Randomized trials provide the gold standard evidence on which rests the decision to approve novel therapeutics for clinical use. They are large and expensive and provide average but unbiased estimates of efficacy and risk. Concern has been expressed about how unrepresentative populations and conditions that pertain in randomized trials might be of the real world, including concerns about the homogeneity of the biomedical and adherence characteristics of volunteers entered into such trials, the dose and constancy of drug administration and the mixture of additional medications that are restricted in such trials but might influence outcome in practice. A distinction has been drawn between trials that establish efficacy and those that demonstrate effectiveness, drugs that patients actually consume in the real world for clinical benefit. However, randomized controlled trials remain the gold standard for establishing efficacy and the testing of effectiveness with less rigorous approaches is a secondary, albeit important consideration. Despite this, there is an appreciation that average results may conceal considerable interindividual variation in drug response, leading to a failure to appreciate clinical value or risk in subsets of patients. Thus, attempts are now being made to individualize risk estimates by modulating those derived from large randomized trials with the individual baseline risk estimates based on demographic and biological criteria—the individual Numbers Needed to Treat to obtain a benefit, such as a life saved. Here, I will consider some reasons why large phase 3 trials—by far the most expensive element of drug development—may fail to address the unmet medical needs, which should justify such effort and investment. (Circ Res. 2014;114:1156-1161.)

Key Words: cholesterol • clinical trial • cyclooxygenase • niacin • rosiglitazone • thromboxane receptor
rigorous approaches is a secondary, albeit important consideration. Despite this, there is an appreciation that average results may conceal considerable interindividual variation in drug response, leading to a failure to appreciate clinical value or risk in subsets of patients.\(^2\) Thus, attempts are being made to individualize risk estimates by modulating those derived from large randomized trials with individual baseline risk estimates based on demographic and biological criteria—the individual Numbers Needed to Treat to obtain a benefit, such as a life saved.\(^4\) Here, I will consider some reasons why large phase 3 randomized trials—by far the most expensive element of drug development—may fail to address the unmet medical needs, which should justify such effort and investment.

**Whose Need Is Unmet?**

Those best positioned to address whether a need is unmet are the relevant patients, and this has most effectively been established by those with AIDS. However, more commonly the definition of whose need is unmet is made by pharmaceutical companies that largely foot the bill for drug development and often do so exclusively for the large-scale randomized trials necessary for their drug’s approval by the Food and Drug Administration (FDA).

In the case of nonsteroidal anti-inflammatory drugs (NSAIDs), the discovery of cyclooxygenase-2 prompted the hypothesis that NSAIDs selective for inhibition of cyclooxygenase-2 would be less likely to result in serious gastrointestinal complications (that were attributed to inhibiting cyclooxygenase-1 in platelets and gastroduodenal epithelium) than that of older drugs that inhibited both cyclooxygenase enzymes. Indeed, this hypothesis was not rejected by the outcomes of 2 randomized trials that showed the selective inhibitors, rofecoxib\(^5\) and lumiracoxib\(^6\) were less likely to cause such complications than their comparators. Although these trials—and those of celecoxib\(^7\) and valdecoxib\(^8\) showing that surrogates of such gastrointestinal end points were less likely to occur than that with mixed inhibitors—were designed to obtain drug approval based on an improved gastrointestinal safety profile, they left unaddressed 2 fundamental issues relevant to clinical practice. The first was whether there was any difference in analgesic or anti-inflammatory efficacy between selective and nonselective NSAIDs. Although preclinical\(^9\) and clinical research\(^10\) were consistent with cyclooxygenase-2 being the major source of prostaglandins mediating pain and inflammation, there is evidence for a variable contribution from cyclooxygenase-1; indeed increased expression of both enzymes is evident in inflammatory tissues, such as synovia from the joints of patients with rheumatoid arthritis.\(^11\) No formal assessments of comparative efficacy were included in the large-scale trials of NSAIDs.

