Chapter 7: Therapy (Randomized Trials)

Michael Walsh; Vlado Perkovic; Braden Manns; Sadeesh Srinathan; Maureen O. Meade; PJ Devereaux; Gordon Guyatt

Clinical Scenario

CLINICAL SCENARIO

A Patient With Peripheral Artery Disease: How Can I Improve Physical Function and Walking?

You are a general internist following up a 62-year-old man with a history of type 2 diabetes mellitus, hypertension, and hyperlipidemia who is taking oral hypoglycemics, a statin, and a thiazide-like diuretic. A vascular surgeon recently evaluated the patient for intermittent claudication and made a diagnosis of peripheral artery disease. The surgeon prescribed low-dose aspirin and pentoxifylline to reduce the patient's risk of vascular events and improve his ability to walk, citing 2 *systematic reviews*: a review of antiplatelet agents in peripheral artery disease that found a decrease in the odds of vascular events (*odds ratio* [OR], 0.78; 95% *confidence interval* [CI], 0.63-0.96) and an increase in walking distance of 59 m (95% CI, 37-81 m) and another review of pentoxifylline in peripheral artery disease that increased maximum walking distance by 59 m (95% CI, 37-81 m).[1,2] Despite the new treatments, the patient is unable to walk more than 2 minutes without pain and finds his quality of life substantially impaired.

Listening to the patient's story of poor response to treatment and ongoing symptoms, you recall seeing an article that may be relevant. You ask him to return in a week for further review of his medications.

Finding the Evidence

You formulate the relevant question for this individual: in a patient with debilitating peripheral vascular disease treated with antiplatelet therapy and not a candidate for surgery, how can we improve symptom-free walking? To conduct a rapid search focused on the most recent preappraised research (see chapter 5, Finding Current Best Evidence), you opt for *ACP Journal Club*, directly accessible through your institution (http://acpjc.acponline.org.ezproxy.library.dal.ca). Typing the terms “peripheral vascular disease” and “intermittent claudication” identifies 7 preappraised editorial summaries of studies, one of which turns out to be your target: Ramipril Improved Walking Times and QOL in Peripheral Artery Disease and Intermittent Claudication.[3] You print a copy of the summary[3] and the original full-text article that reports the results of the trial, Effect of Ramipril on Walking Times and Quality of Life Among Patients With Peripheral Artery Disease and Intermittent Claudication.[4]

This article describes a trial that includes 212 patients with peripheral artery disease and a history of stable intermittent claudication. Participants were *randomly allocated* to ramipril, 10 mg daily, or *placebo* for 24 weeks. The primary *outcome*s were pain-free walking time and maximum walking time.

The Users' Guides

Box 7-1 presents our usual 3-step approach to using an article from the medical literature to guide your practice. You will find these criteria useful for a variety of therapy-related questions, including treating symptomatic illnesses (eg, asthma or arthritis), *prevention* of distant complications of illness (eg, cardiovascular death after myocardial infarction), *screening* for silent but treatable disease (eg, colon cancer screening), and choosing the optimal diagnostic approach (as in randomized trials of alternative diagnostic strategies that address *patient-important outcomes*).

BOX 7-1

Users' Guides for an Article About Therapy

How serious was the risk of bias?
- Did intervention and control groups start with the same prognosis?
- Were patients randomized?
- Was randomization concealed?
- Were patients in the study groups similar with respect to known prognostic factors?
- Was prognostic balance maintained as the study progressed?
- To what extent was the study blinded?
- Were the groups prognostically balanced at the study's completion?
- Was follow-up complete?
- Were patients analyzed in the groups to which they were randomized?
- Was the trial stopped early?

What are the results?
- How large was the treatment effect?
- How precise was the estimate of the treatment effect?

How can I apply the results to patient care?
- Were the study patients similar to my patient?
- Were all patient-important outcomes con- sidered?
- Are the likely treatment benefits worth the potential harm and costs?

If the answer to one key question ("Were patients randomized?") is no, some of the other questions ("Was randomization *concealed*?" "Were patients analyzed in the groups to which they were randomized?") become irrelevant. Nonrandomized *observational studies* typically yield far weaker inferences than *randomized clinical trials* (RCTs). Nevertheless, clinicians must use the best *evidence* available in managing their patients, even if the quality of that evidence is limited (see chapter 2, What Is Evidence-Based Medicine?). The criteria in chapter 14 (Harm [Observational Studies]) will help you assess an observational study that addresses a potential treatment that has not yet been evaluated in an RCT.

How Serious Is the Risk of Bias?

Did Intervention and Control Groups Start With the Same Prognosis?

Were Patients Randomized?

Consider the question of whether hospital care prolongs life. A study finds that more sick people die in the hospital than in the community. We would easily reject the naive conclusion that hospital care kills people because we recognize that hospitalized patients are sicker than patients in the community.

Although the logic of prognostic balance is vividly clear in comparing hospitalized patients with those in the community, it may be less obvious in other contexts. Many people believe that a diet rich in ω3 fatty acids will decrease their risk of a cardiovascular event. This belief arose from many observational studies in which people who ingested larger quantities of ω3 fatty acids had fewer cardiovascular events than those that who ate lesser quantities.[5] However, large randomized trials did not find any benefits with ω3 fatty acid supplementation.[6,7]

Other surprises generated by randomized trials include the demonstration that antioxidant vitamins fail to reduce gastrointestinal cancer[8]—and one such agent, vitamin E, may actually increase all-cause mortality[9]—and that a variety of initially promising drugs increase mortality in patients with heart failure.[10-12] Such surprises occur periodically when investigators conduct randomized trials to test the observations from studies in which patients and physicians determine which treatment a patient receives.

The reason that studies in which patient or physician preference determines whether a patient receives treatment or control (observational studies) often yield misleading results is that morbidity and mortality result from many causes. Treatment studies attempt to determine the impact of an intervention on events such as stroke, myocardial infarction, and death—occurrences that we call the trial's *target outcomes*. A patient's age, the underlying severity of illness, the presence of *comorbidity*, and a host of other factors typically determine the frequency with which a trial's target outcome occurs (*prognostic factors* or *determinants of outcome*). If prognostic factors—either those we know about or those we do not know about—prove unbalanced between a trial's treatment and *control groups*, the study's outcome will be biased, either underestimating or overestimating the treatment's effect. Because known prognostic factors often influence clinicians' recommendations and patients' decisions about taking treatment, observational studies often yield biased results that may get the magnitude or even the direction of the effect wrong.

