Randomized clinical trials (RCTs) are familiar to most clinicians involved in the care of patients with cardiovascular disease. RCTs are the uncontested "gold standard" for establishing causation between a therapy and outcomes for a compelling reason: A large randomized experiment is the only study design that can guarantee that control and intervention subjects are similar in all the known and unknown attributes that influence outcomes. In general, RCTs use a rigorous experimental design with the primary goal of understanding the "average" overall benefit and risk of using a standardized therapy in a selected group of patients. When well conducted, RCTs have internal validity, and a therapy that is found to have a net biological benefit is considered efficacious. Establishing the efficacy of new therapies is critical for advancing medical knowledge.

Yet even when an RCT demonstrates substantial net benefit for a specific therapy, its optimal role in routine clinical practice can remain unclear. Concerns frequently relate to how well the therapy may perform in different clinical settings and in broader patient populations than those studied in traditional RCTs. In addition, it is often unclear how therapies are best applied within the larger environment of the healthcare delivery system where costs and availability become issues for patients and clinicians. These concepts, which refer to the external validity of a clinical trial, become important as new therapies are applied in real-world settings and are a key focus area for many studies in health services and outcomes research. In contrast to efficacy studies like traditional RCTs, effectiveness studies often require a broader range of methodological tools, which can include both experimental and nonexperimental study designs.

In the present article, we review reasons why evidence from traditional RCTs must generally be supplemented by evidence from effectiveness studies to inform best clinical practice. In particular, we discuss (1) why obtaining evidence that considers potential differences between the efficacy (ie, can the treatment work under ideal circumstances?) and effectiveness (ie, will the treatment work in real-world circumstances?) of cardiovascular therapies is essential, and (2) how effectiveness studies can help provide such evidence.

The Achilles’ Heel of RCTs
It would be desirable to rely solely on RCTs to guide clinical practice, but this is simply not feasible. At the root of the problem is the fact that RCTs are typically restricted to evaluating specific discrete interventions one at a time. This restriction limits their ability to (1) directly assess complex interactions within a study arm (ie, interactions between different blood pressure medicines used to obtain tight blood pressure control), (2) continuous relationships (such as what blood pressure or cholesterol levels are optimal), and (3) whether the benefits or harm of a treatment are drug-specific or mechanistic. As a result, traditional RCTs focus more on evaluating the efficacy of simple therapies like drugs and less on the delivery of care. Such issues can only be addressed within an experimental framework when each factor is evaluated in isolation from the others, which can make the costs to conduct RCTs that evaluate all these issues prohibitive. For example, it would take 32 different study arms to examine all the possible combinations of just 5 treatments.

Furthermore, RCTs are principally designed to assess the “average” or mean treatment effect in the overall population that is studied. This forces researchers to decide between conducting the study on a highly selected and homogeneous study population—limiting external validity—or on a diverse study population, a choice that makes the research more expensive to conduct and that is prone to false-negative findings caused by low adherence or heterogeneity in treatment benefit. The desire to optimize the possibility of a positive finding leads most traditional RCTs to use strict study protocols and highly selected patients, and to be conducted at the best medical centers in the country. Furthermore, even within these highly selected patients there can be heterogeneity in the extent of benefit found with a therapy. For example, scenarios may exist where an average treatment effect favors a therapy because of its benefit to a few high-risk patients, although most individuals in the study gain little from it.

Finally, RCTs often require enormous resource investments and may be inherently limited in their ability to investigate certain issues for logistical or ethical reasons. The
Table 1. Potential Reasons for Differences Between the Efficacy and Effectiveness of a Therapy

<table>
<thead>
<tr>
<th>Study design</th>
<th>Patient selection</th>
<th>Therapeutic implementation</th>
<th>Environment of the healthcare delivery system</th>
<th>Inadequate levels of reimbursement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial protocol forces uncommon clinical scenarios</td>
<td>Biases in the patients who are eligible for a therapy</td>
<td>Complex and multifaceted therapies are challenging to implement</td>
<td>Limited availability of providers or resources</td>
<td></td>
</tr>
<tr>
<td>Comparison group does not represent current standard of care</td>
<td>Biases in patients who are ultimately selected for (or agree to) a therapy</td>
<td>Procedural experience of providers influences outcomes</td>
<td>Inadequate levels of reimbursement</td>
<td></td>
</tr>
<tr>
<td>Outcomes include less meaningful end points</td>
<td>Therapeutic implementation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