The second issue was evidence that emerged as these trials began, consistent with a cardiovascular hazard from cyclooxygenase-2 selective NSAIDs\(^12\) and of a pharmacodynamic interaction between NSAIDs that inhibited cyclooxygenase-1 (and cyclooxygenase-2) and low-dose aspirin.\(^13\) Such competitive, active site inhibitors of cyclooxygenase-1 could impede access by aspirin to its target serine residue in the platelet enzyme. This substituted sustained inhibition of thromboxane-dependent platelet function because of covalent modification of cyclooxygenase-1 with transient inhibition during a dosing interval, thereby undermining the cardioprotective effect of low-dose aspirin. This interaction would not be expected with drugs that were selective for inhibition of cyclooxygenase-2 because this form of the enzyme is not extant in mature human platelets.

Because patients with chronic arthritis often have concomitant heart disease, they are commonly receiving low-dose aspirin for cardioprotection. Even low doses of aspirin can confer risk of gastrointestinal adverse events, and this risk adds to that of nonselective NSAIDs.\(^14\) However, the recognition of a cardiovascular risk from selective inhibitors of cyclooxygenase-2 because of suppression of prostacyclin (PGI\(_2\)) formation\(^12\) left practitioners with an unmet medical need: how best to treat their patients with both heart disease and arthritis? No clinical trials designed to gain approval of cyclooxygenase-2 inhibitors addressed that question; either patients who needed aspirin were excluded\(^7\) or the trials were not powered to address comparative drug effect in patients who were versus those who were not on concomitant aspirin.\(^7\)

Just as trials may be designed primarily to gain drug approval rather than address comprehensively an unmet medical need, they can also be designed in a way that protects a franchise. Celecoxib was predicted to confer a similar cardiovascular hazard as rofecoxib.\(^15\) Three randomized trials\(^12\) provide evidence that it does confer such a hazard and this risk was indistinguishable from that of rofecoxib in a recent overview analysis of 700 randomized trials of NSAIDs.\(^16\) Despite this, rofecoxib was withdrawn from the market by its sponsor, whereas celecoxib is available, albeit with a black box warning. Shortly after the FDA advisory committee voted (25-0) in 2005 that celecoxib did indeed confer a cardiovascular hazard, the sponsor announced the Prospective Randomized Evaluation of Celecoxib Integrated Safety Versus Ibuprofen or Naproxen (PRECISION) trial as a statement of their confidence in the comparative cardiovascular safety of celecoxib in patients with or at high risk of cardiovascular disease. This is a noninferiority trial with a power now adjusted to only 80% and an upper bound of confidence of 1.4, thus arguably biased toward the null. Aside from this limitation, this trial was not powered or designed a priori to compare patients randomized to the 3 comparators on and off aspirin. This fundamentally undermines the ability to interpret the results of the trial because administered doses of both ibuprofen and naproxen, but not celecoxib, pharmacodynamically interact with aspirin in the platelet; a win for celecoxib could reflect no more than an asymmetrical impairment of cardioprotection by low-dose aspirin in the other 2 groups. An attempt to circumvent this limitation by urging patients to take their morning dose of aspirin before the NSAID stems from a misinterpretation of the original article on the drug interaction; sufficient ibuprofen remains after the evening dose to sustain the interaction even if aspirin is given first in the morning.\(^17\) These
design flaws were highlighted before the trial was initiated, and it was suggested that such an event-driven trial would last much longer than the few years predicted by its leadership because doctors were unlikely to enlist patients with a high risk of heart disease into a trial, including a drug with such an established risk. Indeed, the European Medicines Agency refused to approve the trial for performance in the European Union and only ≈50% of patients currently enrolled in PRECISION are said to be taking aspirin. Such a prolonged trial might be seen to defend the franchise by deferring an answer while the patent for celecoxib remained extant and it was the only 1 of the 3 comparators that stood to benefit from direct to consumer advertising for arthritis. Similar design defects apply to the Standard Care versus celecoxib Outcome Trial (SCOT). This noninferiority study, purposefully targeting patients at low cardiovascular risk, also failed to account for the possibility of aspirin–NSAID interactions and had originally included an open-label phase of celecoxib administration to determine tolerability before randomization. This was omitted after the first 2 years (2008–2010) because it was deemed not to influence adherence rates and at this time power was adjusted, like PRECISION, to 80% with an upper bound of 1.4. Meantime, back in the United States, celecoxib, sustained by the longest direct to consumer patent life in 2014; in the meantime, the estimated completion date for PRECISION is September 2015. There are other examples where unanticipated prolongation of recruitment into randomized controlled trials has helped defend the franchise. The Improved Reduction of Outcomes: Vytorin Efficacy International Trial (IMPROVE-IT) has sustained interest in the potential cardiovascular efficacy of ezetimibe, despite unimpressive data from smaller studies and its potential association with an excess of cancer deaths. Here, the question is whether patients with relatively low low-density lipoprotein cholesterol on statin therapies might benefit from further reduction in cholesterol by addition of ezetimibe.