Observational studies can theoretically match patients, either in the selection of patients for study or in the subsequent statistical analysis, for known *prognostic factors* (see chapter 14, Harm [Observational Studies], and Chapter 11.1, An Illustration of Bias and Random Error). However, not all prognostic factors are easily measured or characterized, and in many diseases only a few are known. Therefore, even the most careful patient selection and statistical methods are unable to completely address the *bias* in the estimated *treatment effect*. The power of randomization is that treatment and control groups are more likely to have a balanced distribution of known and unknown prognostic factors.

Consider again our example of the ω3 fatty acid studies. What was the cause of bias in the ω3 fatty acids observational studies? People who eat larger amounts of ω3 fatty acids may typically have a higher socioeconomic status than those who eat smaller amounts. In addition, patients who eat larger amounts of ω3 fatty acids may eat fewer unhealthy foods and may be more careful with other important risk factors (eg, smoking and exercise). Their apparent benefit from ω3 fatty acids may reflect their healthier lifestyle. Whatever the explanation, we are now confident that it was their previous *prognosis*, rather than the ω3 fatty acids, that led to lower rates of cardiovascular events.

Although randomization is a powerful technique, it does not always succeed in creating groups with similar prognosis. Investigators may make mistakes that compromise randomization, or randomization may fail because of chance—unlikely events sometimes happen. The next 2 sections address these issues.

When those enrolling patients are unaware and cannot control the arm to which the patient is allocated, we refer to randomization as concealed. In unconcealed trials, those responsible for recruitment may systematically enroll sicker—or less sick—patients to either a treatment or control group. This behavior will compromise the purpose of randomization, and the study will yield a biased result.[13-15] Careful investigators will ensure that randomization is concealed through strategies such as remote randomization, in which the individual recruiting the patient makes a call to a methods center to discover the arm of the study to which the patient is assigned.

Consider, for instance, a trial of β-blockers vs angiotensin-converting enzyme (ACE) inhibitors for hypertension treatment that used opaque numbered envelopes to conceal randomization.[16] At the time the study was conducted, evidence suggested that β-blockers were better for patients with heart disease. Significantly more patients with heart disease were assigned to receive β-blockers ($P = .037$). In addition, evidence suggested that ACE inhibitors were better for patients with diabetes mellitus. Significantly more patients with diabetes were assigned to receive ACE inhibitors ($P = .048$). It is possible that clinicians were opening envelopes and violating the randomization to ensure patients received what the clinicians believed was the best treatment. Thus, the prognostic balance that randomization could have achieved was prevented.

Were Patients in the Treatment and Control Groups Similar With Respect to Known Prognostic Factors?

The purpose of randomization is to create groups whose prognosis, with respect to the target outcomes, is similar. Sometimes, through bad luck, randomization will fail to achieve this goal. The smaller the sample size, the more likely the trial will have prognostic imbalance.

Picture a trial testing a new treatment for heart failure that is enrolling patients classified as having New York Heart Association functional class III and class IV heart failure. Patients with class IV heart failure have a much worse prognosis than those with class III heart failure. The trial is small, with only 8 patients. One would not be surprised if all 4 patients with class III heart failure were allocated to the treatment group and all 4 patients with class IV heart failure were allocated to the control group. Such a result of the allocation process would seriously bias the study in favor of the treatment. Were the trial to enroll 800 patients, one would be startled if randomization placed all 400 patients with class III heart failure in the treatment arm. The larger the sample size, the more likely randomization will achieve its goal of prognostic balance.

You can check how effectively randomization has balanced known prognostic factors by looking for a display of patient characteristics of the treatment and control groups at the study's commencement—the baseline or entry prognostic features. Although we will never know whether similarity exists for the unknown prognostic factors, we are reassured when the known prognostic factors are well balanced.

All is not lost if the treatment groups are not similar at baseline. Statistical techniques permit adjustment of the study result for baseline differences. When both *adjusted analyses* and unadjusted analyses generate the same conclusion, clinicians gain confidence that the *risk of bias* is not excessive.

Was Prognostic Balance Maintained as the Study Progressed?

To What Extent Was the Study Blinded?

If randomization succeeds, treatment and control groups begin with a similar prognosis. Randomization, however, provides no guarantees that the 2 groups will remain prognostically balanced. *Blinding* is the optimal strategy for maintaining prognostic balance.

Box 7-2 describes 5 groups involved in clinical trials that, ideally, will remain unaware of whether patients are receiving the *experimental therapy* or control therapy. Patients who take a treatment that they believe is effective may feel and perform better than those who do not, even if the treatment has no biologic activity. Although the magnitude and consistency of this *placebo effect* remain uncertain,[17-20] investigators interested in determining the biologic impact of a treatment will ensure patients are blind to treatment allocation. Similarly, rigorous research designs will ensure blinding of those caring for participants, as well as those collecting, evaluating, and analyzing data (Box 7-2). Demonstrations of bias introduced by unblinding, such as the results of a trial in multiple sclerosis in which a treatment benefit judged by unblinded outcome assessors disappeared when adjudicators of outcome were blinded,[21] highlight the importance of blinding. The more subjectivity involved in judging whether a patient has had a target outcome, the more important blinding becomes. For example, blinding of an outcome assessor is unnecessary when the outcome is all-cause mortality.

BOX 7-2

Five Groups That Should, if Possible, Be Blind to Treatment Assignment

Patients	To avoid placebo effects
Clinicians	To prevent differential administration of therapies that affect the outcome of interest (cointervention)
Data collectors	To prevent bias in data collection
Adjudicators of outcome	To prevent bias in decisions about whether or not a patient has had an outcome of interest
Data analysts	To avoid bias in decisions regarding data analysis

Finally, differences in patient care other than the intervention under study—*cointerventions*—can, if they affect study outcomes, bias the results. Effective blinding eliminates the possibility of either conscious or unconscious differential administration of effective interventions to treatment and control groups. When effective blinding is not possible, documentation of potential cointerventions becomes important.