recently completed Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation (COURAGE) trial, which compared optimal medical therapy with and without percutaneous coronary intervention (PCI) in 2287 patients, resulted in nearly $60 million in total costs shared by both public and private sponsors (personal communication, William E. Boden, MD, 2008). The logistical challenges of RCTs also make them poorly suited for addressing questions about new therapies in uncommon or heterogeneous conditions such as hypertrophic cardiomyopathy or aortic dissection where it may be difficult to find large numbers of patients suitable for randomization. Ethical considerations are also important to recognize because placebo-controlled comparisons may be inappropriate in certain settings, particularly when the relative advantages of a new therapy are related to cost, convenience, or safety and not efficacy when compared with standard treatments.

Understanding Differences in the Efficacy and Effectiveness of Therapies

Several experts have proposed frameworks for understanding the extent to which findings from RCTs can be applied, or “generalized,” to routine clinical practice. In general, differences between the efficacy and effectiveness of a therapy may result from issues related to (1) study design or trial protocol (including the assessment of outcomes), (2) patient selection, (3) therapeutic implementation, and (4) the larger environment of the healthcare delivery system in which the therapy is being applied. These issues are summarized in Table 1.

Study Design Limitations

To isolate the biological effect of a therapy, an RCT needs to maximize its use in the group randomized to receive it (ie, compliance) and minimize its use in the comparison group (ie, crossover), which can create scenarios in traditional RCTs that, at times, vary substantially from care provided in routine clinical practice. For example, the Danish Multicenter Randomized Study on Fibrinolytic Therapy versus Acute Coronary Angioplasty in Acute Myocardial Infarction (DANAMI-2) trial compared primary PCI with fibrinolytic therapy for ST-elevation myocardial infarction (STEMI). 8

Less than 2% of STEMI patients treated with fibrinolytic therapy underwent rescue PCI and fewer than 20% underwent early revascularization. 8 Although this permitted a more direct comparison between primary PCI and fibrinolytic therapy in STEMI, it does not reflect contemporary practice patterns in the United States where use of rescue PCI and early revascularization after fibrinolytic therapy is more routine. 9,10

Similarly, the Fragmin and Fast Revascularization During Instability in Coronary Artery Disease (FRISC-II) trial, which found a significant benefit with an early invasive strategy when compared with ischemia-driven (or conservative) management after an acute coronary syndrome, had only a 10% crossover rate to coronary angiography in the first 7 days among those patients who underwent conservative management. This rate was due to the strict criteria used for identifying ischemia during stress testing and led to a relatively low overall use of coronary revascularization within the first year. The emphasis of the study therefore shifted from evaluating an early invasive strategy to assessing the efficacy of coronary revascularization. In contrast, the rate of coronary angiography during initial hospitalization in the Invasive versus Conservative Treatment in Unstable Coronary Syndromes (ICTUS) trial was over 50%. 12 This high rate led to higher rates of coronary revascularization in the first year than in FRISC II and is more reflective of contemporary practice patterns in the United States. 13 It is also noteworthy that the ICTUS trial found no differences between an early invasive strategy and conservative management in this patient population, despite treating a higher-risk population.

Another concern that can complicate the interpretability of an RCT is the choice of a comparison group. This comparison group can involve usual care, a placebo arm, or an active control. The ability of an RCT to inform routine clinical practice may be influenced heavily by this issue, especially when the management of a disease is complex and relies on a combination of competing therapies. The extensive literature on antihypertensive therapies illustrates this point. Older trials compared β-blockers and diuretics to placebo whereas recent trials have evaluated newer agents with active controls, resulting in multiple “drug-to-drug” comparisons. It has been argued that newer studies, such as the Anglo-Scandinavian Cardiac Outcomes Trial-Blood Pressure Lowering Arm (ASCOT-BPLA) trial, 14 do not always use the most appropriate comparison therapy 15 or that trial protocols frequently do not achieve similar blood pressure control between treatment arms. 16 In fact, when multiple common practice patterns are present in the medical community, it is not feasible for an RCT to assess all of them, which can lead clinicians to dismiss the RCT as being irrelevant because the “usual care” group did not reflect their own practice.

