The NEW ENGLAND JOURNAL *of* MEDICINE

REVIEW ARTICLE

THE CHANGING FACE OF CLINICAL TRIALS
Jeffrey M. Drazen, M.D., David P. Harrington, Ph.D., John J.V. McMurray, M.D., James H. Ware, Ph.D.,
and Janet Woodcock, M.D., *Editors*

The Primary Outcome Is Positive — Is That Good Enough?

Stuart J. Pocock, Ph.D., and Gregg W. Stone, M.D.

THERE IS A NATURAL TENDENCY TO SIMPLIFY THE FINDINGS OF A CLINI-
cal trial into a binary conclusion: "Was there a positive outcome — or not?"
In order to address this question with some objectivity, attention is typi-
cally focused on whether the prespecified measure of success for the primary
outcome has been met — that is, whether a P value of less than 0.05 has been
achieved for the difference in treatments. In reality, a more nuanced interpretation
requires a thorough examination of the totality of the evidence, including second-
ary end points, safety issues, and the size and quality of the trial. In this article,
which focuses on the evaluation of "positive" studies — as in our previous article,[1]
which focused on the appraisal of "negative" studies — our intent is to facilitate
a more sophisticated and balanced interpretation of trial evidence. Again, we make
our points using examples from trials involving cardiovascular disease (our area
of expertise), but the messages can be easily applied to other subject areas.

From the Department of Medical Statis-
tics, London School of Hygiene and Trop-
ical Medicine, London (S.J.P.); and Colum-
bia University Medical Center, New York
Presbyterian Hospital, and the Cardio-
vascular Research Foundation — all in
New York (G.W.S.). Address reprint re-
quests to Dr. Pocock at the Department
of Medical Statistics, London School of
Hygiene and Tropical Medicine, Keppel St.,
London WC1E 7HT, United Kingdom; or
at stuart.pocock@lshtm.ac.uk.

This article was updated on September 8,
2016, at NEJM.org.

N Engl J Med 2016;375:971-9.
DOI: 10.1056/NEJMra1601511
Copyright © 2016 Massachusetts Medical Society.

KEY QUESTIONS WHEN THE PRIMARY OUTCOME IS POSITIVE

The achievement of statistical significance for the primary outcome is typically a
necessary prerequisite for the adoption of a new therapy, but it is not sufficient.
The totality of trial results will be scrutinized by numerous stakeholders, including
regulators, payers, journal editors and reviewers, clinical experts, guidelines com-
mittees, physicians, patients, and critics. The determination of whether the find-
ings provide evidence that is sufficient to modify medical practice requires in-depth
interpretation of the trial data and the results of earlier, related trials. Answering
the key questions discussed below (and listed in Table 1) may help to identify
which "positive" trials provide evidence that is sufficient to advance clinical practice.
In this regard, we also acknowledge that studies that unexpectedly reveal harm
(e.g., the CAST trial [see box for a list of the complete names of all trials men-
tioned in this article], which revealed that antiarrhythmic therapy was associated
with an increase in mortality[2]) may equally inform medical practice, although
here, too, close scrutiny of the validity of the results is warranted.

IS A P VALUE OF LESS THAN 0.05 GOOD ENOUGH?

A P value of 0.05 carries a 5% risk of a false positive result (i.e., there is no true
difference between treatments). If a trial is meant to provide proof of a genuine
treatment difference beyond reasonable doubt, a much smaller P value — say,
P<0.001 — is required.[3] For instance, the PARADIGM-HF trial[4] of sacubitril–
valsartan versus enalapril in patients with heart failure showed overwhelming ben-
efit (P<0.00001) with respect to the composite primary outcome of cardiovascular

N ENGL J MED 375;10 NEJM.ORG SEPTEMBER 8, 2016

The New England Journal of Medicine
Downloaded from nejm.org at MUSC Library on September 22, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

971

Table 1. Key Questions to Ask When the Primary Outcome Is Positive.

Does a P value of <0.05 provide strong enough evidence?

What is the magnitude of the treatment benefit?

Is the primary outcome clinically important (and internally consistent)?

Are secondary outcomes supportive?

Are the principal findings consistent across important subgroups?

Is the trial large enough to be convincing?

Was the trial stopped early?

Do concerns about safety counterbalance positive efficacy?

Is the efficacy–safety balance patient-specific?

Are there flaws in trial design and conduct?

Do the findings apply to my patients?

Trial Names.

ACCORD: Action to Control Cardiovascular Risk in Diabetes.

ATLAS ACS 2–TIMI 51: Anti-Xa Therapy to Lower Cardiovascular Events in Addition to Standard Therapy in Subjects with Acute Coronary Syndrome–Thrombolysis in Myocardial Infarction 51.

CAST: Cardiac Arrhythmia Suppression Trial.

COMPLETE: Complete vs. Culprit-Only Revascularization to Treat Multi-vessel Disease after Primary PCI for STEMI.

DAPT: Dual Antiplatelet Therapy.

EMPA-REG OUTCOME: Empagliflozin Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients.

EXPEDITION: Na+/H+ Exchange Inhibition to Prevent Coronary Events in Acute Cardiac Condition.

FAME 2: Fractional Flow Reserve versus Angiography for Multivessel Evaluation.

FREEDOM: Future Revascularization Evaluation in Patients with Diabetes Mellitus: Optimal Management of Multivessel Disease.

IMPROVE-IT: Improved Reduction of Outcomes: Vytorin Efficacy International Trial.

LIDO: Levosimendan Infusion versus Dobutamine.

PARADIGM-HF: Prospective Comparison of ARNI (Angiotensin Receptor–Neprilysin Inhibitor) with ACEI (Angiotensin-Converting–Enzyme Inhibitor) to Determine Impact on Global Mortality and Morbidity in Heart Failure Trial.

