View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY VOL. 64, NO. 15, 2014
© 2014 BY THE AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION
ISSN 0735-1097/$36.00
PUBLISHED BY ELSEVIER INC. http://dx.doi.org/10.1016/j.jacc.2014.08.008

THE PRESENT AND FUTURE

STATE-OF-THE-ART REVIEW

Do Current Clinical Trials Meet Society's Needs?

A Critical Review of Recent Evidence

Stuart J. Pocock, PhD,* Bernard J. Gersh, MB, ChB, DPhil†

ABSTRACT

This paper describes some important controversies regarding the current state of clinical trials research in cardiology. Topics covered include the inadequacy of trial research on medical devices, problems with industry-sponsored trials, the lack of head-to-head trials of new effective treatments, the need for wiser handling of drug safety issues, the credibility (or lack thereof) of trial reports in medical journals, problems with globalization of trials, the role of personalized (stratified) medicine in trials, the need for new trials of old drugs, the need for trials of treatment withdrawal, the importance of pragmatic trials of treatment strategies, and the limitations of observational comparative effectiveness studies. All issues are illustrated by recent topical trials in cardiology. Overall, we explore the extent to which clinical trials, as currently practiced, are successful in meeting society's expectations. (J Am Coll Cardiol 2014;64:1615–28)
© 2014 by the American College of Cardiology Foundation.

Randomized clinical trials (RCTs) are accepted as the source of the highest level of evidence for assessing the efficacy and safety of potential new treatments by guidelines and regulatory authorities. Indeed, innumerable important advances in patient care, including the abandonment of biologically plausible but ineffective or unsafe treatments, have been based upon rigorous scrutiny from major pivotal RCTs.

Despite such successes, it is relevant to ask to what extent the whole field of clinical trials research as currently practiced does, in fact, meet society's needs.

Here we focus on several topical controversies from a cardiovascular (CV) perspective, each illustrated by recent clinical trials. The aim throughout is to note deficiencies and encourage improvements, thus enhancing what the public should expect in terms of the extent of clinically-relevant advances in treatment and health derived from RCTs.

PLACEBO EFFECT AND MEDICAL DEVICES

Until the recent SYMPLICITY HTN-3 (Renal Denervation in Patients With Uncontrolled Hypertension) trial revealed its negative findings (1), there was much collective expectation that renal denervation could be a very effective intervention in resistant hypertension. In 2009, SYMPLICITY 1, an uncontrolled trial of 45 patients, found a marked decrease in systolic blood pressure (SBP) after 12 months (2). SYMPLICITY 1 was subsequently expanded to report a mean 22 mm Hg decrease in SBP at 6 months in 86 patients (3). In 2010, SYMPLICITY 2, a randomized, unblinded, controlled trial in 100 patients was equally positive, with mean 6-month SBP reductions

From the *Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom; and the †Department of Cardiovascular Diseases, Mayo Clinic College of Medicine, Rochester, Minnesota. Dr. Pocock has received research grants from GlaxoSmithKline, The Medicine Company, Pfizer, Edwards Lifesciences, Biosensors, Amgen, Boston Scientific, AstraZeneca, and Janssen. Dr. Gersh has been on data safety monitor boards or other trial committees for Medtronic Inc., Baxter Healthcare Corporation, Cardiovascular Research Foundation, Merck & Co, Inc., St. Jude Medical, Ortho-McNeil Janssen Scientific Affairs, TEVA Pharmaceuticals, and Boston Scientific.

Manuscript received June 30, 2014; revised manuscript received July 29, 2014, accepted August 6, 2014.

ABBREVIATIONS AND ACRONYMS

CABG = coronary artery bypass graft

CI = confidence interval

CV = cardiovascular

FDA = U.S. Food and Drug Administration

GPI = glycoprotein IIb/IIIa inhibitor

PCI = percutaneous coronary intervention

RCT = randomized clinical trial

SBP = systolic blood pressure

STEMI = ST-segment elevation myocardial infarction

of 32 and 1 mm Hg, respectively, in the renal denervation and control arms (4).

Because of the recognized potential for bias in these studies, to obtain more rigorous evidence and, specifically, to satisfy the U.S. Food and Drug Administration's (FDA's) regulatory requirements, SYMPLICITY 3 was a larger RCT comparing renal denervation with a sham procedure control group in a 2:1 randomization ratio: patients and those assessing outcomes were blinded as to who got what (1). In the renal denervation (n = 364) and sham procedure (n = 171) arms, mean 6-month SBP reductions were 14 and 12 mm Hg, respectively: a nonsignificant difference of only 2 mm Hg. The difference between SYMPLICITY 3 and the earlier findings is very marked (**Figure 1**).

Advocates of renal denervation are exploring possible deficiencies in SYMPLICITY 3. Did patients not truly have resistant hypertension? Are there specific subsets of patients with hypertension who would benefit? Was drug use different in the 2 arms? Were operators too inexperienced? Are better devices now available? The obvious explanation is that renal denervation appears to be insufficiently effective in reducing SBP in this population and that previous findings reflect a substantial placebo effect, regression to the mean, and the possibility that patients with "refractory hypertension" became adherent to drug therapy once enrolled into the trial. Additional trials would be helpful in establishing the role (or lack thereof) for renal denervation in hypertension. However, the story thus far indicates that the hype of an illusory breakthrough in management of resistant hypertension was perpetuated by inadequately-designed RCTs, which gave exaggerated findings and did not take into account the power of the placebo.

Another excellent example of the placebo effect arises from trials of permanent pacemakers in patients with vasovagal syncope (5). A meta-analysis of 9 trials demonstrated that in unblinded studies, active pacing resulted in a striking reduction in recurrent syncope (odds ratio: 0.09, 95% confidence interval [CI]: 0.04 to 0.22). Nonetheless, when patients with permanent pacemakers were blinded as to whether or not the pacing modes were activated, there was no significant effect (odds ratio: 0.83; 95% CI: 0.41 to 1.70).

These experiences have important wider implications for research into medical devices. Of particular note are the very undemanding requirements for medical device approval in Europe (6). The CE Mark needed to market a device in the European Community does not usually require evidence from RCTs; renal denervation is 1 such example. Relatively small, uncontrolled studies focusing on performance objectives, rather than valid evidence of efficacy and safety, are assessed by Notified Bodies, who are widely recognized as lacking appropriate scientific objectivity. Consequently, such an easy, nonrigorous approval process carries risks that patients could be exposed to ineffective and/or unsafe devices.

There is understandable frustration in the United States that new devices get approved much more slowly than in Europe. Thus, for coronary stents and transcatheter aortic valve replacements, there is the perception that U.S. patients have a substantial delay in access to effective new devices compared with Europeans. Although there may be room for a more expedited approval process within the FDA, critics of the current approach need to recognize that efficacy and safety can be truly determined only after the FDA-required RCTs are performed. The real problem lies in Europe: there is a need for radical reform of how medical devices get approved in Europe, both in terms of the currently inadequate process and the need for well-designed RCTs to be a fundamental part of the mandated evidence base.

PROBLEMS WITH PHARMACEUTICAL TRIALS: THE BIVALIRUDIN EXPERIENCE

The results of the HEAT PPCI (How Effective are Antithrombotic Therapies in Primpary PCI) trial (7), which compared bivalirudin with unfractionated heparin in 1,892 primary percutaneous coronary intervention (PCI) patients followed for 28 days, provoked considerable debate. The primary composite outcome (death, stroke, reinfarction, or unplanned target lesion revascularization) was higher in the bivalirudin group (8.7% vs. 5.7%; p = 0.01), as was stent thrombosis (3.4% vs. 0.9%; p = 0.001), whereas there was no evidence of a difference in major bleeding (3.5% vs. 3.1%; p = 0.59). The use of glycoprotein IIb/IIIa inhibitors (GPIs) was similarly low in both groups (13.5% vs. 15.5%).

This apparent inferiority of bivalirudin seems to contradict evidence from 3 previous trials, each claiming superiority of bivalirudin alone versus heparin + GPI. The ACUITY (Acute Catheterization and Urgent Intervention Triage strategy) trial (8) in 13,819 patients with acute coronary syndrome showed bivalirudin to be noninferior for 30-day composite ischemia (death, myocardial infarction, and revascularization) (7.8% vs. 7.3%) and superior for major bleeding (3.0% vs. 5.7%). The HORIZONS-AMI

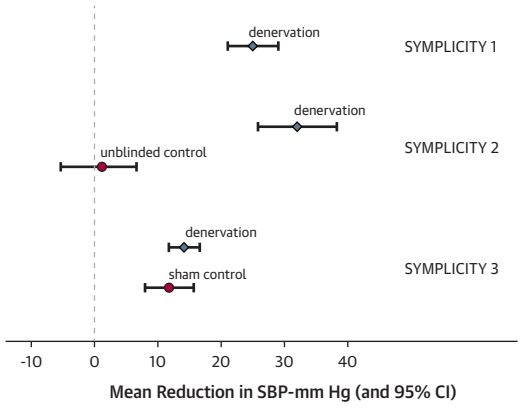

FIGURE 1 A Comparison of Changes in SBP at 6 Months in 3 Trials of Renal Denervation

CI = confidence interval; SBP = systolic blood pressure.

