APPLYING EVIDENCE TO THE CARE OF A PATIENT WITH CARDIOVASCULAR DISEASE: HOW TO DETECT MISLEADING CLAIMS IN CLINICAL RESEARCH REPORTS

Case and Discussion: *Terrence Shaneyfelt, MD, MPH, and Victor M. Montori, MD, MSc*

The standard basis for determining treatment effect is the randomized controlled trial (RCT). RCTs must be designed to reduce the effects of bias on outcomes. However, even when studies are well designed, the interpretation and presentation of the results can be biased and misleading. This could adversely affect a treatment decision, leading a clinician to prescribe—and a patient to take—a therapy that otherwise might not have been considered.

This article uses the case of a patient with a prior myocardial infarction (MI) to apply an evidence-based approach to deciding whether or not the addition of clopidogrel to aspirin is more effective than aspirin alone for reducing the patient's risk for future cardiovascular events. In discussing the evidence-based medicine (EBM) approach to this case, we focus on how to detect misleading claims made by authors in clinical research reports.

Assessing the Clinical Problem

Mrs. Girouard, a 65-year-old woman with known coronary artery disease (CAD), presents to the emergency department (ED) complaining of the recent onset of increasing shortness of breath. Two years earlier, she had a non–ST-segment elevation MI and was discharged after an uncomplicated hospital course on aspirin (81 mg), metoprolol, ramipril, and simvastatin.

Initial evaluation in the ED suggests community-acquired pneumonia as the cause of Mrs. Girouard's new complaints. In the ED, Mrs. Girouard's cardiolo-

gist visits and recommends that, when discharged, she should take clopidogrel in addition to her other prescribed medications to help prevent a recurrent MI. The cardiologist notes that Mrs. Girouard indicated she would be willing to take additional medication if it would help prevent future cardiac events.
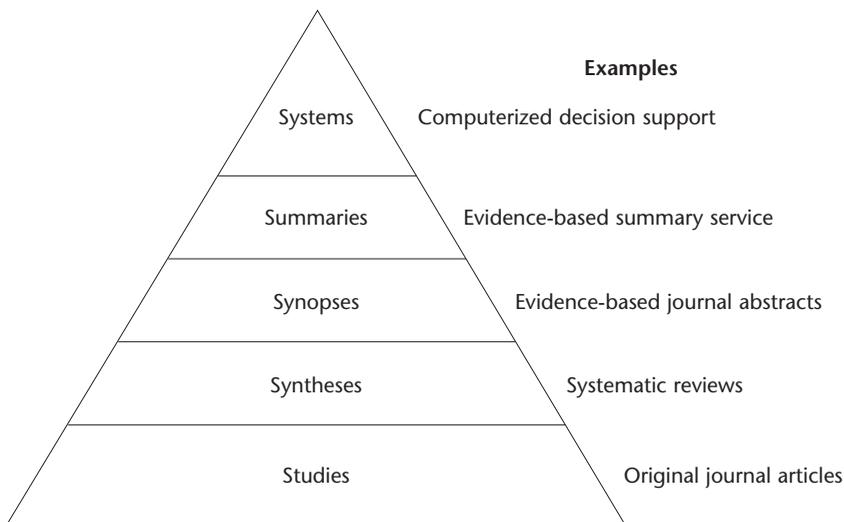
You are a third-year resident who will be admitting Mrs. Girouard to the university's general medical service. You review her ED record and note that her blood pressure and heart rate are well controlled at 118/68 mm Hg and 56 bpm, respectively. Physical examination and chest radiographs confirm right middle lobe pneumonia. During the examination, Mrs. Girouard reports she has not experienced any chest pain even with fairly vigorous exercise, although her exercise capacity has been diminished over the last few days. Laboratory tests and an electrocardiogram confirm that Mrs. Girouard has not had any recent cardiac events. You order appropriate antibiotics, other supporting medications, and continue administering her current cardiac medications.

You present Mrs. Girouard to the ward attending, Dr. Mainieri. He concurs with the treatment plan for Mrs. Girouard's pneumonia. You discuss with Dr. Mainieri whether or not Mrs. Girouard should begin taking clopidogrel as recommended by her cardiologist. In discussing the role of antithrombotic therapy in prevention of cardiovascular disease, you both agree that aspirin has been shown to reduce mortality and recurrence of MI. Furthermore, clopidogrel has shown similar benefits in patients with acute coronary syndromes and following percutaneous coronary intervention, although neither of these factors currently applies to Mrs. Girouard.