PLATO: Study of Platelet Inhibition and Patient Outcomes.

PRAMI: Preventive Angioplasty in Acute Myocardial Infarction.

RITA-3: Randomized Intervention Trial of Unstable Angina.

SAINT I and II: Stroke–Acute Ischemic NXY Treatment.

SPRINT: Systolic Blood Pressure Intervention Trial.

SURVIVE: Survival of Patients with Acute Heart Failure in Need of Intravenous Inotropic Support.

SYMPLICITY HTN-2 and -3: Renal Denervation in Patients with Uncontrolled Hypertension.

SYNTAX: Synergy between Percutaneous Coronary Intervention with Taxus and Cardiac Surgery.

TAPAS: Thrombus Aspiration during Percutaneous Coronary Intervention in Acute Myocardial Infarction Study.

death or hospitalization for heart failure, which justified regulatory approval and the clinical adoption of sacubitril–valsartan. In contrast, in the SAINT I trial[5] of NXY-059, a free-radical–trapping agent, versus placebo for the treatment of acute ischemic stroke, the P value for the primary outcome of disability at 90 days was 0.038 — not strong evidence of efficacy. A second, larger trial (SAINT II[6]) was needed and revealed no significant effect (P=0.33), which prompted the authors to conclude that "NXY-059 is ineffective for the treatment of acute ischemic stroke."

WHAT IS THE MAGNITUDE OF TREATMENT BENEFIT?

Beyond statistical significance, a treatment difference needs to be clinically meaningful — that is, large enough to matter. This determination requires examination of the treatment effect on both a relative scale (e.g., by calculation of the relative risk or the hazard ratio) and an absolute scale (e.g., by calculation of the differences in the rates of events during follow-up and in the number needed to treat). In addition, the extent of uncertainty in the estimated effect should be considered by examining the 95% confidence interval. For instance, if the P value is close to 0.05, the confidence interval will range from almost no effect to an upper boundary that is considerably larger than the point estimate.

An illustration of these concepts is provided in the IMPROVE-IT trial,[7] in which ezetimibe was compared with placebo in patients with acute coronary syndromes who were being treated with simvastatin; the hazard ratio for the composite primary outcome of cardiovascular death, myocardial infarction, unstable angina, revascularization, or stroke was 0.94 (95% confidence interval [CI], 0.89 to 0.98; P=0.016). The 7-year primary event rates were 32.7% with ezetimibe versus 34.7% with placebo (Fig. 1), a difference of 2 percentage points, with a 95% confidence interval that ranged from close to 0 to 4 percentage points. Although the findings for this trial were described as "positive," one might question whether the benefit of ezetimibe is large enough to warrant its cost and potential complications. An advisory panel from the Food and Drug Administration (FDA) recommended against expanding the ezetimibe label to include an indication for a reduction in cardiovascular events.[8]

IS THE PRIMARY OUTCOME CLINICALLY IMPORTANT (AND INTERNALLY CONSISTENT)?
Surrogate Outcomes

Although phase 3 trials are usually powered to achieve clinically relevant outcomes, for some diseases a surrogate primary outcome measure has been accepted (e.g., a reduction in glycated hemoglobin levels as an indication of antiglyce-

mic efficacy in patients with diabetes). However, the findings of some large-scale trials have raised questions about the wisdom of such reliance on surrogate markers.[9] For instance, in the ACCORD trial,[10] intensive therapy resulted in markedly lower glycated hemoglobin levels than standard therapy, but the rate of cardiovascular events was not significantly lower, and mortality was higher. Similarly, in the LIDO trial, levosimendan resulted in greater hemodynamic improvement (the primary outcome) than dobutamine in patients with acute heart failure,[11] which resulted in the regulatory approval of levosimendan in many countries. However, SURVIVE, the larger, subsequent trial of levosimendan versus dobutamine,[12] showed no evidence of a treatment benefit for the primary outcome — 180-day mortality (P=0.40) — and levosimendan was not approved by the FDA.

Composite Outcomes

Positive composite primary outcomes must be carefully inspected to determine which components are driving the result. For instance, in the RITA-3 trial,[13] which assessed the effects of interventional versus conservative management in patients with acute coronary syndromes, fewer patients in the intervention group had a composite primary outcome event of death, myocardial infarction, or refractory angina at 4 months (9.6% vs. 14.5%, P=0.001). When the trial findings were presented at a European Society of Cardiology Congress, a newsletter headline read "RITA-3: First Proof Intervention Saves Lives" — an incorrect interpretation, since this finding was driven by a halving of the rate of refractory angina, with no evidence of a difference in the rates of death or myocardial infarction in the short term. At the time, the question of whether the data justified the use of a routine invasive strategy was debatable, given its up-front risks and costs. Fortunately, a 5-year follow-up study[14] revealed a 22% lower rate of death or myocardial infarction in the intervention group than in the conservative management group (P=0.04), and subsequent meta-analyses[15,16] have supported the use of an early interventional approach in patients with acute coronary syndromes to improve prognosis.