The hesitancy to withhold standard treatments because of ethical concerns has led to a greater use of noninferiority designs in RCTs, but this use has raised a number of conceptual and methodological challenges,9 as well as confusion in reporting. 17 Under ideal circumstances, noninferiority trials require that the new therapy that is being evaluated
have some potential advantages over standard treatments that are unrelated to efficacy, such as lower costs, more convenience, or better safety. Many of these studies, however, may simply be “me-too”–type trials for newer marketable drugs. Another concern is that noninferiority trials may be difficult to interpret because they rely on “historical” information from earlier placebo-controlled trials, as well as on subjective thresholds of clinically tolerable differences. Furthermore, it is possible that establishing that a new therapy is noninferior to standard treatments may make the new therapy the standard treatment in subsequent noninferiority trials. This can lead to a phenomenon of “biocrep” or “regression to mediocrity,” in which the standard treatment is no longer better than placebo.

The types of outcomes that are evaluated by an RCT can also affect its ultimate value in routine clinical practice. Contemporary RCTs may evaluate a number of different end points but often use a composite that is tied to a common biological mechanism (eg, myocardial ischemia) in order to maximize statistical power. However, this can lead to “softer” end points such as revascularization and rehospitalization being measured and weighted the same as “harder” end points such as mortality. Softer end points are often less reliable and may be less relevant to clinicians and patients. These attributes of softer end points become important because softer end points are often what drive differences in the composite end point between therapy and comparison groups. For example, a recent systematic review of 114 RCTs in cardiovascular diseases found that statistically significant differences were most prominent for softer end points and marginal for harder end points. We lack sufficient resources to examine most important medical questions using RCTs adequately powered to examine hard outcomes alone. However, combining end points also becomes problematic when therapies have differing effects on soft and hard end points. This issue was recently demonstrated in the case of drug-eluting stents, which reduce subsequent coronary revascularization but may lead to higher rates of late stent thrombosis and acute myocardial infarction.

Finally, the narrow focus and limited sample sizes of most traditional RCTs make it difficult to anticipate rare but important side effects or complications of a therapy. This difficulty is especially important when (1) the complications are already common in the general population (as with myocardial infarction), (2) their occurrence may be delayed for several years, and (3) a potential exists for interactions with other comorbidities and medications. For example, a recent meta-analysis reported that rosiglitazone, a thiazolidinedione agonist for the peroxisome-proliferator–activated receptor γ, may be associated with an increased risk of myocardial infarction and cardiovascular death. Although RCTs dating back to the mid-1990s have examined the efficacy of this drug in insulin-resistant diabetes mellitus, these studies were too small and brief to individually detect an increased risk for adverse outcomes. Even in combining 42 RCTs, the confidence intervals assessing the increased risk for cardiovascular death were quite wide, extending from 0.98 to 2.74.

**Patient Selection Limitations**

It is well recognized that differences between the efficacy and effectiveness of a therapy may result from the selective recruitment of patients in traditional RCTs. Substantial differences in age, gender, and comorbidities have been reported between patients in traditional RCTs and real-world registries because of explicit exclusion criteria and subtle recruitment biases. By necessity, traditional RCTs usually target a subgroup of individuals from the larger universe of potentially eligible patients on the basis of who is most likely to participate or respond to the therapy. This targeting is done in order to isolate the biological effect of a therapy and to minimize sample size requirements and cost. However, in routine clinical practice, the luxury of not including sicker or more complicated patients is not available. Paradoxically, such targeting can result in a paucity of empirical data on those patients most likely to require the therapy.

For example, ≈60% of patients with heart failure have concomitant renal insufficiency—with 15% having glomerular filtration rates <30 mL/min—and this is associated with a worse prognosis. However, RCTs testing angiotensin-converting enzyme inhibitors, β-blockers, and spironolactone in patients with advanced heart failure have frequently excluded patients with renal insufficiency. Others have reported similar issues for this same group of patients in RCTs of coronary artery disease. The potential danger of generalizing findings from RCTs to real-world settings was recently demonstrated in a large population-based study that found higher rates of hyperkalemia and mortality in hospitalized patients with heart failure who were treated with spironolactone after publication of the Randomized Aldactone Evaluation Study (RALES) trial. These findings were primarily explained by the use of this drug in patients at greater risk for hyperkalemia than those in the RALES trial. Conversely, the lack of data in high-risk patients may also contribute to the well-described “risk-treatment paradox” found for other therapies. Under this scenario, those at greatest risk of death are the least likely to be treated, even when they are likely to receive the greatest absolute benefit.