Where Is the Rationale?
A recent analysis of drug development drawing information from >800 commercial entities estimates that the likelihood of approval of a drug entering phase 1 is on the order of 10%. The likelihood of approval for cardiovascular disease is even lower, ≈6%; here, approval is often configured on death rates and large trials are required. Although most failures occur at the transition from phase 2 to 3, the failure rates in phase 3 cardiovascular trials approximate 50% which, given their cost, ≈60% of the cost of drug development, presents a particular challenge to industry.

Phase 2 classically represents an opportunity to elucidate the mechanism of drug action and to clarify how that might be of therapeutic relevance to a particular clinical condition. However, this requires the integration of preclinical knowledge, a sophisticated approach to deep phenotyping of drug response in small numbers of patients and a comprehensive understanding of clinical medicine and trials methodology. Investigators with such an integrated interdisciplinary background in basic and clinical pharmacology are in short supply, either as scientists within industry or as advisors from the academic sector. An example of a large phase 3 trial performed on the basis of a limited a priori rationale is the PREVEnTToN of CeReBROVAsCulAr and cardiovascular Events of ischemic origin with teRutroban in patients with a history of ischemic strOke or trAnsient ischemic attack (PERFORM) study, a comparison of a thromboxane receptor antagonist (terutroban 30 mg/d) and low-dose (100 mg/d) aspirin. Here, ≈9500 patients who had an ischemic stroke in the previous 3 months or a transient ischemic attack in the previous 8 days were randomized between the treatments.

The first mechanistic rationale for this study was that the antagonist would not suppress PG1, while blocking the platelet aggregatory and vasconstrictive effects of thromboxane A2 (TxA2). A second was that the antagonist would block the effects of incidental ligands at the TxA2 receptor, such as free radical catalyzed isoprostanes, which would be unaffected by cyclooxygenase-1 blockade by low-dose aspirin. Suppression of PG1 formation does have cardiovascular consequences, as we now know from the cyclooxygenase-2 inhibitor saga. However, unlike such NSAIDs, which suppress biosynthesis by 60% to 80%, low-dose aspirin is relatively selective for cyclooxygenase-1 and suppresses prostacyclin by only ≈20%. It might take a large trial indeed to reveal the discriminant benefit of avoiding this modest suppression.

As for isoprostane formation, these makers of lipid peroxidation are markedly elevated during reperfusion after a period of tissue ischemia. However, their levels are barely altered in patients such as those admitted to PERFORM who have not been subjected to therapeutic reperfusion. Thus, the mechanistic basis for the trial design was highly questionable from the outset, and it was perhaps unsurprising that the near $1 billion study was stopped prematurely for futility on the basis of the recommendation of the Data Monitoring Committee.