(Harmonizing Outcomes With Revascularization and Stents in Acute Myocardial Infarction) trial (9) in 3,602 primary PCI patients had coprimary 1-year endpoints of major bleed (5.8% vs. 9.2%; $p < 0.0001$) and net adverse clinical events (15.6% vs. 18.3%; $p = 0.02$). The EUROMAX (European Ambulance Acute Coronary Syndrome Angiography Trial) (10) in 2,218 primary PCI patients had a primary 30-day composite endpoint of death and major bleed (2.6% vs. 6.0%; $p < 0.0001$). Reinfarction (1.7% vs. 0.9%, $p = 0.08$) was removed from this primary outcome after the trial had started. In all of these industry-sponsored trials, there was a high use of GPI in the heparin arm (97% in the ACUITY, 94.5% in the HORIZONS-AMI, and 69.1% in the EUROMAX trial), and this addition of GPI to heparin is what provokes the increased bleeding risk.

Thus, routine practice should now entail much lesser use of GPI (more as a bailout rather than up front), as is reflected in the HEAT-PPCI trial. This trial was investigator-initiated and included all eligible patients in a single center. It is the only head-to-head comparison of bivalirudin and heparin with equal, appropriately low GPI use. It has its critics. Was the bivalirudin dosage too low? Was the delayed consent process ethically appropriate (an important debate, but one that does not affect the actual results)? The main point here is that industry-sponsored trials to date have not directly answered the pivotal question: how does bivalirudin compare with heparin in the context of equal and appropriate background GPI use? The HEAT-PPCI trial findings indicate an urgent need for a further definitive RCT of this issue.

On a more general note, the clinical trial programs that industry undertakes are understandably dedicated to getting their product licensed by regulators (e.g., the FDA and European Medicines Agency [EMA]) and subsequently approved by health technology assessments, such as the National Institute for Health and Care Excellence in the United Kingdom. This process does not necessarily deliver RCTs that are in the best public interest. De Mets and Califf (11) assert in their historical perspective on clinical trials innovation and leadership that, "a better balance between commercial interests and public health is critically needed." For instance, in chronic heart failure, it is common to evaluate any new drug in a pivotal placebo-controlled trial in addition to the multidrug regimens that patients already receive. Specifically, the SHIFT (Systolic Heart Failure Treatment with Ivabradine) trial (12) of ivabradine was placebo-controlled, with beta-blocker therapy as part of such background therapy. A head-to-head comparison of ivabradine versus beta-blocker (both lower heart rate) would be of great interest, but commercial sponsorship of such a trial appears unlikely. These problems are not necessarily due to company policy, but may well arise because regulators require placebo-controlled trials upon a background of current standard therapy.

THE LACK OF HEAD-TO-HEAD TRIALS: NEW ANTICOAGULANTS

In atrial fibrillation, several large trials have compared new anticoagulants against warfarin: the RE-LY (Randomized Evaluation of Long Term Anticoagulant Therapy) (13), ROCKET-AF (Rivaroxaban Once Daily Compared with Vitamin K Antagonism for Prevention of Stroke and Embolism Trial in Atrial Fibrillation) (14), ARISTOTLE (Apixaban for the Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation) (15), and ENGAGE (Effective Anticoagulation with Factor Xa Next Generation in Atrial Fibrillation) (16) trials studied dabigatran, rivaroxaban, apixaban, and edoxaban, respectively. Each trial had stroke and systematic embolism as the composite primary endpoint and major bleeding as the key safety endpoint. Their findings are summarized in **Figure 2**, with all estimates (and 95% CIs) being hazard ratios for new anticoagulant versus warfarin. Collectively, they tell a success story whereby new anticoagulants tend overall to reduce the risk of both the primary endpoint and major bleeding. The next logical question is whether there are real differences in their efficacy and safety, and, if so, which new anticoagulant appears to be best? The indirect evidence is plotted in **Figure 2**.

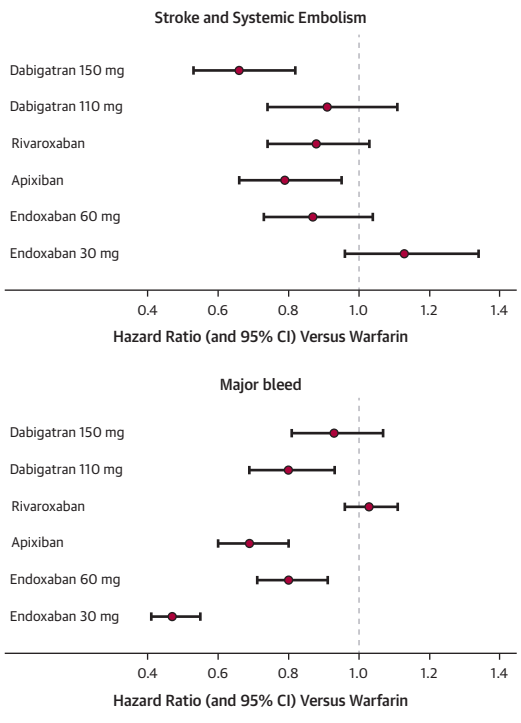

FIGURE 2 Comparisons of 4 New Anticoagulants Versus Warfarin in the RE-LY, ROCKET-AF, ARISTOTLE, and ENGAGE Trials

CI = confidence interval.

At face value, the higher dose of dabigatran has the greatest reduction in risk of stroke and systemic embolism, whereas apixaban has similar efficacy and fewer major bleeds. But, these assertions are unreliable because they rely on indirect comparisons via warfarin across different trials. The trials differ in patient selection, outcome definitions, blinding (RE-LY was not double-blind, and does that matter?), and the degree of international normalized ratio control of warfarin, and it is unclear to what extent these issues affect the validity of indirect comparisons (17).

Unfortunately, there appears to be no prospect of a head-to-head trial of 2 (or more) of these new anticoagulants. Thus, which anticoagulant gets used in practice depends on faster regulatory approval, successful marketing, and costs rather than on reliable estimates of relative efficacy and safety. We cannot expect industry to fund these head-to-head trials, so public funds (e.g., the National Heart, Lung, and Blood Institute) are required. Pathways to motivate, fund, and activate such pragmatic head-to-head trials need to be created. A commendable example is the ISAR-REACT (Intracoronary Stenting and Antithrombotic Regimen: Rapid Early Action for Coronary Treatment) 5 trial (18) comparing ticagrelor and prasugrel in patients with an acute coronary syndrome. How can we stimulate such initiatives to become a more common feature of post-licensing evaluation of effective new agents?

THE CHOICE OF DOSE IN PIVOTAL DRUG TRIALS

A key issue is the choice of dose for each drug in a pivotal Phase III trial. The RE-LY trial in atrial fibrillation is commendable for having studied 2 different doses of dabigatran (150 and 110 mg). The former appears more efficacious and the latter has less bleeding: a logical finding leaving clinicians an appropriate choice as to what is best for the individual patient. Both doses are approved in Europe. However, in the United States, the FDA did not approve the 110 mg dose, opting for 75 mg instead.

The more general point here is that any comparison of drugs is on the basis of the specific doses chosen for their pivotal trials, and the potential for a wrong (or somewhat inferior) choice of dose is always present. For instance, in testing such anticoagulants in patients with an acute coronary syndrome, the APPRAISE 2 (Apixaban for Prevention of Acute Ischemic Events - 2) (19) trial of apixaban was terminated prematurely for safety reasons (perhaps because of too high a dose), whereas the 3-arm ATLAS ACS 2 (Anti-Xa Therapy to Lower Cardiovascular Events in Addition to Standard Therapy in Acute Coronary Syndrome) (20) of rivaroxaban compared 2 doses against placebo, showing that the lower dose appeared superior.