**Therapeutic Implementation Limitations**

When a therapy involves a simple discrete intervention or straightforward strategy—eg, the single administration of an intravenous drug during PCI—the ability to implement it in a real world environment is not difficult to envision. In many studies, however, it is complex and multifaceted therapies that are being tested. Reproducing results in routine clinical practice may be challenging. One framework for generalizing the potential impact of a therapy outside the setting of RCTs suggests that the size of the trial and the simplicity of the therapy are critical to its application. Larger trials and simpler therapies are the most likely to be associated with greater overall effectiveness in routine clinical practice (Figure 1). Several RCTs, for example, have found that comprehensive disease management programs in selected heart failure patients can lead to significant reductions in mortality and morbidity. The overall findings from these studies show clear benefit, but substantial heterogeneity across trials has
been noted in meta-analyses because the types of programs that have been studied vary in modes of communication with patients (telephone, home visit, and clinic visit), frequency and duration of contact, quality of interaction with the primary physician, and the training levels and aptitude of clinical managers.33 The overall effectiveness of these programs depends heavily on how well these complex components can be incorporated into a routine clinical practice.

The performance of specialized surgical and medical procedures can also change as a therapy disseminates outside of RCTs to less-experienced operators and facilities. A classic example of this is carotid artery endarterectomy. Several RCTs have established the efficacy of this procedure in comparison with medical therapy in symptomatic and asymptomatic patients with severe carotid artery stenosis. Wennberg and colleagues, however, identified significantly higher 30-day mortality rates with the procedure in Medicare beneficiaries at hospitals not participating in RCTs compared with those that were involved in the landmark RCTs.34 Indeed, in hospitals that performed carotid artery endarterectomy infrequently, adjusted mortality rates were nearly 2-fold higher: 2.5% versus 1.4%. This difference has led to the recommendation that local surgical expertise be considered when referring patients for carotid artery endarterectomy because higher complication rates after surgery may eliminate the potential therapeutic benefits noted in RCTs.35

Healthcare Policy Limitations

The structure of the healthcare system into which a therapy is being introduced has implications for both its overall use and effectiveness. Many factors, often unrelated to clinical issues, come into play, including cost, reimbursement by third-party payers, geographic availability, and ease of access to the therapy.

As noted earlier, the superiority of primary PCI over fibrinolytic therapy in STEMI patients previously has been demonstrated in several RCTs.36 However, these trials were conducted at high-volume centers with experienced operators, and trials that specifically focused on interhospital transfer for primary PCI were largely conducted within integrated healthcare systems outside of the United States. All of these factors resulted in rapid times to treatment with primary PCI.37 Extrapolating from these findings that primary PCI is universally superior to fibrinolytic therapy in routine clinical practice is challenging. In many parts of the United States, using primary PCI has important limitations because of inadequate prehospital emergency medical systems, delayed access to cardiac catheterization facilities with PCI expertise, and weak cooperation between hospital systems.38,39 Similarly, although early data were favorable for positron emission tomography in diagnosing coronary artery disease, its use was limited in much of the United States because of poor availability of positron emission tomography cameras and radiotracers and unfavorable reimbursement policies.40

Expanding Our Knowledge Base on Effectiveness

Assessing the effectiveness of new therapies outside of traditional RCTs is an important priority for health services and outcomes research. Proposed solutions to address this challenge fall into the following categories: (1) conducting pragmatic clinical trials and improved reporting, (2) synthesizing available studies using meta-analysis and decision analysis techniques, and (3) performing high-quality cohort and observational studies. Key strengths and weaknesses of these different approaches are summarized in Table 2.