Biomarker Beliefs and Magical Thinking
One reason the failure rate in phase 3 trials of cardiovascular disease is so high is that biomarkers measured in phase 2 poorly predict the outcome measured in phase 3, most commonly survival. However, there is also the issue of wishful thinking. A good example here is the development of torcetrapib, an inhibitor of cholesterol ester transfer protein (CETP). Here, studies in phase 2 revealed the desired dose-dependent elevation of high-density lipoprotein (HDL), an apparent biomarker of cardiovascular benefit. However, much confusion has surrounded the therapeutic value of raising HDL. Epidemiological studies have long associated HDL inversely with cardiovascular risk, but these observations are confounded by many aspects of a healthy lifestyle that may cosegregate with HDL. Perhaps the most convincing evidence to the contrary is the failure of genetic variance in HDL levels to relate to cardiovascular risk. Unfortunately, the same studies of torcetrapib revealed a dose-related increase in a much more established biomarker albeit of risk not benefit, blood pressure. Vesting belief in validation of the underlying hypothesis for drug development—an elevation in HDL—rather than in the unexpected increase in the more validated end point—blood pressure—resulted in the close to $1 billion Investigation of Lipid Level Management to Understand its Impact in Atherosclerotic Events (ILLUMINATE) trial. This
demonstrated that torcetrapib actually increased cardiovascular event rates and prompted its abandonment. An immediate question for sponsors of trials of other CETP inhibitors was whether this represented a consequence of CETP inhibition or an off-target effect of torcetrapib. Evidence emerged that torcetrapib might have conferred a cardiovascular risk not only via an increase in blood pressure but also, and perhaps distinctly, from an increase in plasma aldosterone by up-regulation of CYP11B2 (aldoasterone synthase). Importantly, torcetrapib elevated blood pressure in rodents, which lack CETP. This was not a feature of the other 3 CETP inhibitors in development. However, a phase 3 trial in =16,000 patients of another CETP inhibitor, dalcetrapib, failed to reveal a cardiovascular benefit (although unlike torcetrapib, it did not increase risk. Phase 3 trials of 2 other inhibitors, anacetrapib and evacetrapib, continue. The absence of a clear effect of loss of function mutations of CETP—which elevate HDL cholesterol and lower low-density lipoprotein cholesterol—on cardiovascular outcomes and the failure of other genetic variants in HDL alone to influence events heighten the sense of uncertainty in this gamble. The residual hope is that salutary effects of CETP inhibitors independent of their effect on HDL, such as a reduction in low-density lipoprotein cholesterol, lipoprotein (a) and apoB, might deliver the desired outcome. Alternatively, as Rader and deGoma suggest, this may be the first pharmacological mechanism to see 4 different molecules fail in large-scale, >$1 billion phase III trials.

Another example of magical thinking in phase 2 is the response to results of the Integrated Biomarker and Imaging Study (IBIS)-2 of the lipoprotein-associated phospholipase A2 inhibitor, darapladib. Here, 2 primary end points were specified based on intravascular ultrasound palpography; C-reactive protein was among the many secondary end points. Despite darapladib failing to influence either primary end point (or C-reactive protein), the decision was taken to proceed to phase 3. This was based on the attenuation by darapladib of one of the secondary end points in IBIS-2, the increase with time in lipid necrotic core, as assessed by virtual histology. This decision was despite the recognition that (1) the role of lipoprotein-associated phospholipase A2 in cardiovascular biology is controversial (it also degrades platelet-activating factor, suggesting its inhibition might not be desirable) and (2) failure of a loss of function mutation to influence cardiovascular outcome in 26,000 individuals. Meantime, a phase 3 trial of varespladib, an inhibitor of a secretory phospholipase A2, was terminated for lack of efficacy.

