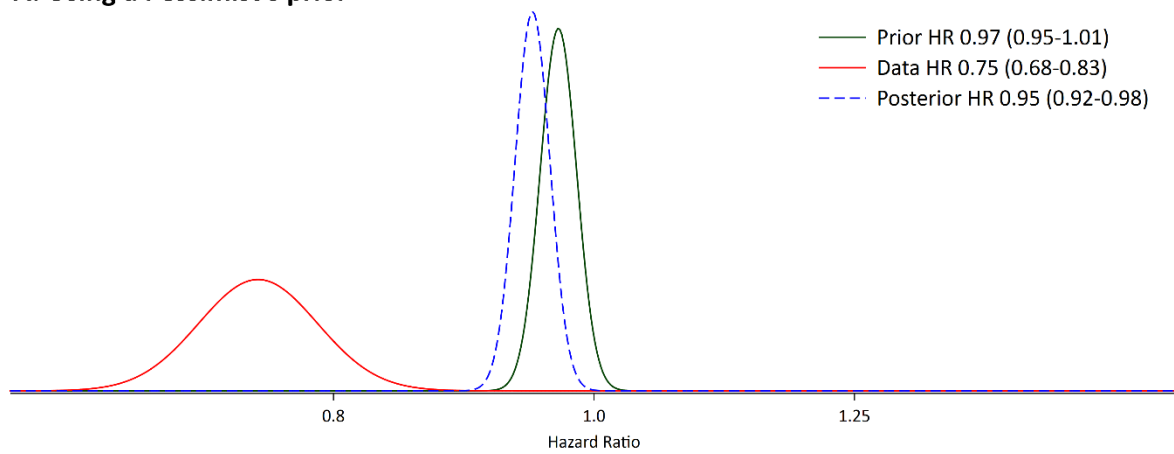
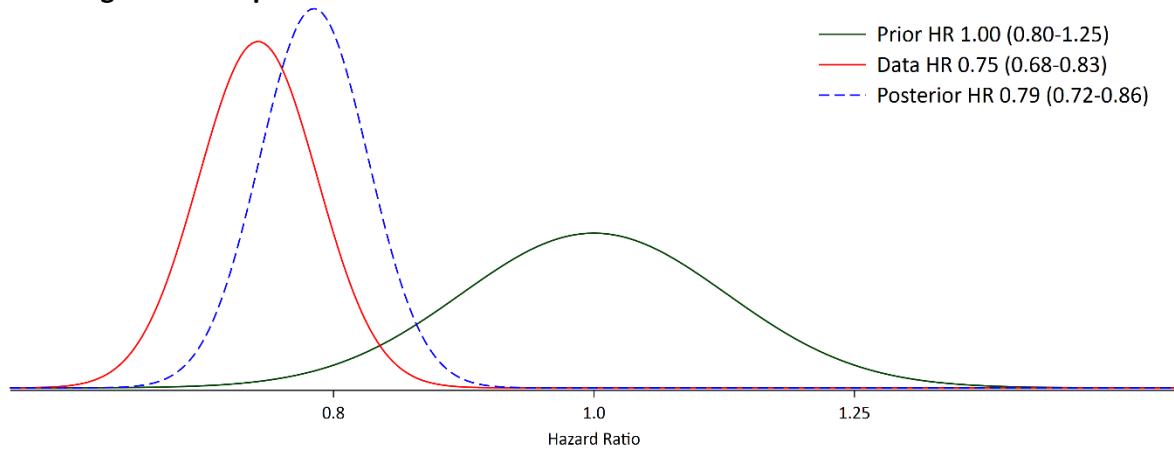


Figure 6: Alternative Bayesian analyses and interpretations of the REDUCE-IT trial results for its primary endpoint*

A: Using a Pessimist's prior



B: Using a Realist's prior



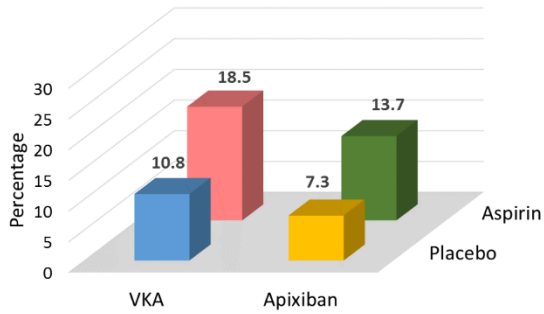
* A: using a pessimist's prior based on a meta-analysis of previous trials of omega 3 fatty acids, the posterior belief remains cautious

B: using a realist's prior which cautiously accepts that some modest benefit (or harm) of EPA is possible, the evidence from REDUCE-IT is very influential producing a posterior belief distribution centred on a 21% risk reduction

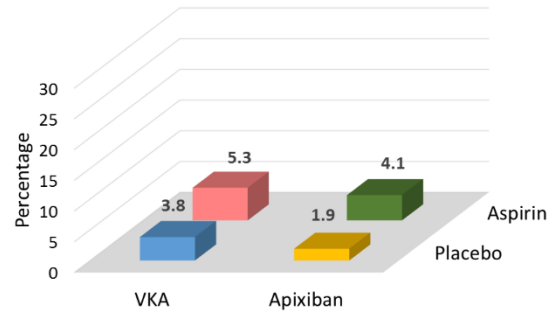
C: using an optimist's vague prior, implying previous negative studies are irrelevant, the posterior belief is the same as the REDUCE-IT evidence as shown by the red distribution curves in A and B.

Figure 7 Main findings from the AUGUSTUS trial for the four randomized treatment options: apixiban+aspirin (green), apixiban+placebo (gold), VKA+aspirin (red), VKA+placebo (blue). Number of patients were 1145, 1143, 1123 and 1126 respectively.

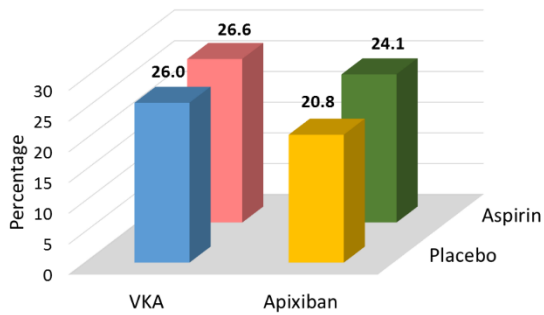
A Major or CRNM bleed



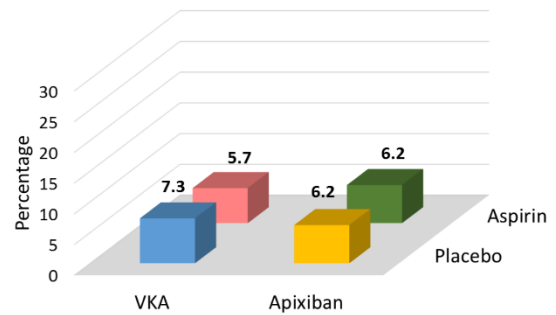
B Major bleed



C Hospitalization



D Death or ischemic events*



* Ischemic events are: MI, stroke, stent thrombosis, urgent revascularization