Pragmatic Clinical Trials and Improved Reporting

After traditional RCTs have established that a treatment can work (ie, efficacy), a pragmatic clinical trial will often be the best method for understanding the effectiveness of that therapy in different clinical settings and in broader patient populations. These trials “address practical questions about the risks, benefits, and costs of a therapy in routine clinical practice,” rather than its biological efficacy.41 They are designed to overcome many of the potential limitations of RCTs by including (1) use of active controls and clinically relevant comparisons, (2) broad study populations, (3) enrollment from diverse clinical environments, (4) evaluations of strategies of treatment rather than single drugs or devices, and (5) measurement of multiple health outcomes including those related to patient symptoms, quality of life, and costs. A well-known pragmatic clinical trial is the landmark Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack trial (ALLHAT), which evaluated multiple antihypertensive drugs, including a low-cost thiazide diuretic, and pravastatin in a broad population of hypertensive and dyslipidemic patients enrolled from multiple centers.42

Although pragmatic clinical trials can be extremely useful in understanding the effectiveness of therapies, they are limited by the enormous amount of resources and time required for completion and still may have study design limitations. The ALLHAT trial, which was primarily funded by the National Institutes of Health, enrolled patients at >600 North American sites and took 8 years to complete. Furthermore, the trial protocol for “step-up” drugs may have produced combination regimens that were somewhat artificial and less clinically relevant.43 Regardless of this limitation, the ALLHAT trial was influential in increasing the role of

Figure 1. Proposed framework for generalizing the results of a randomized clinical on the basis of (1) trial size and (2) the complexity of the therapy. Trials that are larger and that evaluate less complex therapies are the most likely to be applicable in routine clinical practice (ie, efficacy=effectiveness). Those that are smaller and evaluate complex therapies may have important differences between efficacy and effectiveness. Adapted from Flather et al46 with permission of the publisher.
thiazide diuretics as first-line therapy in hypertensive patients. To encourage more of these types of trials, many have argued for offsetting costs by linking payments by insurers to enrollment in clinical trials or registries when the therapy is still considered investigational. Medicare recently used this policy to limit the expansion of carotid artery stenting to broader patient populations, which also encouraged participation in ongoing RCTs evaluating this innovative therapy. Others have suggested specific initiatives to expand federal and industry sources for funding such studies.

Some also argue that improved reporting of existing RCTs could allow for better recognition of differences between the

<table>
<thead>
<tr>
<th>Table 2. Strengths and Weaknesses of Types of Effectiveness Studies in Relation to Traditional Randomized Clinical Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional randomized clinical trial</td>
</tr>
<tr>
<td><strong>Strengths</strong></td>
</tr>
<tr>
<td>Experimental design ensures comparability between treated and nontreated patients</td>
</tr>
<tr>
<td>Standardized methodological approaches to study design and data analysis</td>
</tr>
<tr>
<td><strong>Weaknesses</strong></td>
</tr>
<tr>
<td>Difficulty in assessing multiple or complex therapies simultaneously</td>
</tr>
<tr>
<td>Relies often on highly selected homogenous population at large medical centers</td>
</tr>
<tr>
<td>Smaller sample sizes limit the ability to detect minor treatment effects or rare complications</td>
</tr>
<tr>
<td>High costs and resource investments</td>
</tr>
<tr>
<td>Logistical and ethical challenges, particularly in studying less common but life-threatening diseases</td>
</tr>
<tr>
<td>Pragmatic clinical trials</td>
</tr>
<tr>
<td><strong>Strengths</strong></td>
</tr>
<tr>
<td>Experimental design ensures comparability between treated and nontreated patients</td>
</tr>
<tr>
<td>Standardized methodological approaches to study design and data analysis</td>
</tr>
<tr>
<td>Involves a large and diverse population of patients in real-world settings</td>
</tr>
<tr>
<td><strong>Weaknesses</strong></td>
</tr>
<tr>
<td>Extremely high costs and resource investments</td>
</tr>
<tr>
<td>Smaller sample sizes limit the ability to detect minor treatment effects or rare complications</td>
</tr>
<tr>
<td>Logistical and ethical challenges, particularly in studying less common but life-threatening diseases</td>
</tr>
<tr>
<td>Meta-analysis</td>
</tr>
<tr>
<td><strong>Strengths</strong></td>
</tr>
<tr>
<td>Captures a broader application of therapies in systematic manner</td>
</tr>
<tr>
<td>Pooling of data allows for evaluations of small but important subgroups of patients</td>
</tr>
<tr>
<td>Can provide insights into heterogeneity of treatment effects across studies</td>
</tr>
<tr>
<td><strong>Weaknesses</strong></td>
</tr>
<tr>
<td>Susceptible to study design limitations within individual studies</td>
</tr>
<tr>
<td>Suffers from potential publication and reporting bias of large treatment effects</td>
</tr>
<tr>
<td>Heterogeneity may be inadequately addressed in study selection or analysis</td>
</tr>
<tr>
<td>Difficult to obtain patient-level data from investigators for pooling</td>
</tr>
<tr>
<td>Decision analysis</td>
</tr>
<tr>
<td><strong>Strengths</strong></td>
</tr>
<tr>
<td>Flexible framework for addressing clinically relevant questions using explicit model assumptions</td>
</tr>
<tr>
<td>Can incorporate measurements of quality of life and costs</td>
</tr>
<tr>
<td><strong>Weaknesses</strong></td>
</tr>
<tr>
<td>Depends upon the strength of evidence for explicit model assumptions</td>
</tr>
<tr>
<td>Limited description of model construction and structure in many articles</td>
</tr>
<tr>
<td>Observational studies</td>
</tr>
<tr>
<td><strong>Strengths</strong></td>
</tr>
<tr>
<td>Involves a large and diverse population of patients in real-world settings</td>
</tr>
<tr>
<td>Allows for the ability to examine uncommon diseases and to detect minor treatment effects or rare complications</td>
</tr>
<tr>
<td>Can provide insight into the clinical context in which multiple or complex therapies are delivered</td>
</tr>
<tr>
<td>Relatively inexpensive and can be performed rapidly</td>
</tr>
<tr>
<td><strong>Weaknesses</strong></td>
</tr>
<tr>
<td>Confounding or selection bias of patients—or the two combined—make it difficult to compare treated and nontreated patients</td>
</tr>
<tr>
<td>Multiple methodological approaches are often available but may be inconsistently applied or reported</td>
</tr>
</tbody>
</table>
efficacy and effectiveness of a therapy, particularly as it pertains to important patient subgroups.47 Traditional subgroup analyses are commonly presented in an attempt to assess who may benefit most from treatment, but these are frequently limited by poor statistical power.48 An alternative is to use risk-prediction models to evaluate the benefit of a therapy on the basis of the baseline risk of suffering an adverse outcome ("risk-stratified analysis"). These models, which account for overall risk using a single multivariable risk score, can frequently overcome the usual shortcomings of conventional subgroup analyses (poor statistical power and multiple comparisons) and can be especially helpful when a treatment has even a small incidence of treatment-related harm.47–49 The key limitation of this approach, which is the availability of established risk-prediction models, is fortunately less of an issue in many areas of cardiovascular disease such as STEMI and acute coronary syndromes.