Across all pharmaceutical research, getting the dose right for the pivotal phase III trial of a new product is a major challenge. Wrong choices can have disastrous consequences, both for patients and for a promising drug. A wiser admission of the remaining uncertainty after the phase II trial evidence is in may justify the need for more than 1 dose choice in phase III (although the problem of increased trial size is an issue). Another risk is that, in the race to seek faster licensing approval, phase III trials may be initiated before reliable evidence of correct dose choice exists. For instance, did the excess mortality of moxonidine in chronic heart failure (21) arise because of a hasty choice of dose in the phase III trial before phase II was completed?

THE NEED FOR BETTER HANDLING OF DRUG SAFETY: ANTIDIABETIC AGENTS

Potential safety problems with a drug are often not handled wisely. On the one hand, companies may confidently assert that their drug is safe; on the other

hand, activists who are convinced that a drug is harmful may go to great lengths to promote their case. This can escalate into undue media attention (scare stories) and political involvement, whereby the pressures on regulators make it difficult to achieve a calm, objective assessment of the totality of evidence. A case in point concerns the potential CV risks associated with rosiglitazone in diabetic patients. A meta-analysis published in 2007 (22) reported a 42% increased risk of myocardial infarction ($p = 0.04$) in patients taking rosiglitazone versus control subjects. The evidence was weak, it was from a heterogeneous mix of mostly small trials with short-term follow-up, and myocardial infarction events were not validated. The journal *Nature* referred to this meta-analysis as "a rushed and incomplete examination," but it engendered such a high profile that the FDA had to instigate urgent inquiries into rosiglitazone's CV safety.

The RECORD (Rosiglitazone Evaluated for Cardiovascular Outcomes and Regulation of Glycaemia Diabetes) trial (23) provided the only specifically-designed randomized evidence on the topic, as previously requested by the European Medical Agency. Its primary findings of no overall effect of rosiglitazone on all CV hospitalizations and deaths were accompanied by inconclusive evidence regarding myocardial infarction. The trial was not double-blind, and accusations that it was open to bias held sway in the FDA hearing. The *Lancet's* caution that "alarmist headlines and confident assertions help nobody" was not the predominant mood of the time. However, 3 years later (in 2013), an FDA panel reassessment concluded, "rosiglitazone is not associated with excess myocardial infarction." This relied on an independent readjudication of events in the RECORD trial, with no change in its conclusions. In hindsight, this evolving debate was too focused on the risk of myocardial infarction. The RECORD trial found highly significant excesses of "apparent" heart failure events (perhaps due to fluid retention of uncertain significance) and bone fractures on rosiglitazone compared with control (hazard ratios: 2.10 and 1.57, respectively), but these concerns received little attention in the FDA process. In Europe, rosiglitazone was suspended in 2010 because its benefits no longer outweigh its risks, a verdict based on the overall drug profile.

Largely led by this rosiglitazone story, the FDA guidance for industry on cardiovascular risk in new antidiabetic agents (24) mandated evidence of cardiovascular safety for any such drug. For licensing approval, one needed to rule out an 80% excess cardiovascular risk, to be followed by a subsequent large cardiovascular safety trial to rule out a 30% excess risk. Thus, large noninferiority trials of each such drug versus placebo are in progress, usually with a composite primary endpoint of CV death, myocardial infarction, and stroke, and recruitment of diabetic patients at high risk of CV events. It is curious that heart failure is not included in the primary endpoint, given its apparent excess on rosiglitazone.

Two such trials have now been completed, both concerning gliptins, a different class of drugs from the glitazones. The SAVOR-TIMI 53 (Saxagliptin Assessment of Vascular Outcomes Recorded in Patients with Diabetes Mellitus-Thrombolysis In Myocardial Infarction 53) trial of saxagliptin (25) and the EXAMINE (Examination of Cardiovascular Outcomes with Algorithms versus Standard of Care) trial of alogliptin (26) showed no effects on the primary endpoint. However, both trials had a numerical excess of heart failure hospitalizations compared with placebo: hazard ratios were 1.27 (95% CI: 1.07 to 1.51) and 1.07 (95% CI: 0.79 to 1.46), respectively, the former (saxagliptin) was statistically significant ($p = 0.007$). A cautious interpretation is appropriate: heart failure was one of several secondary hypotheses and the interaction test comparing the drugs' heart failure effects (via an indirect comparison) is not significant. Thus, any claim that saxagliptin alone causes heart failure is unwarranted. The results of the TECOS (Trial Evaluating Cardiovascular Outcomes with Sitagliptin) of sitagliptin (27) are awaited with interest.

We now recognize that cardiovascular safety is a key issue in the development and use of antidiabetic drugs; thus, demonstration of glycemic efficacy alone is not sufficient evidence for regulatory approval by both the FDA and the EMA. The more flexible approach of the latter is, perhaps, more appropriate than the FDA's formal noninferiority trial strictures. What is sadly lacking is an antidiabetic drug that actually reduces CV risk.

There have been other drug safety concerns initiated by claims based on weak evidence. For instance, the associations between calcium-channel blockers and myocardial infarction, ezetimibe and cancer, and angiotensin-receptor blockers and cancer all invoked much research and/or expert investigation leading to conclusions of no safety problem. Rofecoxib was rightly associated with an increased risk of myocardial infarction, although initial relative risk estimates were higher than those subsequently found in more substantive meta-analysis. This is likely to be a recurring theme. That is, concerns are stimulated because initial claims made on the basis of limited evidence are prone to exaggeration. More substantial

evidence to follow usually demonstrates regression to the truth. The challenge is whether that "truth" is a more modest risk or none at all.

In the case of rofecoxib, the risk of myocardial infarction appears to be real, but whether that justified a complete ban is open to debate. Given its marked value in pain relief for patients at low risk of myocardial infarction, a judgment could be made that the low excess absolute risk would be outweighed by the drug's undoubted efficacy.

Two big problems concerning safety signals are: 1) because they are often unanticipated, quality data on harms may not be collected uniformly across trials; and 2) trial exclusion criteria may remove some patients for whom the drug is less safe, whereas such patients subsequently become exposed to the drug in routine practices. This makes post-licensing surveillance of drug safety an important issue.

CREDIBILITY OF CLINICAL TRIAL REPORTS IN MEDICAL JOURNALS

Much attention has been paid by trialists, editors, referees, and guidelines (e.g., CONSORT [28]) to ensure that clinical trial reports in medical journals are of high quality, and improvements have been made over the years. However, we feel that commercial pressures and investigator enthusiasm to assert positive claims combined with publication bias, whereby more "exciting" findings are more likely to reach major journals, means that the clinical trials literature still has an overall tendency toward exaggerated claims. We explore this with a few recent examples.

The CoreValve trial (29) comparing transcatheter aortic valve replacement with surgery concluded that transcatheter aortic valve replacement "was associated with a significantly higher rate of survival at 1 year." The 1-year mortality rates were 14.2% versus 19.1%, and the investigators used an unconventional 1-sided significance test to claim $p = 0.04$. A conventional analysis would report a difference of −4.9% with a 95% CI from −10.3% to +0.5%, $p = 0.08$. These data are suggestive of a potential survival benefit, but the claim by the investigators is overassertive.

A common scenario is when a major clinical trial produces a disappointing neutral finding for the overall pre-defined primary endpoint, but investigators are tempted to make other positive claims in secondary endpoints and/or subgroup analyses. For instance, the BEAUTIFUL (Morbidity-Mortality Evaluation of Ivabradine in Patients with Coronary Disease and Left Ventricular Dysfunction) trial (30) of ivabradine versus placebo in patients with coronary disease and left ventricular systolic dysfunction had almost identical incidence rates for the composite primary endpoint (CV death, myocardial infarction, and congestive heart failure) ($p = 0.94$). The investigators proceeded to emphasize the difference in myocardial infarction only in the subgroup with heart rate ≥70 beats/min: 3.1% versus 4.9% ($p = 0.001$), highlighting this finding in both the abstract and conclusions sections. We feel that it was inappropriate to assert such a positive claim. It is legitimate to document such a secondary finding as data exploration and hypothesis generating, but it should not affect the overall negative conclusions of the trial.