Were the Groups Prognostically Balanced at the Study's Completion?

It is possible for investigators to effectively conceal and blind treatment assignment and still fail to achieve an unbiased result.

Was Follow-up Complete?

Ideally, at the conclusion of a trial, investigators will know the status of each patient with respect to the target outcome. The greater the number of patients whose outcome is unknown—patients *lost to follow-up*—the more a study is potentially compromised. The reason is that patients who are lost to follow-up often have different prognoses from those who are retained—they may disappear because they have adverse outcomes or because they are doing well and so did not return for assessment.[22] The magnitude of the bias may be substantial. A systematic review suggested that up to a third of positive trials reported in high-impact journals may lose significance given plausible assumptions regarding differential loss to follow-up in treatment and control groups.[23]

When does loss to follow-up pose a serious risk of bias? Although you may run across thresholds such as 20% for a serious risk of bias, such rules of thumb are misleading. Consider 2 hypothetical randomized trials, each of which enters 1000 patients into both the treatment and control groups, of whom 30 (3%) are lost to follow-up (Table 7-1). In trial A, treated patients die at half the rate of the control group (200 vs 400), a *relative risk* (RR) of 50%. To what extent does the loss to follow-up threaten our inference that treatment reduces the death rate by half? If we assume the worst (ie, that all treated patients lost to follow-up died), the number of deaths in the *experimental group* would be 230 (23%). If there were no deaths among the control patients who were lost to follow-up, our best estimate of the effect of treatment in reducing the RR of death decreases from 200/400, or 50%, to 230/400, or 58%. Thus, even assuming the worst makes little difference to the best estimate of the magnitude of the treatment effect. Our inference is therefore secure.

Contrast this with trial B. Here, the RR of death is also 50%. In this case, however, the total number of deaths is much lower; of the treated patients, 30 die, and the number of deaths in control patients is 60. In trial B, if we make the same worst-case assumption about the fate of the patients lost to follow-up, the results would change markedly. If we assume that all patients initially allocated to treatment—but subsequently lost to follow-up—die, the number of deaths among treated patients increases from 30 to 60, which is equal to the number of control group deaths. If this assumption is accurate, we would have 60 deaths in both the treatment and control groups, and the effect of treatment would decrease to 0. Because of this marked change in the treatment effect (50% RR if we ignore those lost to follow-up; 100% RR if we assume all patients in the treatment group who were lost to follow-up died), the 3% loss to follow-up in trial B threatens our inference about the magnitude of the RR.

Of course, this worst-case scenario is unlikely. When a worst-case scenario, were it true, substantially alters the results, you must judge the plausibility of a markedly different outcome *event rate* in the treatment and control group patients lost to follow-up. Ideally, investigators would conduct *sensitivity analyses* to deal with this issue. Because they seldom do, guidelines are available to help you should you choose to make your own judgment of the trial's vulnerability to loss to follow-up.[23]

TABLE 7-1

When Does Loss to Follow-up Seriously Increase Risk of Bias?

	Trial A		Trial B	
	Treatment	Control	Treatment	Control
No. of patients randomized	1000	1000	1000	1000
No. (%) lost to follow-up	30 (3)	30 (3)	30 (3)	30 (3)
No. (%) of deaths	200 (20)	400 (40)	30 (3)	60 (6)
RR not counting patients lost to follow-up	0.2/0.4 = 0.50		0.03/0.06 = 0.50	
RR for worst-case scenario[a]	0.23/0.4 = 0.58		0.06/0.06 = 1	

Abbreviation: RR, relative risk.

[a] The worst-case scenario assumes that all patients allocated to the treatment group and lost to follow-up died and all patients allocated to the control group and lost to follow-up survived.

Thus, loss to follow-up may substantially increase the risk of bias. If assuming a worst-case scenario does not change the inferences arising from study results, then loss to follow-up is unlikely a problem. If such an assumption would significantly alter the results, the extent to which bias is introduced depends on how likely it is that treatment patients lost to follow-up fared badly, whereas control patients lost to follow-up fared well. That decision is a matter of judgment.

Was the Trial Stopped Too Early?

Stopping trials early (ie, before enrolling the planned sample size) when one sees an apparent large benefit is risky and may compromise randomization (see Chapter 11.3, Randomized Trials Stopped Early for Benefit). These *stopped early trials* run the risk of greatly overestimating the treatment effect.[24]

A trial designed with too short a follow-up also may compromise crucial information that adequate length of follow-up would reveal. For example, consider a trial that randomly assigned patients with an abdominal aortic aneurysm to either an open surgical repair or a less invasive, endovascular repair technique.[25] At the end of the 30-day follow-up, mortality was significantly lower in the endovascular technique group (*relative risk reduction* [RRR], 0.61; 95% CI, 0.13-0.82). The investigators followed up participants for an additional 2 years and found that there was no difference in mortality between groups after the first year. Had the trial ended earlier, the endovascular technique may have been considered substantially better than the open surgical technique.

Were Patients Analyzed in the Groups to Which They Were Randomized?

Investigators will undermine the benefits of randomization if they omit from the analysis patients who do not receive their assigned treatment or, worse yet, count events that occur in *nonadherent* patients who were assigned to treatment against the control group. Such analyses will bias the results if the reasons for nonadherence are related to prognosis. In a number of randomized trials, patients who did not adhere to their assigned drug regimens fared worse than those who took their medication as instructed, even after taking into account all known prognostic factors.[26-31] When adherent patients are destined to have a better outcome, omitting those who do not receive assigned treatment undermines the unbiased comparison provided by randomization. Investigators prevent this bias when they follow the *intention-to-treat* principle and analyze all patients in the group to which they were randomized irrespective of what treatment they actually received (see Chapter 11.4, The Principle of Intention to Treat and Ambiguous

Dropouts).[32] Following the intention-to-treat principle does not, however, reduce bias associated with loss to follow-up.[33]

USING THE GUIDE

Returning to our opening clinical scenario, did the experimental and control groups begin the study with a similar prognosis? The study was randomized and allocation was concealed; 212 patients participated and 95% were followed up.[4] The investigators followed the intention-to-treat principle, including all patients they had followed up in the arm to which they were randomized, and stopped when they reached the planned sample size. There were more patients who had occlusive arterial disease (39.6% vs 22.7%) in the ramipril group. This finding could bias the results in favor of the placebo group, and the investigators do not provide an adjusted analysis for the baseline differences. Clinicians, patients, data collectors, outcomes assessors, and data analysts were all blind to allocation.