Kent et al used this approach to reanalyze data from the Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries (GUSTO) trial, which compared streptokinase with tissue plasminogen activator in patients with STEMI.50 These investigators found that approximately half the patients in the GUSTO trial were responsible for 85% of the mortality benefit of tissue plasminogen activator. In addition, an important subgroup of patients was identified in whom the increased risk of intracerebral hemorrhage associated with tissue plasminogen activator outweighed its benefits (Figure 2). Similarly, Rothwell and Warlow found that most of the benefit of CEA in the European Carotid Surgery Study occurred in less than 20% of patients with high risk of stroke and low-moderate risk of surgery.51 Thune and colleagues have also reported that the benefits of primary PCI in the DANAMI-2 trial appeared entirely restricted to the 26% of patients with high Thrombolysis in Myocardial Infarction (TIMI) risk scores of 5 or more.52

Meta-Analysis and Decision Analysis
Systematic reviews and meta-analyses are important tools for synthesizing data from multiple RCTs to determine a therapy’s effectiveness. This is important since it is not uncommon for findings from initial RCTs of a therapy to be refuted or challenged over time as it is tested in different populations and settings. In a recent review, nearly one quarter of highly-cited RCTs published between 1990 and 2003 in leading medical journals were subsequently shown to have reported results that initially overstated effects in favor of the therapy.53 Pooling data across studies also allows for evaluations of its impact on specific subgroups of patients and in varying clinical environments.54 Finally, understanding how differences in trial protocols contribute to heterogeneity may give clinicians the ability to determine how best to apply a new therapy in their practice. Caution must be used when combining data from RCTs with small sample sizes or significant clinical or statistical heterogeneity, because summary results may be susceptible to study design limitations within individual studies or sensitive to the analytical approach. A recent reevaluation of data from the meta-analysis on rosiglitazone and cardiovascular death that was noted above highlights these challenges.55