The CURRENT OASIS 7 (Clopidogrel and Aspirin Optimal Dose Usage to Reduce Recurrent Events) trial (31,32) compared standard versus double-dose clopidogrel in patients with an acute coronary syndrome. A *Lancet* paper concluded, "in patients undergoing PCI, double dose was associated with a reduction in CV events" (31). The same weekend, a paper in the *New England Journal of Medicine* concluded, "in patients referred for an invasive strategy, there was no significant difference between double-dose and standard dose" (32). This apparent contradiction in findings came about because the *Lancet* paper concerns a selected subgroup of 17,232 patients who received PCI: the primary composite endpoint of CV death, myocardial infarction, and stroke within 30 days occurred in 3.9% and 4.5% of the double- and standard-dose groups ($p = 0.039$). The *New England Journal of Medicine* paper included all 25,087 patients with a primary endpoint comparison of 4.2% versus 4.4% ($p = 0.37$). The interaction test ($p = 0.03$ for those receiving PCI vs. the rest) suggests a possible heterogeneity of treatment effect, but such secondary evidence favoring double-dose clopidogrel is weak. The *Lancet* paper is problematic because it did not make explicit that it was a subgroup finding, that receiving PCI is an improper subgroup in the sense that it was only determined after randomization, and that such a qualitative interaction (primary endpoint 4.2% vs. 4.9% in the opposite direction for the 7,855 patients not receiving PCI) is rare and implausible.

The SHIFT trial (12) of ivabradine versus placebo in chronic heart failure reported very strong evidence of a reduction in incidence of the primary endpoint, CV death and heart failure hospitalization: hazard ratio: 0.82 ($p < 0.0001$). But, there was an interaction with baseline heart rate, so what appeared in the *Lancet* as an overall positive finding was a more restricted approval by the EMA: "ivabradine is indicated in chronic heart failure patients… whose heart rate is ≥75 bpm."

An overall picture emerges regarding reporting subgroup findings and other secondary analyses. They should be reported in a cautious spirit of exploratory data analysis, with authors and journals exercising restraint, especially when the primary overall trial result is not positive. Conversely, when the overall conclusion is positive, secondary findings (e.g., on subgroups, adverse events) may appropriately place restrictions on the future population of patients who truly benefit from the new treatment.

One concern is that the most important trials are processed as "fast-track" publications, whereby the time from database lock to online release in a major journal is made as short as possible. We fear that such breakneck speed combined with the very limited journal space can occasionally be detrimental to the quality of publication. From a regulatory perspective, journal publications are of limited value because they report such a small, selected subset of findings compared with the much more extensive regulatory dossiers.

THE GLOBALIZATION OF RCTs: DOES GEOGRAPHY MATTER?

Many major trials are multiregional, both to get the large sample size required and to achieve a more globally representative patient population, especially when approval is sought from regulators in different parts of the world. The problem is that the ability to recruit patients can differ markedly across regions: Eastern European centers commonly recruit faster than centers in the United States and Western Europe. This raises concerns over whether such geographic disparities in recruitment affect the validity of the overall trial findings, especially if the future population for the approval and use of new expensive treatment lies in relatively wealthy countries. Although one is obligated to explore trial data for geographic differences in treatment effect, this is a particularly tricky form of subgroup analysis that lacks statistical power. Hence, real geographic heterogeneity may well go undetected. On the other hand, any observed statistically significant heterogeneity across countries or regions is liable to be an exaggeration of the truth.

The TOPCAT (Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist) trial (33) of spironolactone in preserved ejection fraction heart failure is an intriguing example of the challenges faced in conducting and interpreting geographic differences. Overall, the trial showed a modest nonsignificant trend toward the superiority of spironolactone versus placebo: the composite primary endpoint (CV death, heart failure hospitalization, and resuscitated cardiac arrest) had a hazard ratio: 0.89 (95% CI: 0.77 to 1.04), with $p = 0.138$. To facilitate patient recruitment in this National Institutes of Health–funded trial, nearly one-half of the patients (1,678 of 3,445) were in Russia and Georgia. But, the incidence rate of the primary endpoint in the placebo group in these 2 countries was less than one-fifth of that in the other Western countries (United States, Canada, Argentina, and Brazil). Hazard ratios for the primary endpoint were 1.10 in Russia and Georgia and 0.82 in the rest. The latter achieves statistical significance ($p < 0.05$), but the formal test for heterogeneity in hazard ratios across regions is not statistically significant (interaction $p = 0.12$). It is not clear exactly why findings in Russia and Georgia are so discrepant with the rest: is it patient selection, healthcare systems, patterns of follow-up, or data quality issues that contribute? Overall, this trial is a valuable lesson that enhancing recruitment in other regions may not necessarily deliver generalizable findings and can sometimes hamper the ability to detect important treatment differences.

The PLATO (Study of Platelet Inhibition and Patient Outcomes) (34) in patients with acute coronary syndrome is a very different scenario, where apparent geographic heterogeneity cast doubt on whether the overall trial result was generalizable to patients in the United States. Compared with clopidogrel, ticagrelor had fewer primary endpoints (CV death, myocardial infarction, and stroke), with an overall hazard ratio: 0.84 (95% CI: 0.77 to 0.92; $p < 0.001$). However, in North America, the hazard ratio was 1.25 (95% CI: 0.93 to 1.67) and the test for heterogeneity across 4 regions had an interaction $p = 0.05$. Furthermore, the post-hoc interaction test comparing the United States with all other countries combined had $p = 0.009$. The search across more than 50 potential explanatory factors for this geographic anomaly revealed that differences in the maintenance dose of aspirin could be relevant: high-dose aspirin (≥300 mg) was administered to 54% of U.S. patients and to <2% of non-U.S. patients. The primary endpoint hazard ratios for ticagrelor versus clopidogrel were 1.45 and 0.79 in patients on high- and low-dose aspirin, respectively, interaction $p = 0.0006$. After much deliberation, the FDA approved ticagrelor, but with a boxed warning, "use of ticagrelor with aspirin doses exceeding 100 mg/day decreases its effectiveness." However, the main lesson is not about ticagrelor as such, but is to get American practitioners to no longer prescribe high doses of aspirin in such patients. A valid alternative viewpoint is that the geographic anomaly in PLATO could have been due to the play of

chance, because the aspirin dose issue is strongly confounded with region. Such trials are not powered to address potential geographic heterogeneity, and these post-hoc questions can lead to spurious findings (35).

Overall, regional disparities in a global RCT may arise for several reasons: racial/genetic issues; differences in background therapy, healthcare systems, and recruitment patterns; and variation in the quality of trial conduct and data quality. Thus, the greater ease in reaching recruitment targets is offset by these potential complications: caution is warranted.

THE ROLE OF PERSONALIZED (STRATIFIED) MEDICINE IN RCTs

When a report of an RCT reaches a positive conclusion regarding a new treatment's benefit, this may not necessarily apply equally to all eligible patients. Although the perils of subgroup analysis are well known (and mentioned earlier in this review) a treatment's efficacy and safety still need to be considered in the context of the individual patient's risk profile. It is informative to document the absolute treatment differences in the primary endpoint stratified according to patient risk status.

For instance, the EMPHASIS (Eplerenone in Mild Patients Hospitalization and Survival Study in Heart Failure) trial (36) in heart failure patients with mild symptoms compared eplerenone with placebo and found a highly significant reduction in the primary endpoint, heart failure hospitalization or CV death: hazard ratio: 0.63 ($p < 0.0001$), with no evidence of any subgroup interactions. Grouping patients into low-, mid-, and high-risk strata based on a risk score revealed incidence reductions on eplerenone of 2.0, 6.8, and 15.2 primary events per 100 patient-years, respectively. Thus, the absolute treatment benefit is much greater in patients with a poorer prognosis. This approach should be applied more widely to major RCTs, and could better focus attention on high-risk patients, whose absolute benefit is often much greater.

The individual patient's appropriate therapeutic choice becomes more challenging when treatments differ both in efficacy and in safety, but in opposite directions. For instance the TRITON-TIMI 38 (TRial to assess Improvement in Therapeutic Outcomes by optimizing platelet InhibitioN with prasugrel-Thrombolysis In Myocardial Infarction 38) (37) in patients with acute coronary syndrome undergoing PCI found prasugrel superior to clopidogrel for the ischemic primary endpoint, CV death, myocardial infarction, and stroke (9.1% vs. 11.5%, respectively) but inferior to clopidogrel for Thrombolysis In Myocardial Infarction major or minor bleed (5.0% vs. 3.7%, respectively). For each patient, the relative merits of prasugrel and clopidogrel should include consideration of the tradeoff between efficacy and safety, taking account of their individual risks of ischemia and bleeding. Salisbury et al. (38) created multivariable predictive models for ischemic risk and bleeding risk, including some interactions with treatment (e.g., prasugrel's bleeding excess was more pronounced in older patients). The consequence was an ability to predict for any patient whether prasugrel had a net benefit over clopidogrel: this appeared to be true for 42% of patients in the TRITON-TIMI 38 trial. In general, trialists need to do a better job of providing data and tools to apply results to individuals, especially when benefits and risks have differing balance points across the patient population.