The final risk of bias assessment represents a continuum from studies that are at very low risk of bias to others that are at very high risk of yielding a biased estimate of effect. Inevitably, where a study lies in this continuum involves some judgment. In this case, despite uncertainty about baseline differences between the groups, we conclude that the risk of bias is low.

What are the Results?

How Large Was the Treatment Effect?

Most frequently, RCTs monitor *dichotomous* outcomes (eg, "yes" or "no" classifications for cancer recurrence, myocardial infarction, or death). Patients either have such an event or they do not, and the article reports the proportion of patients who develop such events. Consider, for example, a study in which 20% of a control group died but only 15% of those receiving a new treatment died (Table 7-2). How might one express these results?

TABLE 7-2

Results From a Hypothetical Randomized Trial

	Outcome, No. of Patients		
Exposure	**Death**	**Survival**	**Total**
Treatment (experimental)	15	85	100
Control	20	80	100

Control group risk (CGR): 20/100 = 20%

Experimental group risk (EGR): 15/100 = 15%

Absolute risk reduction or risk difference: CGR − EGR, 20% − 15% = 5%

Relative risk: EGR/CGR = (15/100)/(20/100) × 100% = 75%

Relative risk reduction: [1 − (EGR/CGR)] × 100% = 1 − 75% = 25%

Abbreviations: CGR, control group risk; EGR, experimental group risk.

One possibility is the absolute difference (known as the *absolute risk reduction* [ARR] or *risk difference*) between the proportion who died in the control group (*control group risk* [CGR]) and the proportion who died in the experimental group (*experimental group risk* [EGR]), or CGR - EGR = 0.20 − 0.15 = 0.05. Another way to express the impact of treatment is as the RR: the risk of events among patients receiving the new treatment relative to that risk among patients in the control group, or EGR/CGR = 0.15/0.20 = 0.75.

The most commonly reported measure of dichotomous treatment effects is the complement of the RR, the RRR. It is expressed as a percentage: 1 − (EGR/CGR) = 100% = (1 − 0.75) × 100% = 25%. An RRR of 25% means that of those who would have died had they been in the control group, 25% will not die if they receive treatment; the greater the RRR, the more effective the therapy. Investigators may compute the RR during a specified period, as in a *survival analysis*; the relative measure of effect in such a time-to-event analysis is called the *hazard ratio* (see Chapter 9, Does Treatment Lower Risk? Understanding the Results). When people do not specify whether they are talking about RRR or ARR—for instance, "Drug X was 30% effective in reducing the risk of death" or "The efficacy of the vaccine was 92%"—they are almost invariably talking about RRR (see Chapter 9, Does Treatment Lower Risk? Understanding the Results).

How Precise Was the Estimate of the Treatment Effect?

We can never be sure of the true risk reduction; the best estimate of the true treatment effect is what we observe in a well-designed randomized trial. This estimate is called a *point estimate* to remind us that, although the true value lies somewhere in its neighborhood, it is unlikely to be precisely correct. Investigators often tell us the neighborhood within which the true effect likely lies by calculating CIs, a range of values within which one can be confident the true effect lies.[34]

We usually use the 95% CI (see Chapter 10, Confidence Intervals: Was the Single Study or Meta-analysis Large Enough?). You can consider the 95% CI as defining the range that—assuming the study has low risk of bias—includes the true RRR 95% of the time. The true RRR will generally lie beyond these extremes only 5% of the time, a property of the CI that relates closely to the conventional level of *statistical significance* of $P < .05$. We illustrate the use of CIs in the following examples.

Example 1

If a trial randomized 100 patients each to experimental and control groups, and there were 20 deaths in the control group and 15 deaths in the experimental group, the authors would calculate a point estimate for the RRR of 25% [CGR = 20/100 or 0.20, EGR = 15/100 or 0.15, and 1 − EGR/CGR = (1 − 0.75) × 100 = 25%]. You might guess, however, that the true RRR might be much smaller or much greater than 25%, based on a difference of only 5 deaths. In fact, you might surmise that the treatment might provide no benefit (an RRR of 0%) or might even do harm (a negative RRR). And you would be right; in fact, these results are consistent with both an RRR of −38% (that is, patients given the new treatment might be 38% more likely to die than control patients) and an RRR of nearly 59% (that is, patients subsequently receiving the new treatment might have a risk of dying almost 60% less than those who are not treated). In other words, the 95% CI on this RRR is −38% to 59%, and the trial really has not helped us decide whether or not to offer the new treatment (Figure 7-1).

Example 2

What if the trial enrolled 1000 patients per group rather than 100 patients per group, and the same event rates were observed as before, so that there were 200 deaths in the control group (CGR = 200/1000 = 0.20) and 150 deaths in the experimental group (EGR = 150/1000 = 0.15)? Again, the point estimate of the RRR is 25% (1 – EGR/CGR = 1 – [0.15/0.20] × 100 = 25%).

In this larger trial, you might think that our confidence that the true reduction in risk is close to 25% is much greater; again, you would be right. The 95% CI on the RRR for this set of results is all on the positive side of 0 and runs from 9% to 41% (Figure 7-1).

These examples show that the larger the sample size and higher the number of outcome events in a trial, the greater our confidence that the true RRR (or any other measure of effect) is close to what we observed. The point estimate—in this case, 25%—is the one value most likely to represent the true RRR. As one considers values farther and farther from the point estimate, they become less and less likely to represent the truth. By the time one crosses the upper or lower boundaries of the 95% CI, the values are unlikely to represent the true RRR. All of this assumes the study is at low risk of bias.