Computer simulation models used in decision analysis provide a flexible framework for addressing important clinical questions.56 Models are frequently based on a series of underlying assumptions that are related to the efficacy and
safety of a therapy, typically obtained from RCTs or a meta-analysis. One important advantage of decision analysis is that it allows investigators the flexibility to vary assumptions across ranges of potential values to reflect uncertainty, including how the preferred approach might change if advances in treatments occurred in the future (eg, being able to lower the complication rate or increase the effectiveness of a treatment). It also allows for direct modeling of health state utilities such as quality-adjusted life-years that can incorporate composite end points. Chan et al recently used this approach in a study to compare high-dose and conventional-dose statin therapy in patients with coronary artery disease. These investigators found that high-dose statin therapy was highly cost-effective in high-risk patients with acute coronary syndromes but much less so in stable coronary artery disease.

A significant limitation of decision analysis is its dependence on the quality of baseline assumptions entered into the model. However, a well-conducted study can highlight that information on a particularly influential factor is currently insufficient, which should help direct future research to obtain more accurate information and inform clinical practice. Although decision analysis has the potential to produce invalid results, its pitfalls are not so much inherent in the approach itself but rather result from improper application. In particular, the medical literature will often be inadequate for decision analysis to provide a definitive answer, and in these situations the approach should be focused on providing information on what data we need in order to settle a clinical question. An example of the value of these types of “threshold” analyses is a recent study that explicitly quantified the extent of risk reduction in stroke that would be required for catheter-based radiofrequency ablation of atrial fibrillation to be cost-effective. Educating researchers and clinicians in how to assess and use these types of studies is therefore important.

Observational Studies
The use of nonrandomized observational studies is an important tool for determining the effectiveness of a therapy in routine clinical practice. Such studies often have cohort or case-control designs, which allow for the inclusion of broader populations of patients and providers than RCTs. When well conducted, these studies can provide clinicians and patients with a more realistic expectation for outcomes in real-world environments than traditional RCTs. A recent study examining the use of implantable cardioverter defibrillators in a routine clinical practice setting confirmed the substantial mortality benefit associated with these devices. Opportunities for these types of studies are likely to grow even more as information technology and electronic records expand to permit more inexpensive and uniform capture of healthcare utilization and outcomes.

These studies also have the potential to collect richer and varied data elements, which can include detailed socioeconomic data. This information can allow for an improved understanding of the determinants of outcomes in high-risk patients undergoing new therapies. For example, Spertus and colleagues recently reported on the prevalence and predictors of early discontinuation of thienopyridines after drug-eluting stent placement in 500 patients treated at 19 centers across the United States. In addition to identifying psychosocial factors that were linked to patients discontinuing these drugs, these investigators noted a 9-fold higher risk of death at follow-up when this occurred. This suggests that this type of information, which is rarely collected in traditional RCTs, provides valuable data for clinicians to decide on how best to use new therapies as they are disseminated into routine clinical practice.

However, observational studies may often be limited in their ability to account for confounding and bias due to patient selection. The use of hormone replacement therapy for preventing cardiovascular events in postmenopausal women is one of several examples where data from observational studies pointed to a dramatic benefit, but harm was found in an RCT. Even advanced statistical techniques that are used to minimize the potential for confounding have recently been shown to have significant limitations. For example, a recent study by Stukel et al examined the benefits of cardiac catheterization after myocardial infarction within the Cooperative Cardiovascular Project (CCP) and found that its association with improvements in long-term mortality was highly dependent on the analytic approach that was used. Both the standard risk-adjustment models and propensity score analysis used by these investigators were potentially biased and overestimated the impact of this intervention when compared with traditional RCTs, perhaps because of their inability to address unmeasured biases in patient selection for this procedure.

Just as important as the choice of statistical technique is its application and reporting in a specific setting. Application and reporting are critical for observational studies, which often can be examined using an array of complex analytic approaches that may influence the results. For example, Normand and colleagues also examined the benefits of cardiac catheterization after myocardial infarction using the same data set (ie, CCP) and analytic approach (ie, propensity score analysis) described above. Because of minor differences in selection criteria for their patient population and model variables, however, they arrived at potentially less-biased estimates when compared with the report by Stukel et al. A recent study by Schneeweiss and colleagues demonstrates this point further. These investigators examined how the stepwise use of exclusion criteria to define a patient population in an observational study affected the strength of associations between statin use and 1-year mortality. As the patient population became increasingly similar to cohorts enrolled in clinical trials, the results of standard risk-adjustment models approached traditional RCTs. Improving the application and reporting of statistical techniques is the impetus behind recent pushes for “reproducible research” in observational studies through limited sharing of data and programming code.