**How Useful Is the Answer?**

Another drug long known to elevate HDL is niacin, and trials that failed to show its cardiovascular benefit have been criticized as underpowered or for limitations of trial design. The latter was true of prematurely terminated Atherothrombosis Intervention in Metabolic Syndrome with Low HDL/High Triglycerides: Impact on Global Health (AIM-HIGH) that involved administration of lower doses of niacin to the control group to mask the effect of niacin-induced flushing. The Treatment of HDL to Reduce the Incidence of Vascular Events (HPS2-THRIVE) trial was an opportunity to answer these outstanding questions. It was much larger than the antecedent clinical trials (25,673 patients) and suitably powered to detect a reasonable expectation of benefit. However, this was a trial to assess the usefulness of a drug combination, extended-release niacin (2 g) and the prostaglandin D2 receptor antagonist, laropiprant (40 mg), in the reduction of major vascular events (composite of nonfatal myocardial infarction, coronary heart disease death, stroke, or arterial revascularization). The study was not designed to answer the questions of how either drug alone acted or how they compared with the combination. Extended-release niacin/laropiprant treatment reduced low-density lipoprotein cholesterol by an average 10 mg/dL and increased HDL cholesterol by 6 mg/dL. The primary end point was not different between the 2 groups. This study provided a clear answer for the drug combination and the sponsor ceased its further development. However, it is impossible to parse with confidence the differential effect of the 2 elements of this combination and to know how the results might extend to the common use of niacin. The results of ongoing trials of 2 remaining CETP inhibitors and inhibitors of PCSK9 (proprotein convertase subtilisin/kexin type 9) will influence the HDL discussion. Will the performance of a sufficiently large and well-controlled study of niacin in more defined populations, say those relatively resistant to benefit from statins, ever seem like a cost-effective initiative to a commercial sponsor? The answer will likely drive the progressive abandonment of niacin, a drug that has long been a mainstay of cardiovascular therapy, while we still poorly understand its many potentially relevant mechanisms of action and have an incomplete picture of its clinical utility.

**Subversion and Disruption**

The rosiglitazone saga began with an overview analysis in 2007 of heterogeneous trials suggested the possibility of a cardiovascular hazard from the antidiabetic drug rosiglitazone. In contrast to the experience with cyclooxygenase-2 inhibitors, the data that were the basis for pressing the FDA to restrict access to this drug did not derive from placebo-controlled trials, and the limitations of the analysis were much debated. Again, in contrast to the cyclooxygenase-2 story, there was no experimental basis to explain the mechanism of the hazard, never mind why this risk should attach to rosiglitazone more than other members of the class. Considerable political pressure, not least abetted by evidence of bad behavior on the part of the sponsor, resulted in the FDA attaching severe restrictions to the use of rosiglitazone in 2010. Meantime sales of the drug and, more importantly, recruitment into a randomized comparison of rosiglitazone with pioglitazone and other diabetes mellitus drugs began to drop precipitously. Subsequently the FDA turned its attention to the Rosiglitazone Evaluated for Cardiovascular Outcomes and Regulation of Glycemia in Diabetes (RECORD) study, designed to assess the effects of rosiglitazone on cardiovascular outcomes in a randomized trial and commissioned an analysis by Duke University independent of the commercial sponsor. This failed to support the original hypothesis advanced by the 2007 meta-analysis of a cardiovascular hazard. Although the FDA then removed language about an asymmetrical cardiovascular risk from the label, the damage was done: rosiglitazone is highly unlikely ever to recover a substantial market share. There were few heroes in this story, replete as it was with data concealment, and
allegations of intimidation and conflict of interest. However, it illustrates the necessity of obtaining evidence from randomized trials as a basis for drug approval and the limitations of other types of evidence, such as observational studies, preclinical studies in model systems, human genetics, and clinical pharmacology or, as in this case, a meta-analysis of heterogeneous trials. These approaches can raise a question or indeed provide an explanation for the outcome of randomized trials but not substitute for them in establishing risk or benefit.

Conclusions
Randomized trials remain the basis for drug approval, yet their high failure rate imposes an enormous cost on the biomedical enterprise. Attention has been drawn recently to the limitations of preclinical data as predictors of likely clinical response. In particular, the poor reproducibility of some data has led to calls for more rigorous standards in preclinical drug studies in model systems, including randomization, blinding, and public provision of primary data. In some cases, carefully collected preclinical data fail dramatically to predict human biology, while in others they predict it efficiently. Aside from the rigor with which preclinical studies are performed, it is worth remembering that a model is no more than a model and while the same answer drawn from >1 species may enhance the predictive power of such data, there is no species in which drug action unfailingly predicts what happens in humans. Interestingly, although studies in humans supported by industry are biased in favor of drug efficacy when compared with those performed without such support, precisely the opposite bias seems to persist in preclinical studies. Perhaps such preclinical studies are performed more carefully within the industry because they will influence investment decisions directly, but once the drug moves into clinical development, there are pressures from within the company to reinforce that decision.