Not all randomized trials have dichotomous outcomes, nor should they. In a study of respiratory muscle training for patients with chronic airflow limitation, one primary outcome measured how far patients could walk in 6 minutes in an enclosed corridor.[35] This 6-minute walk improved from a mean of 406 to 416 m (up 10 m) in the experimental group receiving respiratory muscle training, and 409 to 429 m (up 20 m) in the control group. The point estimate for improvement in the 6-minute walk due to respiratory muscle training therefore was negative, at −10 m (or a 10-m difference in favor of the control group).

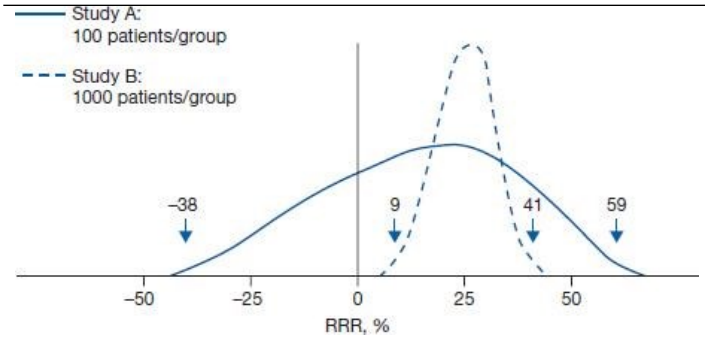
Here, too, you should look for the 95% CIs around this difference in changes in exercise capacity and consider their implications. The investigators tell us that the lower boundary of the 95% CI was −26 (ie, the results are consistent with a difference of 26 m in favor of the control treatment) and the upper boundary was 5 m. Even in the best of circumstances, patients are unlikely to perceive adding 5 m to the 400 recorded at the start of the trial as important, and this result effectively excludes an important benefit of respiratory muscle training as applied in this study.

FIGURE 7-1

Confidence Intervals in Trials of Various Sample Size

Abbreviation: RRR, relative risk reduction.

Two studies with the same point estimate, a 25% RRR, but different sample sizes and correspondingly different CIs. The x-axis represents the different possible RRR, and the y-axis represents the likelihood of the true RRR having that particular value. The solid line represents the CI around the first example, in which there were 100 patients per group, and the number of events in the active and control groups were 15 and 20, respectively. The dashed line represents the CI around the second example, in which there were 1000 patients per group, and the number of events in the active and control groups were 150 and 200, respectively.



Having determined the magnitude and precision of the treatment effect, clinicians can turn to the final question of how to apply the article's results to their patients.

USING THE GUIDE

Using the numbers provided in the article,[4] patients in the ramipril group walked 75 seconds (95% CI, 60-89 seconds) longer without pain than the placebo group and 255 seconds (95% CI, 215-295 seconds) longer overall. The effect of ramipril is convincing given that the 95% CIs are narrow and the lower boundaries are far from showing no effect (ie, 0 seconds). The clinical importance of walking 75 seconds without pain is likely noticeable given that they could walk a mean of 140 seconds without pain at baseline. This finding is consistent with a substantial improvement in a secondary outcome, a measure of health-related quality of life, for patients in the ramipril group.

How Can I Apply the Results to Patient Care?

Were the Study Patients Similar to the Patient in My Practice?

If the patient before you would have qualified for enrollment in the study, you can apply the results with considerable confidence or consider the results *generalizable*. Often, your patient has different attributes or characteristics from those enrolled in the trial and would not have met a study's eligibility criteria. Patients may be older or younger, may be sicker or less sick, or may have comorbid disease that would have excluded them from participation in the study.

A study result probably applies even if, for example, adult patients are 2 years too old for enrollment in the study, had more severe disease, had previously been treated with a competing therapy, or had a comorbid condition. A better approach than rigidly applying the study *inclusion* and *exclusion criteria* is to ask whether there is some compelling reason why the results do not apply to the patient. You usually will not find a compelling reason, in which case you can generalize the results to your patient with confidence.

A related issue has to do with the extent to which we can generalize findings from a study using a particular drug to another closely (or not so closely) related agent. The issue of drug *class effects* and how conservative one should be in assuming class effects remains controversial (see Chapter 28.4, Understanding Class Effects). Generalizing findings of surgical treatment may be even riskier. Randomized trials of carotid endarterectomy, for instance, demonstrate much lower perioperative rates of stroke and death than one might expect in one's own community, which may reflect on either the patients or surgeons (and their relative expertise) selected to participate in randomized trials.[36] An example of how expertise might be considered is provided below.

Expertise in Procedural Interventions

Unlike pharmacologic interventions in which we expect the intervention to vary minimally between patients, procedural interventions may differ substantially based on the expertise of the physician and the technology available to deliver the intervention.

For example, it is suggested that “off-pump” coronary artery bypass surgery reduces the risk of postoperative complications compared with the traditional “on-pump” technique. When the 2 techniques are compared in a randomized trial, one must be careful interpreting the results because of potential differences in expertise. For example, if surgeons participating in the trial are, on average, less skilled with the off-pump technique, the outcomes of patients in the off-pump group may reflect surgeon inexperience more than the true risks and merits of the technique. Further, surgeons may choose to switch from off-pump to on-pump technique more frequently than they would switch from on-pump to off-pump. This will bias the result toward demonstrating no difference between the techniques. One way of preventing these misleading results is by ensuring only surgeons with sufficient expertise in both on-pump and off-pump techniques are allowed to participate in the trial, as was done in the CABG Off or On Pump Revascularization Study (CORONARY) trial.[37] Another method of preventing this differential expertise bias is to randomize patients to a surgeon with expertise in one technique or to a surgeon with expertise in the alternate technique rather than randomize the patient to a surgeon who will perform either procedure to which the patient is randomized.[38]

A final issue arises when a patient fits the features of a subgroup of patients in the trial report. We encourage you to be skeptical of *subgroup analyses*.[39] The treatment is likely to benefit the subgroup more or less than the other patients only if the difference in the effects of treatment in the subgroups is large and unlikely to occur by chance. Even when these conditions apply, the results may be misleading, particularly when investigators did not specify their hypotheses before the study began, if they had a large number of hypotheses, or if other studies fail to replicate the finding.[40]

Were All Patient-Important Outcomes Considered?