These ideas also apply to major trials comparing 2 different doses of a new drug. In the RE-LY (13) and ENGAGE (16) trials of dabigatran (150 mg vs. 110 mg) and edoxaban (60 mg vs. 30 mg) in atrial fibrillation, the higher doses were more effective in stroke prevention, but also caused more major bleeds. For the RE-LY trial, Eikelboom et al. (39) concluded, "the similar overall benefits of the two doses versus warfarin support individualizing the dose based on patient characteristics and physician and patient preferences." We need further quantitative research to facilitate this concept.

A more provocative use of the term "personalized medicine" concerns the search for new biomarker- or genotype-specific treatments (40). Can we find a biomarker that, if positive, indicates a poor patient prognosis (e.g., pro-B-type natriuretic peptide in heart failure) and that also facilitates a new, targeted treatment that is especially effective in biomarker-positive patients? For instance, in some acute coronary syndrome trials, therapies work well in troponin-positive patients but not at all in others. There have been successes in cancer research and treatment in this form of personalized medicine (e.g., herceptin is specific to HER2-positive breast cancer), but we are skeptical whether there will be major breakthroughs in cardiology. Mutations are central to the pathophysiology of cancers, but not to most CV diseases. Nevertheless, the ideas might apply to monogenetic CV diseases like hypertrophic cardiomyopathy, or to mutations affecting clopidogrel or warfarin sensitivity.

The huge multiplicity of potential biomarkers being investigated and the lack of clarity on what "biomarker positive" means carry a risk of pursuing seemingly attractive, but false positive leads. Also,

prognosis in most CV conditions is multifactorial, so the hope for a new "magic bullet" biomarker needs to be seen in the context of the complex of other, already established risk factors. Time will tell whether any biomarker-specific targeted treatments emerge from rigorously-conducted RCTs in cardiology.

THE NEED FOR NEW TRIALS OF OLD DRUGS: THE CASE OF BETA-BLOCKERS

Once a drug (or class of drugs) has achieved regulatory approval there may be some post-approval trials (e.g., safety trials) supported by the sponsor, but sponsors do not usually have any long-term plans for future clinical trials research. Indeed, once the patent duration approaches (and passes) its end, sponsors have even less motivation to stimulate and fund new research. Yet, new, important questions that would be best answered by initiating new clinical trials may arise many years after any effective class of drugs first came into use. The dilemma is how to get such trials publicly funded and activated, given the inevitable lack of commercial sponsorship. We illustrate this problem with 3 outstanding issues concerning beta-blockers where new RCTs would be of great value.

The role of beta-blockers after a myocardial infarction was well studied years ago, but those trials were all done before primary PCI became the standard practice for ST-segment elevation myocardial infarction (STEMI) patients. Thus, an interesting unresolved issue is whether intravenous metoprolol administered before reperfusion in STEMI patients can enhance prognosis. A pilot trial, METOCARD-CNIC (Effect of Metoprolol in Cardioprotection During an Acute Myocardial Infarction) (41), showed reduced infarct size and improved left ventricular function compared with untreated control subjects. A longer-term, double-blind, placebo-controlled RCT of intravenous metoprolol administered before the primary PCI is now needed. Because prevention of post-infarction heart failure is the goal, the proposed primary composite endpoint is heart failure hospitalization and cardiac death over 2 years of follow-up, requiring a trial of around 4,000 patients.

An ongoing controversy concerns the value (or lack thereof) of beta-blocker treatment in patients undergoing noncardiac surgery. Previous guidelines, both in Europe and in the United States, recommended beta-blockers in this setting, but 2 key trials have since been discredited because of alleged scientific misconduct. Furthermore, POISE (Perioperative Ischemic Evaluation Study) (42), the largest remaining trial comparing metoprolol versus placebo, is controversial, with a 30% decrease in nonfatal myocardial infarction ($p < 0.001$) contrasting with a 33% increase in total mortality ($p = 0.03$), perhaps because the beta-blocker dose was too high. Meta-analyses provide inconclusive evidence on the topic. The new European Society of Cardiology/European Society of Anaesthesiology guidelines on noncardiac surgery (43) conclude that, "high priority needs to be given to new randomized clinical trials to better clarify which patients derive benefit from beta-blocker therapy in the preoperative setting, and to determine the optimal method of beta-blockade."

A further issue concerning beta-blocker treatment is whether and when they can be withdrawn in patients with stable coronary disease. This issue is discussed in a broader setting in the following section.

THE NEED FOR TRIALS OF TREATMENT WITHDRAWAL

Patients with stable coronary disease usually take many drugs including, for example, aspirin, beta-blockers, statin, ACE inhibitor, plus other drugs if they are hypertensive, diabetic, and/or have other comorbid conditions. In this context, industry-sponsored trials seeking regulatory approval for a new class of drugs are often conducted as placebo-controlled trials in addition to all of these treatments. Thus, effective new agents may cause an accumulating and potentially excessive polypharmacy, especially in elderly patients.

One concern is that such trials are of limited duration, whereas in routine practice, treatments, both new and established, may be administered open-endedly over many years. As patients age and their renal function declines, the potential for adverse drug interactions and/or drug side effects arises. The reduced efficacy of some drugs over longer medication periods is an important issue that largely remains unexplored research territory. As part of this broader topic of long-term polypharmacy, we need clinical trials that can investigate the withdrawal of certain established medications to see whether such withdrawal induces patient benefit, harm, or no difference compared with continued medication. Which drugs, for which patients, and for which key outcomes are trials of treatment withdrawal ripe for investigation?

The long-term effectiveness of beta-blockers has been questioned. Therefore, a trial of withdrawal of beta-blocker treatment could be along the following lines: patients undergoing PCI for stable coronary disease and who are already on a beta-blocker could be randomized at their post-procedure follow-up visit (typically 6 weeks post-PCI) to withdrawal or

continuation of their beta-blocker. The composite primary endpoint could be all-cause death, myocardial infarction, and hospitalization for heart failure over 1 year of follow-up. Hypotheses in reference to the superiority, inferiority, and noninferiority of withdrawal would all be examined. Data on quality of life, angina symptoms, adverse events, compliance, and other cardiac events would also be evaluated.

Another controversial topic is the duration of dual antiplatelet therapy (e.g., aspirin + clopidogrel) administered to patients undergoing PCI with a drug-eluting stent. Stopping too soon may increase the risk of stent thrombosis and possibly other cardiac events, whereas continuing too long accentuates the risk of bleeding events. This topic has been hotly debated, with various trials of dual antiplatelet therapy duration being initiated. DAPT (Dual Antiplatelet Therapy) (44), the largest of these trials, recruited 20,000 patients, the main comparison being between 12 and 30 months duration, a time comparison now thought to be less relevant. The real question is whether or not dual antiplatelet therapy can be administered for <12 months. OPTIMIZE (Optimized Duration of Clopidogrel Therapy Following Treatment With the Endeavor Zotarolimus-Eluting Stent in the Real World Clinical Practice) (45) is 1 such trial, comparing 3 months versus 12 months of dual antiplatelet therapy after a zotarolimus-eluting stent in 3,119 patients. They showed noninferiority for the composite primary endpoint of death, myocardial infarction, stroke, or major bleed over 12 months. But, the real concerns are excess stent thrombosis if dual antiplatelet therapy is stopped too soon and excess major bleeding if dual antiplatelet therapy is continued too long. In OPTIMIZE there were 4 versus 1 stent thromboses contrasting with 3 versus 6 major bleeds in the 3- and 12-month dual antiplatelet therapy groups, respectively. Thus, this trial is too small to reach definitive conclusions. We await the findings of other trials, several comparing 6 months versus 12 months of dual antiplatelet therapy in over 11,000 total patients receiving a newer generation of stents.

THE IMPORTANCE OF TRIALS INTO ALTERNATIVE TREATMENT STRATEGIES

In practical terms, RCTs comparing different drugs or devices (e.g., stents) are often relatively easy to undertake, but many important therapeutic questions are more fundamental in that they involve a choice of substantially different alternative strategies in patient management. Here, the challenge is to conduct RCTs of sufficient quality, size, and representativeness. A common difficulty is to successfully recruit sufficient patients into such strategic trials. Each patient is faced with 2 markedly different therapeutic options (e.g., surgery vs. some alternative management, invasive vs. conservative management). Therefore, eliciting patient consent to be randomized is more of a challenge. Also, recruiting sufficient investigators in a state of clinical equipoise to accept randomization is no easy matter, especially when it requires collaboration across different specialties, for example, surgeons and interventional cardiologists.