You are unclear on whether or not the addition of clopidogrel to aspirin will further reduce Mrs. Girouard's risk of another MI. Dr. Mainieri reports that a recent trial might answer this question and asks you to present the results of that trial on rounds the following morning.

Terrence Shaneyfelt, MD, MPH, Division of General Internal Medicine, Department of Medicine, University of Alabama School of Medicine, and Department of Veterans Affairs Medical Center, Birmingham, AL; and Victor M. Montori, MD, MSc, Division of Endocrinology, Diabetes, Metabolism, and Nutrition, Department of Medicine, Mayo Clinic College of Medicine, Rochester, MN.

Examples

Systems — Computerized decision support

Summaries — Evidence-based summary service

Synopses — Evidence-based journal abstracts

Syntheses — Systematic reviews

Studies — Original journal articles

Figure 1. The "5S" levels of organization of evidence from health care research. (Adapted with permission from Haynes RB. Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based health care decisions [editorial]. ACP J Club 2006; 145:A8.)

In this case, we face the common scenario of a patient for whom secondary prevention of cardiovascular disease is clinically indicated. Although Mrs. Girouard appears to be doing well on her current treatment regimen, every clinical encounter provides us with an opportunity to assess whether we can improve the quality of health care we are delivering. At this visit, an important new issue has been raised: can we optimize Mrs. Girouard's cardiovascular care and further reduce her risk of future cardiovascular events? This case highlights several issues that should be considered when assessing a clinical problem, including our learning needs (how accurate, complete, and up-to-date is our knowledge?), the state of the medical literature (is there compelling new evidence on the topic?), and the patient's preferences, values, and needs (what does our patient value most, and how can we optimize her health?).

Asking a Focused Clinical Question

While awaiting your next admission, you set out to find the study Dr. Mainieri suggested. You estimate you have 30 to 45 minutes before needing to return to the ED to admit your next patient.

Before searching the medical literature, it is often helpful to develop a focused clinical question using the PICO format (patient/problem, intervention or diagnostic test of interest, comparison intervention, and outcome). Developing a focused clinical question can help make our search for evidence more efficient and, when more than 1 study exists on a topic, can help guide us to the study that best answers our question. The PICO components for this case may be defined as:

P = adults with previous MI
I = clopidogrel in addition to aspirin
C = aspirin alone
O = recurrent MI or death

Using this framework, our question regarding Mrs. Girouard would be:

In **adults with previous MI**, does the **addition of clopidogrel to aspirin**, compared with **aspirin alone**, reduce the risk of **recurrent MI or death**?

Acquiring the Current Best Evidence

Several medical textbooks are lying around the on-call room but are unlikely to contain any recent study data. Applying the "5S pyramid" for selecting evidence resources (**Figure 1**) [1], you recognize that no first-level resource (system of evidence) is available, as your hospital does not have an evidence-integrated electronic medical record system, so you move to the second level (evidence summaries). The hospital library subscribes to *Clinical Evidence*, and a search of this online resource reveals 1 RCT on the addition of clopidogrel to aspirin; however, the patients studied were those with acute coronary syndromes [2].

You move to the next level (synopses of evidence) and open a link to *ACP Journal Club*. Entering the terms "clopidogrel and myocardial infarction" into the search field yields 38 results, one of which is entitled "Clopidogrel plus aspirin did not differ

from aspirin alone for reducing MI, stroke, and CV death in high-risk atherothrombosis" [3]. You print the structured abstract [3] and commentary [4] for future reference and then read both.

The synopsis is of the CHARISMA (Clopidogrel for High Atherothrombotic Risk and Ischemic Stabilization, Management, and Avoidance) trial [5]. According to the conclusion in the structured abstract, the study showed that in patients with or at high risk for vascular disease, "long-term treatment with clopidogrel plus aspirin did not differ from aspirin alone for reducing the composite endpoint of myocardial infarction, stroke, and cardiovascular death" [3]. The expert commentary, however, indicates that "dual therapy provided no benefit for patients with multiple risk factors alone but a small risk reduction in the primary endpoint for patients who had established vascular disease (6.9% versus 7.9%)" [4]. Because Mrs. Girouard has CAD, you print the full text of the study to review in detail after confirming that this is the study to which Dr. Mainieri had referred.

The "5S" Approach to Seeking Evidence

The latest conceptual model for organizing evidence from health care research includes 5 levels (Figure 1) [1]. Our search for evidence should begin at the top and end at the highest-available level.