More strikingly, the EXPEDITION trial[17] of cariporide versus placebo in high-risk patients undergoing coronary-artery bypass grafting (CABG) had a very positive result (P=0.0002) for the
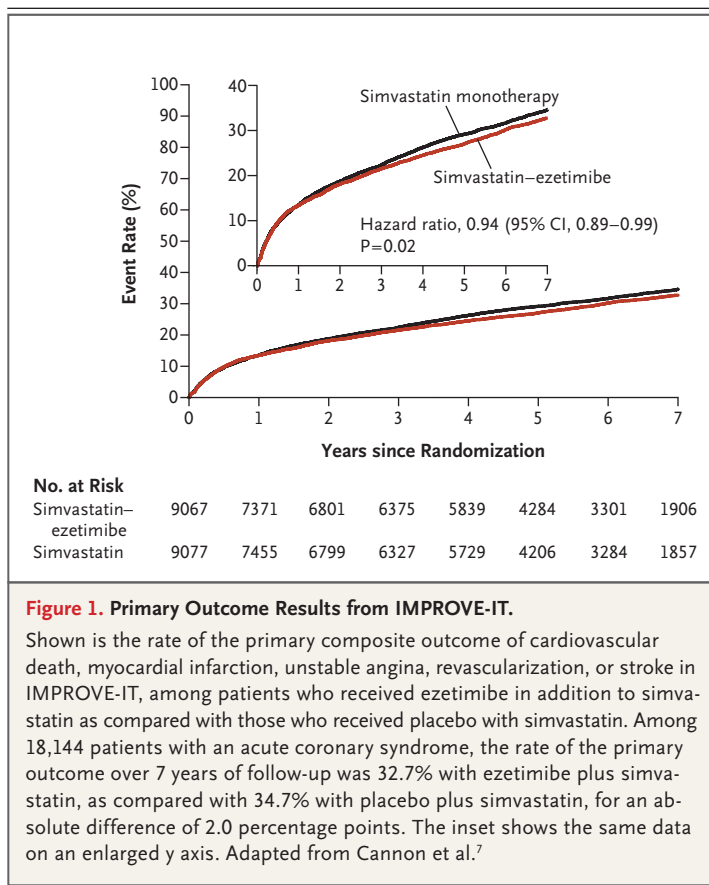

Figure 1. Primary Outcome Results from IMPROVE-IT.
Shown is the rate of the primary composite outcome of cardiovascular death, myocardial infarction, unstable angina, revascularization, or stroke in IMPROVE-IT, among patients who received ezetimibe in addition to simvastatin as compared with those who received placebo with simvastatin. Among 18,144 patients with an acute coronary syndrome, the rate of the primary outcome over 7 years of follow-up was 32.7% with ezetimibe plus simvastatin, as compared with 34.7% with placebo plus simvastatin, for an absolute difference of 2.0 percentage points. The inset shows the same data on an enlarged y axis. Adapted from Cannon et al.[7]

primary composite outcome of death or myocardial infarction. However, this outcome was driven by a reduction in myocardial infarction (P=0.000005), whereas mortality was higher with cariporide (P=0.02), as was the rate of cerebrovascular events (P<0.001). These findings led to the abandonment of cariporide for this indication.

ARE SECONDARY OUTCOMES SUPPORTIVE?

Confidence in the overall "positivity" of a trial is enhanced if prespecified secondary outcomes also show a treatment benefit. Conversely, if secondary outcomes show no hint of benefit, doubts will materialize. For instance, in the SAINT I trial[5] of NXY-059 in acute ischemic stroke, no evidence of benefit existed for two key secondary outcomes — scores on the National Institutes of Health Stroke Scale and the Barthel Index. This absence of evidence enhanced suspicion regarding the "positive" primary outcome, a suspicion that was reinforced by the negative result in the sequel trial, SAINT II.[6]

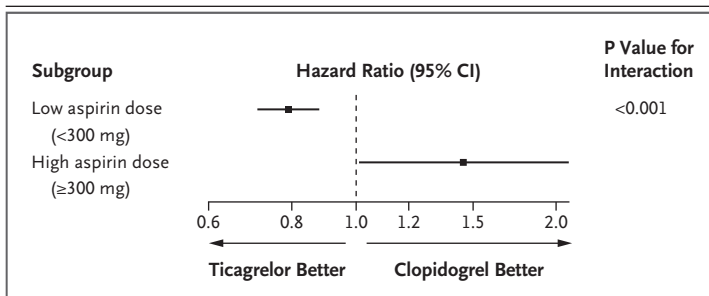In contrast, in the EMPA-REG OUTCOME trial[18]

Figure 2. Apparent Interaction between a Maintenance Dose of Aspirin and Randomization to Ticagrelor or Clopidogrel and the Primary Efficacy End Point from the PLATO Trial.

In the PLATO trial, which involved 18,624 patients who presented with an acute coronary syndrome, ticagrelor was more effective than clopidogrel with respect to the primary composite outcome of cardiovascular death, myocardial infarction, or stroke among patients who were on a low maintenance dose of aspirin (<300 mg) but not among those who were on a high maintenance dose of aspirin (≥300 mg), a qualitative interaction that was statistically significant. Data are from Carroll et al.[22]

of empagliflozin versus placebo in type 2 diabetes, the benefit of empagliflozin with respect to the composite primary outcome (cardiovascular death, myocardial infarction, or stroke) was of borderline significance, with a hazard ratio of 0.86 (95% CI, 0.74 to 0.99; P=0.04). However, this finding was driven by a robustly lower rate of cardiovascular death (hazard ratio, 0.62; 95% CI, 0.49 to 0.77; P<0.001) and was reinforced by similar findings regarding all-cause death (P<0.001) and hospitalizations for heart failure (P=0.002). Thus, the effect of empagliflozin was observed mainly in the secondary outcomes, although the positive finding for the primary outcome imparted initial credibility.

ARE FINDINGS CONSISTENT ACROSS IMPORTANT SUBGROUPS?

Relative treatment effects may vary according to patient characteristics. Alternatively, a consistent relative treatment effect across all patient types may be observed, but certain high-risk subgroups may have greater absolute benefits, as has been seen with statins in patients with multiple cardiac risk factors.[19] Long-term statin use in primary prevention is thus commonly confined to patients with a sufficiently high baseline risk.