Treatments are indicated when they provide important benefits. Demonstrating that a bronchodilator produces small increments in forced expiratory volume in patients with chronic airflow limitation, that a vasodilator improves cardiac output in heart failure patients, or that a lipid-lowering agent improves lipid profiles does not provide sufficient justification for administering these drugs (see Chapter 13.4, Surrogate Outcomes). In these instances, investigators have chosen *substitute outcomes* or *surrogate outcomes* rather than those that patients would consider important. What clinicians and patients require is evidence that the treatments improve outcomes that are important to patients, such as reducing shortness of breath during the activities required for daily living, avoiding hospitalization for heart failure, or decreasing the risk of a major stroke.[41]

Trials of the impact of antiarrhythmic drugs after myocardial infarction illustrate the danger of using substitute outcomes or end points. Because abnormal ventricular depolarizations were associated with a high risk of death and antiarrhythmic drugs demonstrated a reduction in abnormal ventricular depolarizations (the substitute end point), it made sense that they should reduce death. A group of investigators performed randomized trials on 3 agents (encainide, flecainide, and moricizine) that were previously found to be effective in suppressing the substitute end point of abnormal ventricular depolarizations. The investigators had to stop the trials when they discovered that mortality was substantially higher in patients receiving antiarrhythmic treatment than in those receiving placebo.[42,43] Clinicians relying on the substitute end point of arrhythmia suppression would have continued to administer the 3 drugs, to the considerable detriment of their patients (for additional examples of misleading surrogates, see Chapter 11.2, Surprising Results of Randomized Trials).

Even when investigators report favorable effects of treatment on a patient-important outcome, you must consider whether there may be deleterious effects on other outcomes. For instance, cancer chemotherapy may lengthen life but decrease its quality. Randomized trials often fail to adequately document the toxicity or adverse effects of the experimental intervention.[44]

Composite end points represent a final dangerous trend in presenting outcomes (see Chapter 12.4, Composite End Points). Like surrogate outcomes, composite end points are attractive for reducing sample size and decreasing length of follow-up. Unfortunately, they can mislead. For example, a trial that reduced a composite outcome of death, nonfatal myocardial infarction, and admission for an acute coronary syndrome actually demonstrated a trend toward increased mortality with the experimental therapy and convincing effects only on admission for an acute coronary syndrome.[45] The composite outcome would most strongly reflect the treatment effect of the most common of the components, admission for an acute coronary syndrome, even though there is no convincing evidence the treatment reduces the risk of death or myocardial infarction.

Another long-neglected outcome is the resource implications of alternative management strategies. Health care systems face increasing resource constraints that mandate careful attention to *economic analysis* (see Chapter 28.2, Economic Analysis).

Are the Likely Treatment Benefits Worth the Potential Harm and Costs?

If the results of a study apply to your patient and the outcomes are important to your patient, the next question concerns whether the probable treatment benefits are worth the associated *risks*, *burden*, and resource requirements. A 25% reduction in the RR of death may sound impressive, but its impact on your patient may nevertheless be minimal. This notion is illustrated by using a concept called *number needed to treat* (NNT), the number of patients who must receive an intervention of therapy during a specific period to prevent 1 adverse outcome or produce 1 positive outcome.[46]

The impact of a treatment is related not only to its RRR but also to the risk of the adverse outcome it is designed to prevent. One large trial in myocardial infarction suggests that clopidogrel in addition to aspirin reduces the RR of death from a cardiovascular cause, nonfatal myocardial infarction, or stroke by approximately 20% in comparison to aspirin alone.[47] Table 7-3 considers 2 patients presenting with acute myocardial infarction without elevation of ST segments on their electrocardiograms.

In the first case, a 40-year-old man presents with electrocardiographic findings that suggest an inferior myocardial infarction without ST-segment elevation. You find no signs of heart failure; the patient is in normal sinus rhythm, with a rate of 80/min; and he does not have elevated troponin. This individual's risk of death or recurrent myocardial infarction in the next year is estimated to be 5.3%. Compared with aspirin alone, clopidogrel in addition to aspirin would reduce this risk by 20% to 4.2%, an ARR of 1.1% (0.011). The inverse of this ARR (ie, 100 divided by the ARR expressed as a percentage) is equal to the number of such patients we would have to treat to prevent 1 event (ie, 1 death, or recurrent myocardial infarction after a mild myocardial infarction in a low-risk patient), the NNT. In this case, we would have to treat approximately 91 such patients to prevent 1 recurrent myocardial infarction or save 1 life (100/1.1 = 91). Given the small decrease in the outcome of death, recurrent myocardial infarction, or stroke (most noticeably recurrent myocardial infarction) with clopidogrel, the small increased risk of major bleeding associated with clopidogrel, and its additional cost, many clinicians might prefer aspirin alone in this patient.

In the second case, a 70-year-old man presents with electrocardiographic signs of anterior myocardial infarction with pulmonary edema and cardiogenic shock. His risk of dying or having a recurrent myocardial infarction in the subsequent year is approximately 36%. A 20% RRR of death in such a high-risk patient generates an ARR of 7.2% (0.072), and we would have to treat only 14 such individuals to avert a recurrent myocardial infarction or death (100/7.2 = 13.8). Many clinicians would consider clopidogrel in addition to aspirin.

TABLE 7-3

Considerations in the Decision to Treat 2 Patients With Myocardial Infarction With Clopidogrel and Aspirin or Aspirin Alone

	Risk of Death or MI 1 Year After MI With Aspirin Alone (CER)	Risk With Clopidogrel Plus Aspirin (EGR) (ARR = CGR – EGR)	NNT (100/ARR When ARR Is Expressed as a Percentage)
40-year-old man with small MI	5.3%	4.2% (1.1% or 0.011)	91
70-year-old man with large MI and heart failure	36%	28.8% (7.2% or 0.072)	14

Abbreviations: ARR, absolute risk reduction; CER, control event rate; CGR, control group risk; EGR, experimental group risk; MI, myocardial infarction; NNT, number needed to treat.