When carefully applied, observational studies remain a reliable and even invaluable source for informing routine clinical practice. It is only through observational studies that we have been able to understand the association between smoking and heart disease, between aspirin and Reyes syndrome, and between some diet pills and increased risk of
pulmonary hypertension. A traditional RCT, for example, could not have identified the association between sudden cardiac death and the concurrent use of erythromycin and strong inhibitors of the cytochrome P-450 3A isozymes or the relationship between major congenital malformations and angiotensin-converting enzyme inhibitor use in the first trimester.70,71

Furthermore, once RCTs demonstrate that a specific drug therapy works, observational studies can often be conducted expeditiously to examine whether its benefits are similar to other medications in the same class, how this association varies with dose, and whether the association differs across patient populations. Often when observational studies suggest treatment heterogeneity or a lack of generalizability, the best answer is to conduct a new RCT to confirm or refute this finding, but ignoring the evidence from observational studies is not a viable option. In fact, a strong argument can be made that, when carefully conducted and treatment is not heavily biased by patient selection, observational analyses are at least as reliable as small RCTs.72

Bridging the Gap Between Efficacy and Effectiveness

Although it is important to understand the strengths and limitations of both efficacy and effectiveness studies, this exercise ultimately represents a limited perspective when performed in isolation. All types of evidence for new therapies rely primarily on the rigor with which individual studies were conducted, regardless of the methodological approach, and the care with which they are interpreted. For example, Avorn has wisely pointed out that choosing between RCTs and observational studies is really a “nonchoice.”73 In the end, informing best clinical practice requires a broader outlook that requires us to synthesize multiple sources of evidence. Rather than viewing data as competing, a more productive strategy would be to understand the value each type of evidence brings toward moving quality of patient care forward.

In 2002, Califf et al introduced such a view through his conceptual model of the “cycle of quality.”74 This model integrated the concepts of new drug and device development from traditional RCTs with contemporary health services and outcomes research aimed at measuring their effective delivery. The importance of health services and outcomes research is not only in evaluating the impact of therapies in the real world, but in spurring further innovation in the development of new therapies which ultimately closes the cycle.75 Recent advancements in the care for STEMI patients over the last 15 years are a real-world example of this cycle working successfully.76 Traditional RCTs in the early 1990s found that primary PCI was superior to fibrinolytic therapy for reperfusion therapy. These traditional RCTs were followed by effectiveness studies which found that primary PCI in the real world was frequently limited by delays in its timely delivery leading to worse outcomes. Investigators subsequently identified hospital-based strategies that could improve its timely delivery. At the same time, these effectiveness studies have encouraged the development of new RCTs aimed at improving fibrinolytic therapy and exploring combination approaches to reperfusion. For STEMI patients, the availability of large hospital networks participating in RCTs and registries, clinical guidelines and performance measures, and national quality improvement efforts has been instrumental in this regard.

Conclusion

RCTs are performed with the primary goal of understanding the efficacy of a new therapy in a selected group of patients. Even when evidence from these studies strongly favors a new therapy, its role in routine clinical practice may be unclear because of considerable differences between its efficacy and its effectiveness. Improving our ability to understand when and why these differences exist and how they influence the management of patients with cardiovascular disease is a challenging but important goal of health services and outcomes research exemplified by effectiveness studies. The methods used in effectiveness studies are frequently complex, and assessing them typically requires more detailed examination and expertise than are required by traditional RCTs. However, the careful conduction and evaluation of such studies is essential to translating findings from RCTs into high-quality evidence that can guide routine clinical practice.

Disclosures

None.

References

11. Invasive compared with non-invasive treatment in unstable coronary-artery disease: FRISC II prospective randomised multicentre study.


Beyond the Randomized Clinical Trial: The Role of Effectiveness Studies in Evaluating Cardiovascular Therapies

Brahmajee K. Nallamothu, Rodney A. Hayward and Eric R. Bates

Circulation. 2008;118:1294-1303
doi: 10.1161/CIRCULATIONAHA.107.703579

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2008 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/118/12/1294

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation 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 is online at:
http://circ.ahajournals.org/subscriptions/