More challenging is the situation in which subgroup analyses in a "positive" trial identify patients who do not appear to benefit from the new treatment. Caution is warranted, since spurious findings can arise when multiple sub-

groups are analyzed.[20] Nonetheless, protecting such patients from a treatment that appears to be ineffective (or harmful) may be warranted, depending on the strength of the statistical interaction and its biologic plausibility.

For instance, in the PLATO trial[21,22] involving patients with an acute coronary syndrome, the risk of the composite primary outcome of cardiovascular death, myocardial infarction, or stroke was 16% lower with ticagrelor than with clopidogrel in the overall study population (P<0.001). However, among patients receiving a high maintenance dose of aspirin, the risk was 45% higher with ticagrelor than with clopidogrel, whereas among patients receiving a low maintenance dose, ticagrelor was associated with a 21% lower risk (P=0.0006 for the interaction) (Fig. 2). Although the validity of this observation is still disputed (it arose from numerous exploratory subgroup analyses and lacks obvious biologic plausibility), the FDA issued a warning that a maintenance dose of more than 100 mg of aspirin reduces the effectiveness of ticagrelor and should be avoided.

IS THE TRIAL LARGE ENOUGH TO BE CONVINCING?

When a small clinical trial achieves statistical significance for its primary outcome, cautious interpretation is warranted. Small trials lack power, so positive treatment effects are susceptible to exaggeration, and false positives may occur.

For instance, in a trial of N-acetylcysteine versus placebo to prevent nephropathy induced by radiocontrast agents,[23] 1 of 41 patients receiving N-acetylcysteine had a primary outcome event, as compared with 9 of 42 patients who were receiving placebo, resulting in a relative risk with N-acetylcysteine of 0.10 (95% CI, 0.02 to 0.90; P=0.01). On the basis of this small trial, the stated conclusion that N-acetylcysteine is "an effective means of preventing renal damage" is too strong; a more appropriate statement is that N-acetylcysteine "may be effective." Such a conclusion would motivate the conduct of a larger, more definitive trial. Unfortunately, a subsequent meta-analysis[24] of 10 randomized trials (1916 patients) reported that the evidence was too weak and heterogeneous to support use of N-acetylcysteine for this indication.

Similarly, in the PRAMI trial, 465 patients with ST-segment elevation myocardial infarction (STEMI) and multivessel disease who were un-

dergoing primary percutaneous coronary intervention (PCI) were randomly assigned to receive preventive angioplasty (complete revascularization during the index procedure) or initial angioplasty of the infarct artery only.[25] The hazard ratio for the composite primary outcome (refractory angina, myocardial infarction, or cardiac death) with preventive angioplasty versus angioplasty of the infarct artery only was 0.35 (95% CI, 0.21 to 0.58; P<0.001), with similarly lower risks for each of the three components. This controversial finding was based on relatively few primary events (21 in the intervention group vs. 53 in the group receiving standard care) and selective enrollment (recruitment took 5 years and was stopped early), giving the impression that the 65% reduction in hazard was too good to be true. Two subsequent, similarly sized trials[26,27] showed mixed results. Thus, more evidence is needed to justify such a radical change in STEMI management, and the results of the ongoing, large-scale COMPLETE trial[28] are awaited with interest.

WAS THE TRIAL STOPPED EARLY?

Sometimes a trial is stopped early because interim results show strong evidence of treatment superiority, which is often a newsworthy event. Unfortunately, this practice tends to exaggerate treatment efficacy.[29] As a trial progresses, the estimated treatment effect varies randomly in relation to the true effect. If the interim estimate is based on a randomly high indication of efficacy, it is more likely to cross a statistical stopping boundary and to convince a data and safety monitoring board that overwhelming evidence of benefit exists. Stopping early also truncates evidence for important secondary (and safety) outcomes.

In the FAME 2 trial,[30] for example, PCI was compared with medical therapy alone in patients with stable coronary artery disease and hemodynamically significant lesions (as assessed by means of fractional flow reserve). The trial was stopped early because the hazard ratio for the primary outcome (all-cause death, myocardial infarction, or urgent revascularization) favoring PCI was 0.39 (95% CI, 0.26 to 0.57; P<0.001). This benefit with PCI was driven by fewer urgent revascularizations, an arguably "soft" outcome in an unblinded trial. The rate of death or myocardial infarction, although lower with PCI (hazard ratio 0.79; 95% CI, 0.49 to 1.29; P=0.35),

was inconclusive. Had the trial continued to its intended completion, a significantly lower rate of death or myocardial infarction may have emerged, which would have greatly enhanced the value of the trial.

Another recent example is the SPRINT trial[31] of intensive versus standard blood-pressure control, which had a composite primary outcome of myocardial infarction, acute coronary syndrome, stroke, heart failure, or cardiovascular death. The trial was stopped early at a median of 3.26 years rather than at the intended 5-year follow-up, and at the time the trial was stopped, the hazard ratio for the primary outcome with intensive control was 0.75 (95% CI, 0.64 to 0.89; P<0.001). The exceptional speed to publication was surprising[32]; only 4 weeks lapsed between the time that the trial was stopped and the time at which the manuscript was submitted for publication. The quality and completeness of any interim database are inevitably imperfect — there will be outstanding primary (and other) events yet to be ascertained and adjudicated. In addition, the moment at which a trial is stopped is the time at which an exaggerated estimate of efficacy is more likely to be present. Orderly trial closure after early stoppage takes several months and is necessary to achieve robust interpretation of all the evidence. Earlier reporting of preliminary, incomplete results is usually unwise.

DO CONCERNS ABOUT SAFETY COUNTERBALANCE POSITIVE EFFICACY?

When a new treatment has superior efficacy, it is important to identify concerns about safety that might offset the benefits. A balanced account of both efficacy and safety must be provided.[33] Absolute benefits and risks should be presented in terms of differences in percentages. Consideration of the number needed to treat for benefit versus the number needed to harm may provide a guide to net clinical benefit.