A key element of the decision to start therapy, therefore, is to consider the patient's risk of the event if left untreated. For any given RRR, the higher the probability that a patient will experience an adverse outcome if we do not treat, the more likely the patient will benefit from treatment and the fewer such patients we need to treat to prevent 1 adverse outcome (see Chapter 9, Does Treatment Lower Risk? Understanding the Results). Knowing the NNT assists clinicians in helping patients weigh the benefits and downsides associated with their management options.

Trading off benefits and risks also requires an accurate assessment of the adverse effects of treatment. Randomized trials with relatively small sample sizes are unsuitable for detecting rare but catastrophic adverse effects of therapy. Clinicians often must look to other sources of information—often characterized by higher risk of bias—to obtain an estimate of the adverse effects of therapy (see Chapter 14, Harm [Observational Studies]).

When determining the optimal treatment choice based on the relative benefits and harms of a therapy, the *values and preferences* of each individual patient must be considered. How best to communicate information to patients and how to incorporate their values into clinical decision making remain areas of active investigation in *evidence-based medicine* (see Chapter 27, Decision Making and the Patient).

CLINICAL SCENARIO RESOLUTION

The study that we identified found an increase in pain-free and total walking time of patients with peripheral arterial disease treated with ramipril compared with placebo.[4] The authors did not describe any harmful effects of ramipril other than more withdrawals due to cough than placebo-treated patients. This finding may leave some uncertainty as to the net benefits to patients. In particular, there is no mention of kidney failure or hyperkalemia-induced cardiac arrest, the most serious adverse effects associated with ramipril. However, there is a large body of literature on patients with other types of vascular disease that suggests that ramipril, at the dose used in this study, is well tolerated and safe, particularly if clinicians monitor patients periodically for the precursors to these adverse effects (ie, changes in kidney function or serum potassium).

Your patient is significantly limited by his intermittent claudication. He is similar to patients included in this study. Given the treatment effect on walking time and the observed effect on health-related quality of life, as well as an apparently minimal side effect profile, the study suggests patient-important benefits to taking ramipril.

The patient finds his limited walking ability and the pain he experiences debilitating. He believes that being able to walk for 1 additional minute would be worthwhile. He is, however, under financial stress and is concerned that ramipril costs $1.20 per pill, or approximately $450 in the next year. You explain that the investigators' choice of medication leaves some doubt about the best drug to use. The investigators could have chosen lisinopril, an ACE inhibitor with marginal differences from ramipril, which the patient can purchase for approximately one-third the price. Ultimately, implicitly accepting a class effect, the patient chooses the lisinopril.

References

1. Momsen AH, Jensen MB, Norager CB, Madsen MR, Vestersgaard-Andersen T, Lindholt JS. Drug therapy for improving walking distance in intermittent claudication: a systematic review and meta-analysis of robust randomised controlled studies. *Eur J Vasc Endovasc Surg*. 2009;38(4):463-–474. [PubMed: 19586783]

2. Wong PF, Chong L-Y, Stansby G. Antiplatelet therapy to prevent cardiovascular events and mortality in patients with intermittent claudication. *JAMA*. 2013;309(9):926-–927. [PubMed: 23462789]

3. Jaar BG. ACP Journal Club. Ramipril improved walking times and QOL in peripheral artery disease and intermittent claudication. *Ann Intern Med*. 2013;158(12):JC7. [PubMed: 23778931]

4. Ahimastos AA, Walker PJ, Askew C, et al. Effect of ramipril on walking times and quality of life among patients with peripheral artery disease and intermittent claudication: a randomized controlled trial. *JAMA*. 2013;309(5):453-–460. [PubMed: 23385271]

5. Hu FB, Bronner L, Willett WC, et al. Fish and omega-3 fatty acid intake and risk of coronary heart disease in women. *JAMA*. 2002;287(14):1815-–1821. [PubMed: 11939867]

6. Kotwal S, Jun M, Sullivan D, Perkovic V, Neal B. Omega 3 fatty acids and cardiovascular outcomes: systematic review and meta-analysis. *Circ Cardiovasc Qual Outcomes*. 2012;5(6):808-–818. [PubMed: 23110790]

7. Bosch J, Gerstein HC, Dagenais GR, et al; ORIGIN Trial Investigators. n-3 fatty acids and cardiovascular outcomes in patients with dysglycemia. *N*

Engl J Med. 2012;367(4):309-–318. [PubMed: 22686415]

8. Bjelakovic G, Nikolova D, Simonetti RG, Gluud C. Antioxidant supplements for prevention of gastrointestinal cancers: a systematic review and meta-analysis. *Lancet*. 2004;364(9441):1219-–1228. [PubMed: 15464182]

9. Miller ER III, Pastor-Barriuso R, Dalal D, Riemersma RA, Appel LJ, Guallar E. Meta-analysis: high-dosage vitamin E supplementation may increase all-cause mortality. *Ann Intern Med*. 2005;142(1):37-–46. [PubMed: 15537682]

10. Ikram H, Crozier IG. Xamoterol in severe heart failure. *Lancet*. 1990;336(8713):517-–518. [PubMed: 1975033]

11. Califf RM, Adams KF, McKenna WJ, et al. A randomized controlled trial of epoprostenol therapy for severe congestive heart failure: The Flolan International Randomized Survival Trial (FIRST). *Am Heart J*. 1997;134(1):44-–54. [PubMed: 9266782]

12. Hampton JR, van Veldhuisen DJ, Kleber FX, et al; Second Prospective Randomised Study of Ibopamine on Mortality and Efficacy (PRIME II) Investigators. Randomised study of effect of ibopamine on survival in patients with advanced severe heart failure. *Lancet*. 1997;349(9057):971-–977. [PubMed: 9100622]

13. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273(5):408-–412. [PubMed: 7823387]

14. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*. 1998;352(9128):609-–613. [PubMed: 9746022]

15. Balk EM, Bonis PA, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA*. 2002;287(22):2973-–2982. [PubMed: 12052127]

16. Hansson L, Lindholm LH, Niskanen L, et al. Effect of angiotensin-converting-enzyme inhibition compared with conventional therapy on cardiovascular morbidity and mortality in hypertension: the Captopril Prevention Project (CAPPP) randomised trial. *Lancet*. 1999;353(9153):611-–616. [PubMed: 10030325]