Consider the case of multivessel primary PCI in STEMI patients. The PRAMI (Preventive Angioplasty in Myocardial Infarction Trial) (46) identified on angiogram those patients to be treated by primary PCI who also had 1 or more other significantly stenosed arteries in addition to the culprit lesion causing the myocardial infarction. Such patients were then randomized to either stenting of the culprit lesion only or to preventive angioplasty, whereby stents were also placed in other stenosed arteries. This multicenter U.K. trial took almost 5 years to randomize 465 patients, reflecting recruitment difficulties in such pragmatic trials.

Nevertheless, the interim results showed a marked treatment benefit of multivessel primary PCI, and the Data and Safety Monitoring Board recommended that the PRAMI trial be stopped early for superiority. Over a mean 23-month follow-up, the composite primary endpoint (refractory angina, nonfatal myocardial infarction, and cardiac death) had a hazard ratio of 0.35 (95% CI: 0.21 to 0.58; $p < 0.001$) in favor of multivessel versus culprit-vessel-only primary PCI. Findings for all 3 components of this primary endpoint were similar, and there was a rapid divergence in risk soon after randomization.

There has been much debate over these findings. The apparently huge treatment benefit is based on rather few primary events (53 vs. 21), and so one feels that the 65% reduction in hazard is too good to be true. Also, the trial was terminated early, perhaps on a "random high," which may exaggerate the true benefit. The trial was not blinded, which may contribute to a potential bias. So, these apparently impressive findings are, perhaps, not sufficient to justify a radical change in the routine management of STEMI patients with additional stenosed lesions, but are a motivation for other RCTs of preventive angioplasty to be undertaken.

The COMPLETE (Complete vs Culprit-only Revascularisation to Treat Multi-vessel Disease after Primary PCI for STEMI) trial (47) is also randomizing patients to revascularization of culprit lesion only versus a strategy of complete revascularization using drug-eluting stents of all suitable noninfarct-related

lesions. The latter will be staged during the same hospitalization period, rather than simultaneous with the culprit lesion PCI as was done in the PRAMI trial. The primary endpoint is restricted to CV death or new myocardial infarction over a minimum 2 years of follow-up, and hence, the intended sample size is 3,900 patients. The trial has recently started, and results are expected in the year 2020. That is a long time ahead, and one wonders (as in all trials of lengthy recruitment) whether changes in background therapies in the interim could negate the generalization of trial results to clinical practice.

RCTs VERSUS OBSERVATIONAL REGISTRIES: WHAT (NOT) TO BELIEVE?

Whether patients with multivessel coronary disease should receive coronary artery bypass graft (CABG) or PCI is a challenging question that has stimulated clinical trials research for over 20 years. The need for reintervention and the risk of myocardial infarction is higher after PCI, but the short-term risk of stroke is higher after CABG, which is, of course, more invasive and incapacitating for a few weeks. But, the key issue is whether there is a difference in patient survival. **Table 1** summarizes the variety of evidence available.

A meta-analysis of 10 older trials of CABG versus PCI before the use of drug-eluting stents (48) provided data on 7,812 patients over a median 5.9 years follow-up. Overall, there was a hint of a survival benefit for CABG: hazard ratio: 0.91 (p = 0.12). This became clearer in the subgroup of 1,233 diabetic patients (hazard ratio: 0.70) compared with nondiabetic subjects (hazard ratio: 0.98), interaction p = 0.014.

However, this needs updating to reflect the modern use of drug-eluting stents. Thus far, there are 2 such published trials: SYNTAX (SYNergy Between PCI With TAXUS and Cardiac Surgery) (49) in 1,800 patients with 3-vessel or left main disease, and FREEDOM (Future Revascularization Evaluation in Patients With Diabetes Mellitus: Optimal Management of Multivessel Disease) (50) in 1,900 diabetic patients with multivessel disease. Both trials show 5-year mortality trends favoring CABG: SYNTAX 11.4% versus 13.9% (p = 0.10); and FREEDOM 10.9% versus 16.3% (p = 0.049), but this leaves considerable uncertainty with wide CIs around the effect estimates. We await findings from the EXCEL (Evaluation of Xience Prime or Xience V versus Coronary Artery Bypass Surgery for Effectiveness of Left Main Revascularization) trial (51), which, although intended to recruit 2,600 patients, has, at the sponsor's request, recently curtailed recruitment at 1,905 patients, with primary results anticipated in 2016. Note that no trial of CABG versus PCI has ever recruited more than 2,000 patients. Such trials demand collaboration between cardiac surgeons and interventional cardiologists in each center regarding who is eligible for either procedure, and given the contrasting nature of the 2 procedures, the approach to eliciting informed patient consent is a challenge. Hence, there is always valid concern as to whether the patients in these trials are sufficiently representative of the broader population of patients in routine practice. A few trials (e.g., BARI [Bypass Angioplasty Revascularization Investigation]) (52) have included registries of nonrandomized patients to address this issue. Overall, limited size and doubts regarding their generalizability are often present when interpreting such trials of alternative strategies.

Comparative effectiveness studies based on observational registries provide a very different approach and show mortality differences between CABG and PCI in patients with multivessel disease. The ASCERT (ACCF and STS Database Collaboration on the Comparative Effectiveness of Revascularization Strategies) registry (53) compared 86,244 CABG patients with 103,549 PCI patients age ≥65 years over a median 2.67 years of follow-up. Adjustment for potential confounders was done using a propensity score with inverse probability weighting, giving adjusted 4-year mortality rates of 16.4% versus 20.8%, hazard ratio: 0.79 (95% CI: 0.76 to 0.82; p < 0.00000001). An alternative approach by Hlatky et al. (54) studied Medicare beneficiaries

TABLE 1 Summary of 5-Year Mortality in Both RCTs and Registries Comparing CABG and PCI

		5-Year Mortality		
	Number of Patients	CABG	PCI	p Value
10 older trials				
All patients	7,812	8.4	10.0	0.12
Diabetic	1,233	12.3	20.0	Interaction
Nondiabetic	6,561	7.5	8.1	0.014
SYNTAX	1,800	11.4	13.9	0.10
FREEDOM all diabetic	1,900	10.9	16.3	0.049
ASCERT registry†	189,793	16.4*	20.8*	0.000001
CMS registry†				
All patients	210,312	25.9	28.1	0.000001
Diabetic		30.4	33.9	Interaction
Nondiabetic		23.8	25.3	0.002

Values are % unless otherwise indicated. *4-year mortality. †Mortality rates are higher in the 2 registries, mainly because they only include patients age >65 years.

CABG = coronary artery bypass graft; PCI = percutaneous coronary intervention; RCT = randomized controlled trial.

TOPIC	MAIN EXAMPLES
Problems with medical device trials	Renal denervation in resistant hypertension
Balancing commercial interests and public health needs	Bivalirudin in STEMI patients
Lack of head-to-head trials of new drugs	Anticoagulants in atrial fibrillation
Need for better handling of drug safety concerns	Anti-diabetic agents
Credibility of medical journal publications	COREVALVE, BEAUTIFUL, CURRENT OASIS 7 and SHIFT trials
Geographic heterogeneity in trial findings	TOPCAT and PLATO trials
The role of personalized (stratified) medicine	EMPHASIS and TRITON-TIMI 38 trials
The need for new trials of old drugs	Beta blockers pre PPCI and during non-cardiac surgery
The need for trials of treatment withdrawal	Beta blockers in stable CHD and DAPT in ACS patients
Pragmatic trials of treatment strategies	Multivessel vs culprit-vessel PPCI
Credibility of non-randomised registries	PCI vs CABG in multi-vessel disease

CENTRAL ILLUSTRATION Review of Clinical Trials Research

ACS = acute coronary syndrome(s); CABG = coronary artery bypass graft; CHD = coronary heart disease; DAPT = dual antiplatelet therapy; PCI = percutaneous coronary intervention; PPCI = primary percutaneous coronary intervention; STEMI = ST-segment elevation myocardial infarction.

age ≥66 years and derived 105,156 propensity score-matched pairs of CABG/PCI patients with a median 4.3 years follow-up. Five-year mortality rates were 25.9% versus 28.1%, hazard ratio: 0.92 (95% CI: 0.90 to 0.95; $p < 0.00000001$). Both studies suggest a very highly-significant survival benefit for CABG, but the sizes of these effects differ enormously: comparing the 2 hazard ratios, 0.79 and 0.92, the interaction test has $p < 0.01$. So, we have very precise estimates from 2 very large registries that differ substantially in magnitude. Can either of them be trusted?

The problem with all such comparative effectiveness studies is that, in the absence of random allocation, there is potentially strong selection bias as to which patients get which treatments (55,56). Although propensity score or covariate adjustment methods can adjust for measured confounders, it would be naive to assume that potential residual confounding (i.e., bias) does not exist. One cannot expect the database to capture all of the complex reasons why one patient gets CABG and another gets PCI.