In the DAPT trial,[34] for example, an additional 18 months of dual antiplatelet therapy versus aspirin alone beginning 1 year after the implantation of a drug-eluting stent resulted in rates of major adverse cardiac and cerebrovascular events and of stent thromboses (the two primary efficacy outcomes) that were lower by 1.6% and 1.0%, respectively. However, this benefit came at the cost of higher rates of major bleeding events. According to Global Use of Strategies to Open Occluded Arteries (GUSTO)
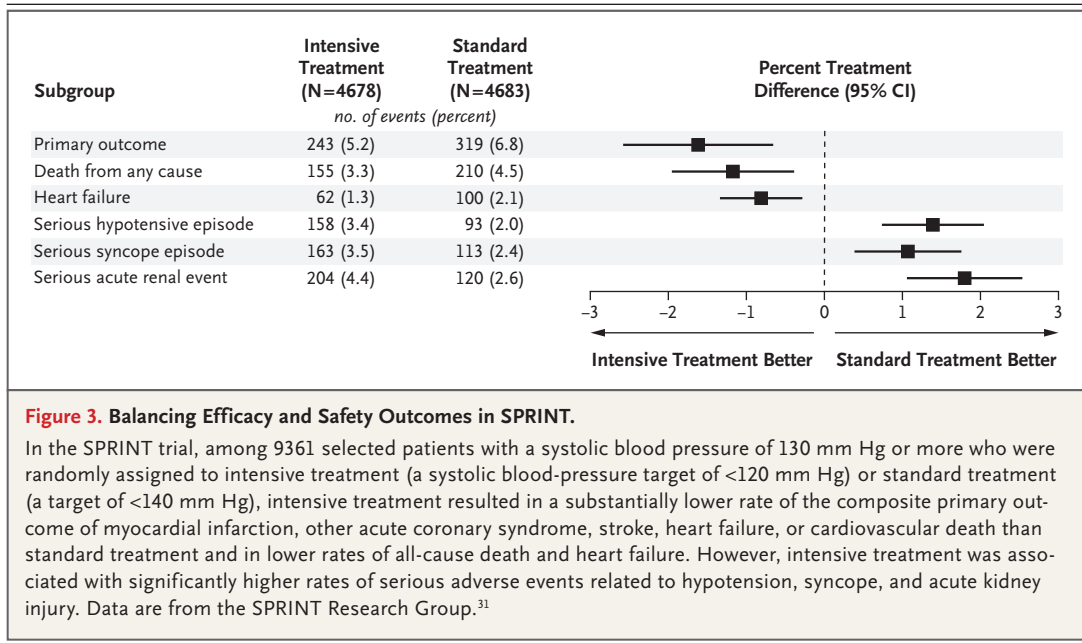
Subgroup	Intensive Treatment (N=4678)	Standard Treatment (N=4683)	Percent Treatment Difference (95% CI)
	no. of events (percent)		
Primary outcome	243 (5.2)	319 (6.8)	
Death from any cause	155 (3.3)	210 (4.5)	
Heart failure	62 (1.3)	100 (2.1)	
Serious hypotensive episode	158 (3.4)	93 (2.0)	
Serious syncope episode	163 (3.5)	113 (2.4)	
Serious acute renal event	204 (4.4)	120 (2.6)	

−3 −2 −1 0 1 2 3

Intensive Treatment Better   Standard Treatment Better

Figure 3. Balancing Efficacy and Safety Outcomes in SPRINT.

In the SPRINT trial, among 9361 selected patients with a systolic blood pressure of 130 mm Hg or more who were randomly assigned to intensive treatment (a systolic blood-pressure target of <120 mm Hg) or standard treatment (a target of <140 mm Hg), intensive treatment resulted in a substantially lower rate of the composite primary outcome of myocardial infarction, other acute coronary syndrome, stroke, heart failure, or cardiovascular death than standard treatment and in lower rates of all-cause death and heart failure. However, intensive treatment was associated with significantly higher rates of serious adverse events related to hypotension, syncope, and acute kidney injury. Data are from the SPRINT Research Group.[31]

criteria, the rate of "moderate or severe" bleeding events was 0.9% higher with continued dual antiplatelet therapy, and according to Bleeding Academic Research Consortium (BARC) criteria, the rate of bleeding that required medical attention was 2.7% higher. All-cause mortality was 0.5% higher with prolonged dual antiplatelet therapy (P=0.05), and this change was attributed primarily to greater noncardiovascular mortality (P=0.002), although some experts have argued that the higher rates of death may have been due to chance. Debate ensues: Is the net effect of prolonged dual antiplatelet therapy in this population beneficial or harmful?

Similarly, in the SPRINT trial,[31] intensive lowering of blood pressure resulted in a rate of the primary composite cardiovascular outcome that was 1.6 percentage points lower and a rate of death that was 1.2 percentage points lower than the rates with standard blood-pressure control during a median follow-up period of 3.26 years. However, these benefits must be weighed against rates of hypotension, syncope, and acute kidney injury, which were higher (by 1.4 percentage points, 1.1 percentage points, and 1.8 percentage points, respectively) with intensive blood-pressure control (Fig. 3). Note that all these benefits and risks, although statistically significant, represent small absolute differences. Thus, guideline committees, treating physicians, and patients face a challenge when trying to determine which strategy to adopt.

IS THE BALANCE OF EFFICACY AND SAFETY PATIENT-SPECIFIC?