17. Kaptchuk TJ. Powerful placebo: the dark side of the randomised controlled trial. *Lancet*. 1998;351(9117):1722-–1725. [PubMed: 9734904]

18. Hróbjartsson A, G⊘tzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *N Engl J Med*. 2001;344(21):1594-–1602. [PubMed: 11372012]

19. McRae C, Cherin E, Yamazaki TG, et al. Effects of perceived treatment on quality of life and medical outcomes in a double-blind placebo surgery trial. *Arch Gen Psychiatry*. 2004;61(4):412-–420. [PubMed: 15066900]

20. Rana JS, Mannam A, Donnell-Fink L, Gervino EV, Sellke FW, Laham RJ. Longevity of the placebo effect in the therapeutic angiogenesis and laser myocardial revascularization trials in patients with coronary heart disease. *Am J Cardiol*. 2005;95(12):1456-–1459. [PubMed: 15950570]

21. Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yetisir E, Roberts R. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology*. 1994;44(1):16-–20. [PubMed: 8290055]

22. Ioannidis JP, Bassett R, Hughes MD, Volberding PA, Sacks HS, Lau J. Predictors and impact of patients lost to follow-up in a long-term randomized trial of immediate versus deferred antiretroviral treatment. *J Acquir Immune Defic Syndr Hum Retrovirol*. 1997;16(1):22-–30. [PubMed: 9377121]

23. Akl EA, Briel M, You JJ, et al. Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. *BMJ*. 2012;344:e2809. [PubMed: 22611167]

24. Montori VM, Devereaux PJ, Adhikari NK, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005;294(17):2203-–2209. [PubMed: 16264162]

25. Greenhalgh RM, Brown LC, Powell JT, Thompson SG, Epstein D, Sculpher MJ; United Kingdom EVAR Trial Investigators. Endovascular versus open repair of abdominal aortic aneurysm. *N Engl J Med*. 2010;362(20):1863-–1871. [PubMed: 20382983]

26. The Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N Engl J Med*. 1980;303(18):1038-–1041. [PubMed: 6999345]

27. Asher WL, Harper HW. Effect of human chorionic gonadotrophin on weight loss, hunger, and feeling of well-being. *Am J Clin Nutr*. 1973;26(2):211-–218. [PubMed: 4574032]

28. Hogarty GE, Goldberg SC. Drug and sociotherapy in the aftercare of schizophrenic patients: one-year relapse rates. *Arch Gen Psychiatry*. 1973;28(1):54-–64. [PubMed: 4345664]

29. Fuller R, Roth H, Long S. Compliance with disulfiram treatment of alcoholism. *J Chronic Dis*. 1983;36(2):161-–170. [PubMed: 6337171]

30. Pizzo PA, Robichaud KJ, Edwards BK, Schumaker C, Kramer BS, Johnson A. Oral antibiotic prophylaxis in patients with cancer: a double-blind randomized placebo-controlled trial. *J Pediatr*. 1983;102(1):125-–133. [PubMed: 6336781]

31. Horwitz RI, Viscoli CM, Berkman L, et al. Treatment adherence and risk of death after a myocardial infarction. *Lancet*. 1990;336(8714):542-–545. [PubMed: 1975045]

32. Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ*. 2001;165(10):1339-–1341. [PubMed: 11760981]

33. Alshurafa M, Briel M, Akl EA, et al. Inconsistent definitions for intention-to-treat in relation to missing outcome data: systematic review of the methods literature. *PLoS One*. 2012;7(11):e49163. [PubMed: 23166608]

34. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. *Br Med J (Clin Res Ed)*. 1983;286(6376):1489-–1493. [PubMed: 6405856]

35. Guyatt G, Keller J, Singer J, Halcrow S, Newhouse M. Controlled trial of respiratory muscle training in chronic airflow limitation. *Thorax*. 1992;47(8):598-–602. [PubMed: 1412115]

36. Walker MD, Marler JR, Goldstein M, et al; Executive Committee for the Asymptomatic Carotid Atherosclerosis Study. Endarterectomy for asymptomatic carotid artery stenosis. *JAMA*. 1995;273(18):1421-–1428. [PubMed: 7723155]

37. Lamy A, Devereaux PJ, Prabhakaran D, et al; CORONARY Investigators. Off-pump or on-pump coronary-artery bypass grafting at 30 days. *N Engl J Med*. 2012;366(16):1489-–1497. [PubMed: 22449296]

38. Devereaux PJ, Bhandari M, Clarke M, et al. Need for expertise based randomised controlled trials. *BMJ*. 2005;330(7482):88. [PubMed: 15637373]

39. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med*. 1992;116(1):78-–84. [PubMed: 1530753]

40. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? updating criteria to evaluate the credibility of subgroup analyses. *BMJ*. 2010;340:c117. [PubMed: 20354011]

41. Guyatt G, Montori V, Devereaux PJ, Schünemann H, Bhandari M. Patients at the center: in our practice, and in our use of language. *ACP J Club*. 2004;140(1):A11-–A12. [PubMed: 14711297]

42. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo: The Cardiac Arrhythmia

Suppression Trial. *N Engl J Med*. 1991;324(12):781-–788. [PubMed: 1900101]

43. Rogers W, Epstein A, Arciniegas J, et al; The Cardiac Arrhythmia Suppression Trial II Investigators. Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. *N Engl J Med*. 1992;327(4):227-–233. [PubMed: 1377359]

44. Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA*. 2001;285(4):437-–443. [PubMed: 11242428]

45. Pfisterer M, Buser P, Osswald S, et al; Trial of Invasive versus Medical therapy in Elderly patients (TIME) Investigators. Outcome of elderly patients with chronic symptomatic coronary artery disease with an invasive vs optimized medical treatment strategy: one-year results of the randomized TIME trial. *JAMA*. 2003;289(9):1117-–1123. [PubMed: 12622581]

46. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med*. 1988;318(26):1728-–1733. [PubMed: 3374545]

47. Yusuf S, Zhao F, Mehta SR, Chrolavicius S, Tognoni G, Fox KK; Clopidogrel in Unstable Angina to Prevent Recurrent Events Trial Investigators. Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. *N Engl J Med*. 2001;345(7):494-–502. [PubMed: 11519503]