Of interest are the exploratory subgroup analyses for diabetes in Hlatky et al. (54). For diabetic and nondiabetic patients, the hazard ratios were 0.88 and 0.95, respectively (interaction $p = 0.002$). This is broadly compatible with findings from the meta-analysis of older trials mentioned previously. Perhaps the survival benefit of CABG is truly more pronounced in diabetic patients, although with modern-day practice (drug-eluting stents), the gap may have narrowed. The issue may be clouded further by the emergence of an ever-newer generation of drug-eluting stents, as in EXCEL, but not SYNTAX and FREEDOM (57). In any field characterized by rapid technological changes and in which trials require prolonged follow-up, perhaps every trial is vulnerable to the accusation of "obsolescence" by the time it is published, but this does not obviate the rationale for such trials.

Although the data from registries may be more applicable to clinical practice at large, they are subject to selection bias. RCTs are, however, subject to "entry bias," in that patients are highly selected on the basis of clinical equipoise with regard to the findings on angiography. Registries and trials may complement each other and emphasize differences between the process of randomization versus the selection of a therapeutic strategy based upon physician judgment and patient preference (52). Note that in the EAST (Emory Angioplasty versus Surgery Trial) trial of percutaneous transluminal coronary angioplasty versus coronary bypass surgery, college-educated patients were less likely to accept randomization (58).

We need to be continually cautious of making therapeutic claims from observational data. For instance, a meta-analysis of 21 randomized trials (59) comparing drug-eluting and bare-metal stents in 8,867 patients with mean 2.9 years follow-up found no difference in mortality: hazard ratio: 0.97 ($p = 0.72$). The corresponding meta-analysis of 31 observational registries had far more patients, 169,595 with mean 2.5 years follow-up, and revealed significantly lower mortality after a drug-eluting stent: hazard ratio: 0.81 ($p < 0.00001$).

All observational studies are vulnerable to the influence of confounding variables. In this respect, randomized controlled trials are crucial. This is exemplified by the epidemiologic data that demonstrated a protective effect of estrogen on the heart and bone (60). In contrast, data from the Women's Health Initiative (61,62), a set of 2 hormone therapy trials in healthy post-menopausal women, showed a

number of adverse outcomes, including an excess risk of coronary heart disease, stroke, venous thromboembolism, and breast cancer. Similar examples of epidemiological evidence and biological plausibility that did not withstand the rigorous scrutiny of RCTs are provided by studies of vitamin C, vitamin E, and beta-carotene.

Thus, observational research has its uses, but can sometimes be prone to make false claims on the relative efficacy and relative safety of alternative treatments.

CONCLUSIONS

This paper is not meant to be a comprehensive review of the whole field of clinical trials research applied to cardiology. Rather, it has attempted to highlight several important controversies and challenges. The **Central Illustration** summarizes the issues that were tackled. Although we have much to celebrate in the past achievements of cardiology trials in generating many important treatment advances, there remains considerable room for improvement if future RCTs are truly to achieve their full potential in enhancing public health and optimizing cardiological patient care.

REPRINT REQUESTS AND CORRESPONDENCE: Prof. Stuart J. Pocock, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom. E-mail: stuart.pocock@lshtm.ac.uk.

REFERENCES

1. Bhatt DL, Kandzari DE, O'Neill WW, et al., for the SYMPLICITY HTN-3 Investigators. A controlled trial of renal denervation for resistant hypertension. N Engl J Med 2014;370:1393–401.

2. Krum H, Schlaich M, Whitbourn R, et al. Catheter-based renal sympathetic denervation for resistant hypertension: a multicentre safety and proof-of-principle cohort study. Lancet 2009;373:1275–81.

3. Symplicity HTN-1 Investigators. Catheter-based renal sympathetic denervation for resistant hypertension: durability of blood pressure reduction out to 24 months. Hypertension 2011;57:911–7.

4. Esler MD, Krum H, Sobotka PA, et al., for the Symplicity HTN-2 Investigators. Renal sympathetic denervation in patients with treatment-resistant hypertension (The Symplicity HTN-2 Trial): a randomised controlled trial. Lancet 2010;376:1903–9.

5. Sud S, Massel D, Klein GJ, et al. The expectation effect and cardiac pacing for refractory vasovagal syncope. Am J Med 2007;120:54–62.

6. Kramer DB, Xu S, Kesselheim AS. Regulation of medical devices in the United States and European Union. N Engl J Med 2012;366:848–55.

7. Sahzad A, Kemp I, Mars C, et al., for the HEAT-PPCI Trial Investigators. Unfractionated heparin versus bivalirudin in primary percutaneous coronary intervention (HEAT-PPCI): an open-label, single centre, randomised controlled trial. Lancet 2014 July 4 [E-pub ahead of print].

8. Stone GW, McLaurin BT, Cox DA, et al., for the ACUITY Investigators. Bivalirudin for patients with acute coronary syndromes. N Engl J Med 2006;355:2203–16.

9. Stone GW, Lansky AJ, Pocock SJ, et al., for the HORIZONS-AMI Trial Investigators. Paclitaxel-eluting stents versus bare-metal stents in acute myocardial infarction. N Engl J Med 2009;360:1946–59.

10. Steg PG, van't Hof A, Hamm CW, et al., for the EUROMAX Investigators. Bivalirudin started during emergency transport for primary PCI. N Engl J Med 2013;369:2207–17.

11. DeMets DL, Califf RM. A historical perspective on clinical trials innovation and leadership: where have the academics gone? JAMA 2011;305:713–4.

12. Swedberg K, Komajda M, Bohm M, et al. Ivabradine and outcomes in chronic heart failure (SHIFT): a randomised placebo-controlled study. Lancet 2010;376:875–85.

13. Connolly SJ, Ezekowitz MD, Yusuf S, et al., for the RE-LY Steering Committee and Investigators. Dabigatran versus warfarin in patients with atrial fibrillation. N Engl J Med 2009;361:1139–51.

14. Patel MR, Mahaffey KW, Garg J, et al., for the ROCKET AF Investigators. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. N Engl J Med 2011;365:883–91.

15. Granger CB, Alexander JH, McMurray JJ, et al., for the ARISTOTLE Committees and Investigators. Apixaban versus warfarin in patients with atrial fibrillation. N Engl J Med 2011;365:981–92.

16. Giugliano RP, Ruff CT, Braunwald E, et al., for the ENGAGE AF-TIMI 48 Investigators. Edoxaban versus warfarin in patients with atrial fibrillation. N Engl J Med 2013;369:2093–104.

17. Glenny AM, Altman DG, Song F, et al., for the International Stroke Trial Collaborative Group. Indirect comparisons of competing interventions. Health Technol Assess 2005;9:1–134, iii–iv.

18. Schulz S, Angiolillo DJ, Antoniucci D, et al., for the Intracoronary Stenting and Antithrombotic Regimen: Rapid Early Action for Coronary Treatment (ISAR-REACT) 5 Trial Investigators. Randomized comparison of ticagrelor versus prasugrel in patients with acute coronary syndrome and planned invasive strategy—design and rationale of the iNtracoronary Stenting and Antithrombotic Regimen: Rapid Early Action for Coronary Treatment (ISAR-REACT) 5 trial. J Cardiovasc Transl Res 2014;7:91–100.

19. Alexander JH, Lopes RD, James S, et al., for the APPRAISE-2 Investigators. Apixaban with antiplatelet therapy after acute coronary syndrome. N Engl J Med 2011;365:699–708.

20. Mega JL, Braunwald E, Wiviott SD, et al., for the ATLAS ACS 2-TIMI 51 Investigators. Rivaroxaban in patients with a recent acute coronary syndrome. N Engl J Med 2012;366:9–19.

21. Cohn JN, Pfeffer MA, Rouleau J, et al., for the MOXCON Investigators. Adverse mortality effect of central sympathetic inhibition with sustained-release moxonidine in patients with heart failure (MOXCON). Eur J Heart Fail 2003;5:659–67.

22. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. N Engl J Med 2007;356:2457–71.

23. Home PD, Pocock SJ, Beck-Nielsen H, et al., for the RECORD Study Team. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. Lancet 2009;373:2125–35.

24. Food and Drug Administration. Diabetes mellitus: evaluating cardiovascular risks of new antidiabetic therapies. 2008. Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071627.pdf. Accessed August 6, 2014.

25. Scirica BM, Bhatt DL, Braunwald E, et al., for the SAVOR-TIMI 53 Steering Committee and Investigators. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. N Engl J Med 2013;369:1317–26.