The net clinical benefit of a new treatment may be patient-specific — that is, worthwhile for those at an increased risk for the primary efficacy outcome but deleterious for those at an increased risk for adverse events. Calculating the individual patient trade-offs between efficacy and safety is not straightforward, and statistical modeling techniques may be useful.[35]

For the DAPT trial,[34] multivariable models were developed for predicting the risk of myocardial infarction or stent thrombosis and the risk of major bleeding.[36] In addition to accounting for the duration of antiplatelet treatment (12 months vs. 30 months), these models included 9 patient and procedural characteristics and were designed to allow determination of the relative risks of ischemia versus bleeding in individual patients. Limitations included the omission of some variables that are known to predict the risk of ischemia and bleeding, the failure to directly consider the predictors of mortality, and the absence of external validation from a contemporary data set (as yet). Nonetheless, this analysis represents a useful development toward individualizing patient care.

ARE THERE FLAWS IN TRIAL DESIGN OR CONDUCT?

A highly significant result for the primary outcome goes a long way toward substantiating the view that findings cannot be attributed to chance.

Nonetheless, biases in the design and conduct of the trial must be ruled out before a genuine benefit can be acknowledged.

For instance, the first randomized trial of renal denervation in treatment-resistant hypertension, SYMPLICITY HTN-2,[37] showed that at 6 months, systolic blood pressure was markedly lower in the treatment group than in the control group (mean difference, 31 mm Hg; P<0.0001). However, the absence of blinding introduced major issues[38] (e.g., placebo and Hawthorne effects, ascertainment bias, and regression to the mean). In the subsequent, sham-controlled trial, SYMPLICITY HTN-3,[39] renal denervation appeared to be ineffective, thereby emphasizing the potential unreliability of unblinded trials.

Limitations in the completeness and quality of the data can also corrupt the validity of a trial. Not all patients fully comply with intended treatment regimens, and some withdraw from follow-up. Judgment is required in determining whether the extent of nonadherence or withdrawals casts doubt on the legitimacy of a trial. For instance, the ATLAS ACS 2–TIMI 51 trial[40] of rivaroxaban versus placebo in patients with acute coronary syndromes showed highly significant between-group differences in favor of the lower dose of rivaroxaban with respect to both the primary outcome (cardiovascular death, myocardial infarction, or stroke) and cardiovascular death alone. But 27.6% of the patients discontinued treatment prematurely, and data on vital status were missing for 7.2% of the patients — factors that introduced uncertainty. These problems appeared to be greater in this trial than in other large trials that address acute coronary syndromes and contributed to the FDA decision to withhold its approval of rivaroxaban for this indication.[41]

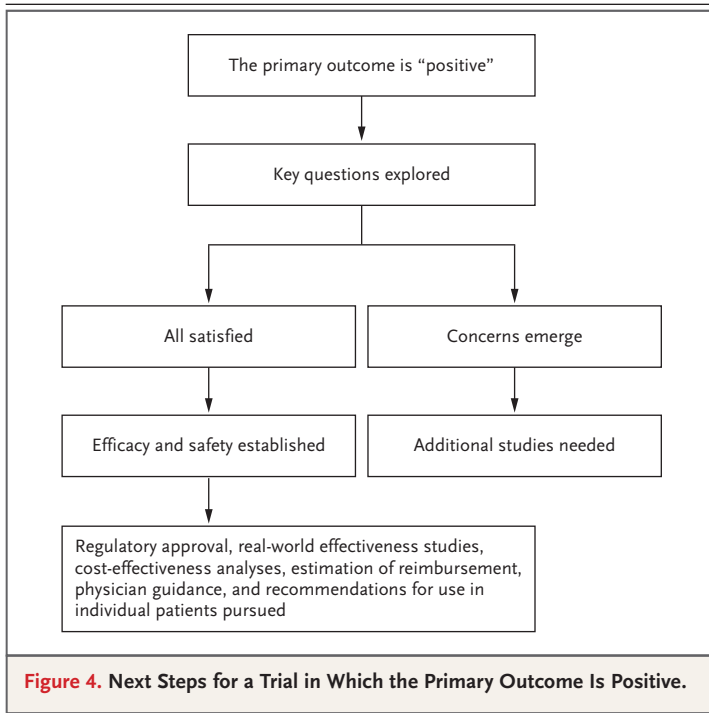### DO THE FINDINGS APPLY TO MY PATIENTS?

The findings of any trial apply to the specific patients enrolled and the therapies administered (both background and experimental). The question of whether the results can be generalized to other patients must be considered. For instance, the SPRINT trial[31] excluded patients younger than 50 years of age and those with diabetes or a history of stroke. The trial results (Fig. 3) thus apply to only approximately 20% of all patients with hypertension who are seen in practice.[42] Investigators in the ACCORD trial[43] previously considered the use of intensive blood-pressure control exclusively in patients with type 2 diabetes and reported no effect on cardiovascular events as compared with standard therapy. Whether the divergent results of the SPRINT and ACCORD trials are due to differences in patient characteristics, medications, trial methods, or other factors remains undetermined.

The geographic representation in a trial may also affect the generalizability of its results. Many major trials are multinational, which is an advantage in conferring global meaning. But health care practices may vary across regions (e.g., the use of primary PCI vs. fibrinolysis in patients with STEMI). If patient recruitment is dominated by one region, worldwide applicability may be limited. In addition, genetic, anatomical, environmental, and dietary differences among peoples sometimes make outcomes difficult to generalize across countries.