At the top of the pyramid are *systems* of evidence in which research evidence is integrated with the electronic medical record. Unfortunately, most practitioners do not have access to this level of evidence. As a result, most searches will begin with clinical topic *summaries* that integrate the best evidence from the lower levels of the pyramid on that topic. An example would be *Clinical Evidence*, which summarizes the current state of knowledge about prevention and treatment of common clinical conditions, based on a thorough appraisal of the medical literature. Evidence at the next 3 levels of the pyramid—*synopses* (eg, a structured abstract in *ACP Journal Club)*, *syntheses* (eg, a Cochrane systematic review), and *studies* (eg, an original research study found using the Clinical Queries function on PubMed)—typically addresses a single clinical question about a specific health concern (eg, does metformin lower the risk for cardiovascular complications in patients with diabetes?).

This case illustrates an important advantage of identifying evidence at a higher level of the pyramid, in this instance a synopsis (level 3). *ACP Journal Club*, published bimonthly by the American College of Physicians, requires that included articles meet

Table 1. Criteria for Appraising and Applying an Individual Therapy Study

---

Are the results valid?

Did experimental and control groups begin the study with a similar prognosis?
- Were patients randomized?
- Was randomization concealed (blinded or masked)?
- Were patients analyzed in the groups to which they were randomized?
- Were patients in the treatment and control groups similar with respect to known prognostic variables?

Did experimental and control groups retain a similar prognosis after the study started?
- Were patients aware of group allocation?
- Were clinicians aware of group allocation?
- Were outcome assessors aware of group allocation?
- Was follow-up complete?

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 patients?
- Were all clinically important outcomes considered?
- Are the likely treatment benefits worth the potential harm and costs?

Adapted with permission from Guyatt G, Rennie D, editors. Users' guides to the medical literature: a manual for evidence-based clinical practice. The Evidence-Based Medicine Working Group. Chicago: AMA Press; 2002:58.

strict methodologic standards. In the case of therapeutic studies, participants must be randomly allocated to comparison groups and achieve at least 80% follow-up, and outcome measures must be clinically important. Because of these strict methodologic standards, readers of *ACP Journal Club* can feel secure that included articles do not contain serious methodologic flaws. Thus, the step of critical appraisal can be skipped, saving valuable time.

Appraising Evidence for Validity and Importance

Because the synopsis of the CHARISMA trial was published in *ACP Journal Club*, you do not need to critically appraise the full-text copy of the study using the criteria for therapy articles recommended in the *Users' Guides to the Medical Literature* (**Table 1**) [6]. However, you read the study methods and results

Table 2. Composite and Individual End Points of the CHARISMA Trial

End Point	Clopidogrel plus Aspirin (N = 7802) n (%)	Placebo plus Aspirin (N = 7801) n (%)	Relative Risk (95% CI)	P Value
Efficacy end points				
Primary efficacy end point	534 (6.8)	573 (7.3)	0.93 (0.83–1.05)	0.22
Death from any cause	371 (4.8)	374 (4.8)	0.99 (0.86–1.14)	0.90
Death from cardiovascular causes	238 (3.1)	229 (2.9)	1.04 (0.87–1.25)	0.68
Myocardial infarction (nonfatal)	146 (1.9)	155 (2.0)	0.94 (0.75–1.18)	0.59
Ischemic stroke (nonfatal)	132 (1.7)	163 (2.1)	0.81 (0.64–1.02)	0.07
Stroke (nonfatal)	150 (1.9)	189 (2.4)	0.79 (0.64–0.98)	0.03
Secondary efficacy end point*	1301 (16.7)	1395 (17.9)	0.92 (0.86–0.995)	0.04
Hospitalization for unstable angina, transient ischemic attack, or revascularization	866 (11.1)	957 (12.3)	0.90 (0.82–0.98)	0.02
Safety end points				
Severe bleeding	130 (1.7)	104 (1.3)	1.25 (0.97–1.61)	0.09
Fatal bleeding	26 (0.3)	17 (0.2)	1.53 (0.83–2.82)	0.17
Primary intracranial hemorrhage	26 (0.3)	27 (0.3)	0.96 (0.56–1.65)	0.89
Moderate bleeding	164 (2.1)	101 (1.3)	1.62 (1.27–2.08)	< 0.001

CI = confidence interval. (Adapted with permission from Bhatt DL, Fox KA, Hacke W, et al. Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. CHARISMA Investigators. N Engl J Med 2006;354:1712. Copyright © 2006 Massachusetts Medical Society. All rights reserved.)