26. White WB, Cannon CP, Heller SR, et al., for the EXAMINE Investigators. Alogliptin after acute coronary syndrome in patients with type 2 diabetes. N Engl J Med 2013;369:1327–35.

27. Bethel AM, Green J, Califf RM, et al. Rationale and design of the Trial Evaluating Cardiovascular Outcomes with Sitagliptin (TECOS) (abstr). Diabetes 2009;58 Suppl 1:A587.

28. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010;152:726–32.

29. Adams DH, Popma JJ, Reardon MJ, et al., for the U.S. CoreValve Clinical Investigators. Transcatheter aortic-valve replacement with a

self-expanding prosthesis. N Engl J Med 2014;370:1790–8.

30. Fox K, Ford I, Steg PG, Tendera M, et al., for the BEAUTIFUL Investigators. Ivabradine for patients with stable coronary artery disease and left-ventricular systolic dysfunction (BEAUTIFUL): a randomised, double-blind, placebo-controlled trial. Lancet 2008;372:807–16.

31. Mehta SR, Tanguay JF, Eikelboom JW, et al., for the CURRENT-OASIS 7 Trial Investigators. Double-dose versus standard-dose clopidogrel and high-dose versus low-dose aspirin in individuals undergoing percutaneous coronary intervention for acute coronary syndromes (CURRENT-OASIS 7): a randomised factorial trial. Lancet 2010;376:1233–43.

32. Mehta SR, Bassand JP, Chrolavicius S, et al., for the CURRENT-OASIS 7 Investigators. Dose comparisons of clopidogrel and aspirin in acute coronary syndromes. N Engl J Med 2010;363:930–42.

33. Pitt B, Pfeffer MA, Assmann SF, et al., for the TOPCAT Investigators. Spironolactone for heart failure with preserved ejection fraction. N Engl J Med 2014;370:1383–92.

34. Wallentin L, Becker RC, Budaj A, et al., for the PLATO Investigators. Ticagrelor versus clopidogrel in patients with acute coronary syndromes. N Engl J Med 2009;361:1045–57.

35. Pocock S, Calvo G, Marrugat J, et al. International differences in treatment effect: do they really exist and why? Eur Heart J 2013;34:1846–52.

36. Collier TJ, Pocock SJ, McMurray JJV, et al. The impact of eplerenone at different levels of risk in patients with systolic heart failure and mild symptoms. Eur Heart J 2013;34:2823–9.

37. Wiviott SD, Braunwald E, McCabe CH, et al., for the TRITON-TIMI 38 Investigators. Prasugrel versus clopidogrel in patients with acute coronary syndromes. N Engl J Med 2007;357:2001–15.

38. Salisbury AC, Wang K, Cohen DJ, et al. Selecting antiplatelet therapy at the time of percutaneous intervention for an acute coronary syndrome: weighing the benefits and risks of prasugrel versus clopidogrel. Circ Cardiovasc Qual Outcomes 2013;6:27–34.

39. Eikelboom JW, Connolly SJ, Hart RG, et al. Balancing the benefits and risks of 2 doses of dabigatran compared with warfarin in atrial fibrillation. J Am Coll Cardiol 2013;62:900–8.

40. U.S. Department of Health and Human Services. U.S. Food and Drug Administration. Paving the way for personalized medicine. 2013. Available at: http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/PersonalizedMedicine/UCM372421.pdf. Accessed August 6, 2014.

41. Ibanez B, Macaya C, Sanchez-Brunete V, et al. Effect of early metoprolol on infarct size in ST-segment-elevation myocardial infarction patients undergoing primary percutaneous coronary intervention: the Effect of Metoprolol in Cardioprotection During an Acute Myocardial Infarction (METOCARD-CNIC) trial. Circulation 2013;128:1495–503.

42. Devereaux PJ, Yang H, Yusuf S, et al., for the POISE Study Group. Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. Lancet 2008;371:1839–47.

43. Authors/Task Force Members, Kristensen SD, Knuuti J, Saraste A, et al. 2014 ESC/ESA Guidelines on non-cardiac surgery: cardiovascular assessment and management: The Joint Task Force on non-cardiac surgery: cardiovascular assessment and management of the European Society of Cardiology (ESC) and the European Society of Anaesthesiology (ESA). Eur Heart J 2014;35:2383–431.

44. Mauri L, Kereiakes DJ, Normand SL, et al. Rationale and design of the dual antiplatelet therapy study, a prospective, multicenter, randomized, double-blind trial to assess the effectiveness and safety of 12 versus 30 months of dual antiplatelet therapy in subjects undergoing percutaneous coronary intervention with either drug-eluting stent or bare metal stent placement for the treatment of coronary artery lesions. Am Heart J 2010;160:1035–41, 1041.e1.

45. Feres F, Costa RA, Abizaid A, et al., for the OPTIMIZE Trial Investigators. Three vs twelve months of dual antiplatelet therapy after zotarolimus-eluting stents: the OPTIMIZE randomized trial. JAMA 2013;310:2510–22.

46. Wald DS, Morris JK, Wald NJ, et al., for the PRAMI Investigators. Randomized trial of preventive angioplasty in myocardial infarction. N Engl J Med 2013;369:1115–23.

47. ClinicalTrials.gov. Complete vs Culprit-only Revascularisation to Treat Multi-vessel Disease after Primary PCI for STEMI (COMPLETE). April 24, 2013. Available at: http://clinicaltrials.gov/show/NCT01740479. Accessed August 6, 2014.

48. Hlatky MA, Boothroyd DB, Bravata DM, et al. Coronary artery bypass surgery compared with percutaneous coronary interventions for multivessel disease: a collaborative analysis of individual patient data from ten randomised trials. Lancet 2009;373:1190–7.

49. Mohr FW, Morice MC, Kappetein AP, et al. Coronary artery bypass graft surgery versus percutaneous coronary intervention in patients with three-vessel disease and left main coronary disease: 5-year follow-up of the randomised, clinical SYNTAX trial. Lancet 2013;381:629–38.

50. Farkouh ME, Domanski M, Sleeper LA, et al., for the FREEDOM Trial Investigators. Strategies for multivessel revascularization in patients with diabetes. N Engl J Med 2012;367:2375–84.

51. Stone GW. Milestone for left main PCI: upcoming EXCEL trial 2008. 2008. Available at: http://www.summitmd.com/pdf/pdf/1168_EXCEL 3.pdf. Accessed August 6, 2014.

52. Detre KM, Guo P, Holubkov R, et al. Coronary revascularization in diabetic patients: a comparison of the randomized and observational components of the Bypass Angioplasty Revascularization Investigation (BARI). Circulation 1999;99:633–40.

53. Weintraub WS, Grau-Sepulveda MV, Weiss JM, et al. Comparative effectiveness of revascularization strategies. N Engl J Med 2012;366:1467–76.

54. Hlatky MA, Boothroyd DB, Baker L, et al. Comparative effectiveness of multivessel coronary bypass surgery and multivessel percutaneous coronary intervention: a cohort study. Ann Intern Med 2013;158:727–34.

55. Brown ML, Gersh BJ, Holmes DR, et al. From randomized trials to registry studies: translating data into clinical information. Nat Clin Pract Cardiovasc Med 2008;5:613–20.

56. Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? N Engl J Med 2000;342:1907–9.

57. Windecker S, Stortecky S, Stefanini GG, et al. Revascularisation versus medical treatment in patients with stable coronary artery disease: network meta-analysis. BMJ 2014;348:g3859.

58. King SB 3rd, Barnhart HX, Kosinski AS, et al. Angioplasty or surgery for multivessel coronary artery disease: comparison of eligible registry and randomized patients in the EAST trial and influence of treatment selection on outcomes. Emory Angioplasty versus Surgery Trial Investigators. Am J Cardiol 1997;79:1453–9.

59. Kirtane AJ, Gupta A, Iyengar S, et al. Safety and efficacy of drug-eluting and bare metal stents: comprehensive meta-analysis of randomized trials and observational studies. Circulation 2009;119:3198–206.

60. Mendelsohn ME, Karas RH. The protective effects of estrogen on the cardiovascular system. N Engl J Med 1999;340:1801–11.

61. Anderson GL, Limacher M, Assaf AR, et al., for the Women's Health Initiative Steering Committee. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. JAMA 2004;291:1701–12.

62. Rossouw JE, Anderson GL, Prentice RL, et al., for the Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. JAMA 2002;288:321–33.

KEY WORDS clinical trial, placebo effect, public health, random allocation, reporting, research design, statistical analysis