Similarly, the results from single-center trials must be viewed with caution. Center-specific effects, such as particular systems of care and the background therapies used, may preclude the generalizability of the findings, and single-center trials often lack quality-control measures. Results from single-center studies, even those with a reasonable sample size, should rarely serve as the basis for changing guidelines unless the results have been validated in subsequent multicenter trials. For example, the single-center TAPAS trial[44] of thrombus aspiration during primary PCI, which involved 1071 patients with STEMI, showed dramatically lower mortality at 1 year after PCI and thrombus aspiration than after conventional PCI (hazard ratio, 0.60; 95% CI, 0.36 to 0.98; P=0.04). In hindsight, this outcome was unrealistic given the modest benefit in reperfusion success, which was the primary outcome. However, this study led to widespread adoption of thrombus aspiration for many years. Two multicenter trials involving more than 17,000 patients have now convincingly shown that routine thrombus aspiration offers no advantages with regard to mortality or cardiovascular events.[45,46]

Finally, by the time the long-term findings for the primary outcome of a randomized trial become available, advances in care may have lessened their relevance to contemporary practice. For example, in the SYNTAX and FREEDOM trials,[47,48] patients with left main or multivessel disease were assigned to PCI with first-generation drug-eluting stents or to CABG. However,

If the efficacy and safety outcomes of the trial are convincingly met, the next step is to evaluate its overall quality and internal validity. It must also be determined whether the findings translate into treatment effectiveness (and net clinical benefit) in real-world patients. However, caution is required when data from nonrandomized registries are used to confirm or refute trial findings, given the potential for selection bias and residual confounding when such registries are used. At the same time, gauging the cost-effectiveness of treatments across different types of health care systems will determine the level of reimbursement (which in turn will affect the adoption of a new therapy).

For a new drug, the question of whether the results of a pivotal trial will satisfy the requirements for approval established by regulators (e.g., FDA and the European Medicines Agency) depends on the totality of the evidence from the trial and from all previous, related studies. Often, further evidence is required to clarify the safety profile of a new drug, and approval may depend on the sponsor's willingness to undertake additional safety studies (by means of randomized trials or observational registries).

Societal guideline committees play an important role in synthesizing the knowledge base and classifying the strength of evidence for new treatments; their endorsements strongly affect practice. Ultimately, however, physicians at the point of care bear the final responsibility for accurately interpreting clinical trial results and for integrating regulatory and guideline recommendations in order to make the best treatment decisions for each patient in their care.

Figure 4. Next Steps for a Trial in Which the Primary Outcome Is Positive.

contemporary drug-eluting stents represent a substantial improvement from first-generation devices,[49,50] a fact that diminishes the applicability of these trials to current practice.

DISCUSSION

A significance level of 5% for the primary efficacy outcome is the minimum requirement if a trial is to be declared "positive," and this level of significance should prompt deeper inspection into study processes and outcomes. Determining whether the evidence justifies the announcement of a major advance in medical care or a more cautionary note that further studies are warranted requires a comprehensive approach to all available evidence by various stakeholders (Fig. 4).

REFERENCES

1. Pocock SJ, Stone GW. The primary outcome fails — what next? N Engl J Med 2016;375:861-70.
2. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. N Engl J Med 1989;321:406-12.
3. Pocock SJ, Ware JH. Translating statistical findings into plain English. Lancet 2009;373:1926-8.
4. McMurray JJV, Packer M, Desai AS, et al. Angiotensin–neprilysin inhibition versus enalapril in heart failure. N Engl J Med 2014;371:993-1004.
5. Lees KR, Zivin JA, Ashwood T, et al. NXY-059 for acute ischemic stroke. N Engl J Med 2006;354:588-600.
6. Shuaib A, Lees KR, Lyden P, et al. NXY-059 for the treatment of acute ischemic stroke. N Engl J Med 2007;357:562-71.
7. Cannon CP, Blazing MA, Giugliano RP, et al. Ezetimibe added to statin therapy after acute coronary syndromes. N Engl J Med 2015;372:2387-97.
8. O'Riordan M. Ezetimibe fails to garner support for CV-event reduction claim from FDA advisory committee. December 15, 2015 (http://www.tctmd.com/show.aspx?id=133383).
9. Guidance for industry: diabetes mellitus — evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes. Silver Spring, MD: Food and Drug Administration, December 2008 (http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm071627.pdf).
10. The Action to Control Cardiovascular Risk in Diabetes Study Group. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med 2008;358:2545-59.
11. Follath F, Cleland JG, Just H, et al. Ef-

ficacy and safety of intravenous levosimendan compared with dobutamine in severe low-output heart failure (the LIDO study): a randomised double-blind trial. Lancet 2002;360:196-202.

12. Mebazaa A, Nieminen MS, Packer M, et al. Levosimendan vs dobutamine for patients with acute decompensated heart failure: the SURVIVE randomized trial. JAMA 2007;297:1883-91.

13. Fox KAA, Poole-Wilson PA, Henderson RA, et al. Interventional versus conservative treatment for patients with unstable angina or non-ST-elevation myocardial infarction: the British Heart Foundation RITA 3 randomised trial. Lancet 2002; 360:743-51.

14. Fox KAA, Poole-Wilson P, Clayton TC, et al. 5-Year outcome of an interventional strategy in non-ST-elevation acute coronary syndrome: the British Heart Foundation RITA 3 randomised trial. Lancet 2005; 366:914-20.

15. Bavry AA, Kumbhani DJ, Rassi AN, Bhatt DL, Askari AT. Benefit of early invasive therapy in acute coronary syndromes: a meta-analysis of contemporary randomized clinical trials. J Am Coll Cardiol 2006;48:1319-25.

16. Fox KAA, Clayton TCC, Damman P, et al. Long-term outcome of a routine versus selective invasive strategy in patients with non-ST-segment elevation acute coronary syndrome a meta-analysis of individual patient data. J Am Coll Cardiol 2010;55:2435-45.

17. Mentzer RM Jr, Bartels C, Bolli R, et al. Sodium-hydrogen exchange inhibition by cariporide to reduce the risk of ischemic cardiac events in patients undergoing coronary artery bypass grafting: results of the EXPEDITION study. Ann Thorac Surg 2008;85:1261-70.