There has been a call to perform phase 1 studies under randomized, double-blind conditions, so better to provide interpretable data on which to base the decision to proceed. Other advances, such as microdosing to derive estimates more safely of both pharmacokinetic and pharmacodynamic responses to new chemical entities and genomic and metabolomic information to identify more accurately populations susceptible to drug action, may also improve the efficiency of the process. However, the examples in this review focus on problems further downstream. These include how we decide what questions to address in clinical trials, how we interpret the rationale for such decisions, how vested interest in the evolution or cessation of a program in drug development can cloud judgment, particularly when it comes to the interpretation of suggestive, but inconclusive results in phase 2.

The opportunity and the challenge is to rely on a more thorough understanding of the mechanism of drug action, dependent on reliable biomarkers, and deep phenotyping approaches in phase 2 and to design more targeted, smaller, and faster trials in phase 3 that are more likely to succeed. An example of such a strategy is the success of ivacaftor in improving forced expiratory volume by +10% in patients with cystic fibrosis with the G551D mutation in the cystic fibrosis gene. Similarly, such a strategy can lead to a faster failure or the redesign of a trial based on such a failure. An example is the failure of ivacaftor in a subsequent study of 69 patients with the R117H mutation. Here, a prespecified analysis of the 50 patients aged ≥18 years suggested benefit, a possibility that can be addressed directly and rapidly in a further study.

Several further initiatives may serve to strengthen the trials process. These include (1) more vigorous engagement with patients in the determination of both unmet clinical need and the effectiveness of efficacious therapies; (2) provision of resource to perform trials that are in the interest of the public health but not of a commercial sponsor; (3) a safe haven, provided by the FDA, for systems pharmacology and incentives to explore of the full breath of drug action—not just what is thought to relate to the primary indication being pursued—early in human development, and (4) a strategic expansion of investigators trained in translational medicine and therapeutics—spanning the preclinical to clinical divide.

These initiatives should be coupled with an integrative approach to mechanistic data—drawn from preclinical models and humans—with pharmacoepidemiology to develop hypotheses relating to adverse effects because randomized trials are typically underpowered to detect them. Examples of where such an integrated approach together with refined trial design has been helpful are in detection and elucidation of the cardiovascular hazard from NSAIDs and in probing the cardiovascular adverse effects of erythropoietin, and its potential to undermine efficacy in the treatment of cancer. Finally, the process would benefit from a more detailed and accessible provision to the public of incentives provided by industry to academia—both to investigators and to their institutions via indirect costs—than is presently extant and a recognition that conflict of interest can relate to fame as well as to fortune.

Sources of Funding
Dr FitzGerald is the McNeil Professor in Translational Medicine and Therapeutics. This work was supported by the institutional Clinical and Translational Science Award (UL1 TR000003).

Disclosures
None.

References
21.

Ther non-steroidal anti-inflammatory drugs: meta-analyses of individual par-

versus celecoxib outcome trial (SCOT).

BMJ versus traditional non-steroidal anti-inflammatory drugs in patients

randomised, open, blinded endpoint streamlined safety study of cele-

SCOT Study Group Collaborators. Methodology of a large prospective,

with cerebral ischaemic events (PERFORM): a randomised, double-blind,

Bousser M, Economidou C, FitzGerald GA. Emotion recollected in tranquility: lessons


Grosser T, Fries S, FitzGerald GA. Biological basis for the cardiovascular

consequences of COX-2 inhibition: therapeutic challenges and opportuni-


Patrano C, García Rodríguez LA, Landolfi R, Baigent C. The impact of COX-2 inhibi-


Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial


Djulbegovic B, Hozo I, Ioannidis JP. Improving the drug development pro-


Testing Cardiovascular Drug Safety and Efficacy in Randomized Trials
Garret A. FitzGerald

Circ Res. 2014;114:1156-1161
doi: 10.1161/CIRCRESAHA.114.301809
Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/114/7/1156

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation Research can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation Research is online at:
http://circres.ahajournals.org/subscriptions/