18. Zinman B, Wanner C, Lachin JM, et al. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. N Engl J Med 2015;373:2117-28.

19. Cholesterol Treatment Trialists' (CTT) Collaborators. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. Lancet 2012;380:581-90.

20. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine — reporting of subgroup analyses in clinical trials. N Engl J Med 2007;357:2189-94.

21. Wallentin L, Becker RC, Budaj A, et al. Ticagrelor versus clopidogrel in patients with acute coronary syndromes. N Engl J Med 2009;361:1045-57.

22. Carroll KJ, Fleming TR. Statistical evaluation and analysis of regional interactions: the PLATO trial case study. Stat Biopharm Res 2013;5:91-101.

23. Tepel M, van der Giet M, Schwarzfeld C, Laufer U, Liermann D, Zidek W. Prevention of radiographic-contrast-agent–induced reductions in renal function by acetylcysteine. N Engl J Med 2000;343: 180-4.

24. Sun Z, Fu Q, Cao L, Jin W, Cheng L, Li Z. Intravenous N-acetylcysteine for prevention of contrast-induced nephropathy: a meta-analysis of randomized, controlled trials. PLoS One 2013;8(1):e55124.

25. Wald DS, Morris JK, Wald NJ, et al. Randomized trial of preventive angioplasty in myocardial infarction. N Engl J Med 2013;369:1115-23.

26. Engstrøm T, Kelbæk H, Helqvist S, et al. Complete revascularization versus treatment of the culprit lesion only in patients with ST-segment elevation myocardial infarction and multivessel disease (DANAMI-3–PRIMULTI): an open-label, randomised controlled trial. Lancet 2015; 386:665-71.

27. Gershlick AH, Khan JN, Kelly DJ, et al. Randomized trial of complete versus lesion-only revascularization in patients undergoing primary percutaneous coronary intervention for STEMI and multivessel disease: the CvLPRIT trial. J Am Coll Cardiol 2015;65:963-72.

28. ClinicalTrials.gov. Complete vs culprit-only revascularization to treat multivessel disease after primary PCI for STEMI (COMPLETE) (https://clinicaltrials.gov/show/NCT01740479).

29. Pocock S, White I. Trials stopped early: too good to be true? Lancet 1999; 353:943-4.

30. De Bruyne B, Fearon WF, Pijls NHJ, et al. Fractional flow reserve–guided PCI for stable coronary artery disease. N Engl J Med 2014;371:1208-17.

31. The SPRINT Research Group. A randomized trial of intensive versus standard blood-pressure control. N Engl J Med 2015; 373:2103-16.

32. Drazen JM, Morrissey S, Campion EW, Jarcho JA. A SPRINT to the finish. N Engl J Med 2015;373:2174-5.

33. Ioannidis JP, Evans SJ, Gøtzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 2004;141: 781-8.

34. Mauri L, Kereiakes DJ, Yeh RW, et al. Twelve or 30 months of dual antiplatelet therapy after drug-eluting stents. N Engl J Med 2014;371:2155-66.

35. Pocock SJ, Stone GW, Mehran R, Clayton TC. Individualizing treatment choices using quantitative methods. Am Heart J 2014;168:607-10.

36. Yeh RW, Secemsky EA, Kereiakes DJ, et al. Development and validation of a prediction rule for benefit and harm of dual antiplatelet therapy beyond 1 year after percutaneous coronary intervention. JAMA 2016;315:1735-49.

37. 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.

38. Howard JP, Shun-Shin MJ, Hartley A, Bhatt DL, Krum H, Francis DP. Quantifying the 3 biases that lead to unintentional overestimation of the blood pressure-lowering effect of renal denervation. Circ Cardiovasc Qual Outcomes 2016;9: 14-22.

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

40. Mega JL, Braunwald E, Wiviott SD, et al. Rivaroxaban in patients with a recent acute coronary syndrome. N Engl J Med 2012;366:9-19.

41. Krantz MJ, Kaul S. The ATLAS ACS 2-TIMI 51 trial and the burden of missing data: (Anti-Xa Therapy to Lower Cardiovascular Events in Addition to Standard Therapy in Subjects With Acute Coronary Syndrome ACS 2-Thrombolysis In Myocardial Infarction 51). J Am Coll Cardiol 2013;62:777-81.

42. Fuster V. Unraveling the complexities of statistical presentation: why it is important. J Am Coll Cardiol 2015;66:2909-10.

43. The ACCORD Study Group. Effects of intensive blood-pressure control in type 2 diabetes mellitus. N Engl J Med 2010;362: 1575-85.

44. Vlaar PJ, Svilaas T, van der Horst IC, et al. Cardiac death and reinfarction after 1 year in the thrombus aspiration during percutaneous coronary intervention in acute myocardial infarction study (TAPAS): a 1-year follow-up study. Lancet 2008;371: 1915-20.

45. Fröbert O, Lagerqvist B, Olivecrona GK, et al. Thrombus aspiration during ST-segment elevation myocardial infarction. N Engl J Med 2013;369:1587-97.

46. Jolly SS, Cairns JA, Yusuf S, et al. Randomized trial of primary PCI with or without routine manual thrombectomy. N Engl J Med 2015;372:1389-98.

47. 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.

48. Farkouh ME, Domanski M, Sleeper LA, et al. Strategies for multivessel revascularization in patients with diabetes. N Engl J Med 2012;367:2375-84.

49. Palmerini T, Benedetto U, Biondi-Zoccai G, et al. Long-term safety of drug-eluting and bare-metal stents: evidence from a comprehensive network meta-analysis. J Am Coll Cardiol 2015;65: 2496-507.

50. 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.

Copyright © 2016 Massachusetts Medical Society.