*The secondary efficacy end point was the first occurrence of myocardial infarction, stroke, death from cardiovascular causes, or hospitalization for unstable angina, a transient ischemic attack, or a revascularization procedure (coronary, cerebral, or peripheral).

to be better prepared to discuss the study at rounds the next morning.

The CHARISMA trial was a multicenter, blinded, randomized, placebo-controlled trial comparing the efficacy and safety of clopidogrel plus aspirin to aspirin alone in patients at high risk for a cardiovascular event [5]. The study involved 15,603 patients with 1 of the following: multiple atherothrombotic risk factors, coronary disease, cerebrovascular disease, or symptomatic peripheral arterial disease. Patients were randomized to clopidogrel plus aspirin or matching placebo plus aspirin, and patient follow-up was 99.5%. The primary efficacy end point was a composite of MI, stroke, or cardiovascular death. Bleeding was also monitored.

Satisfied with the methodology, you proceed to consider the results of the study, which are summarized in a table (**Table 2**). The addition of clopidogrel to aspirin did not reduce the primary efficacy end point when compared with aspirin; however, clopidogrel did significantly reduce the risk of nonfatal stroke by 21% (relative risk [RR], 0.79 [95% confidence interval {CI}, 0.64–0.98]). This was offset by a significant 62% increase in the risk of bleeding when clopidogrel was added to aspirin (RR, 1.62 [95% CI, 1.27–2.08]).

What Do the Results Mean?

An in-depth discussion of relative risk reduction (RRR) and number needed to treat (NNT) is beyond the scope of this article and is provided in an earlier article in this series [7]. Briefly, RRR expresses the proportional reduction in the rate of a bad outcome, in this case stroke, by the experimental therapy (clopidogrel plus aspirin) compared with the control therapy (aspirin alone). It is easily calculated from the relative risk (RR) using the following formula: RRR = 100% − RR. In this case, there is a 21% reduction in the risk of stroke when clopidogrel is added to aspirin (RRR = 100% − 79% = 21%). Unfortunately, RRR can be misleading because it expresses a proportional reduction and not an absolute reduction in risk of an adverse event. In addition, the clinical importance of the RRR depends on the patient's baseline risk of developing an outcome. For example, a drug that reduces a patient's risk of stroke from 30% to 15%

(RRR of 50%) is clinically important, whereas one that reduces the patient's risk of stroke from 0.05% to 0.025% (also an RRR of 50%) is not.

NNT is a useful measure of the burden of care for a particular intervention, in that it tells us how many patients will need to receive the new intervention to see 1 additional success compared with the alternative treatment. NNT is the inverse of the absolute risk reduction (ARR) or 100% divided by the ARR, where ARR is equal to the absolute difference in outcome rates between the treatment and control groups (ie, experimental event rate [EER] − control event rate [CER]). Using results data from the CHARISMA trial (Table 2), the rate of stroke was 1.9% in the clopidogrel plus aspirin group and 2.4% in the aspirin group. Thus, the ARR was |1.9% − 2.4%|, which equals 0.5%, and the NNT for stroke (100% divided by 0.5%), was 200, meaning 200 patients would need to be given clopidogrel in addition to aspirin for the duration of the trial to prevent 1 additional stroke.

When compared with the number needed to harm (NNH), NNT can be used to understand the trade-offs between benefits and harms of a therapy. NNH, which is calculated when a bad outcome occurs more frequently in the experimental group, is 100% divided by the absolute risk increase (ARI), which is calculated in the same fashion as the ARR. Since clopidogrel increased the risk of bleeding compared with aspirin alone, we need to calculate the NNH for bleeding. Again using results data from the CHARISMA trial (Table 2), the rate of moderate bleeding was 2.1% in the clopidogrel plus aspirin group and 1.3% in the aspirin alone group. Thus, the ARI was 2.1% − 1.3%, which equals 0.8%, and the NNH for moderate bleeding (100% divided by 0.8%) was 125. In this case, the NNT for stroke was 200, whereas the NNH for moderate bleeding was 125. Thus, we expect that for every stroke clopidogrel prevents, it will cause 1.6 excess moderate bleeding events (200 divided by 125).

Unfortunately, there is no rule of thumb for what constitutes a good NNT, but the smaller the better. Conversely, for NNH, the larger the better. When comparing NNTs and NNHs, we must take into account the outcomes and their relative consequences. We also must consider patient values and costs. For example, we may be willing to cause 10 patients to have moderate bleeding to prevent 1 patient from suffering a stroke [8].

In the results section of the CHARISMA study, you note that the authors performed several subgroup analyses to see if there were any differential effects of
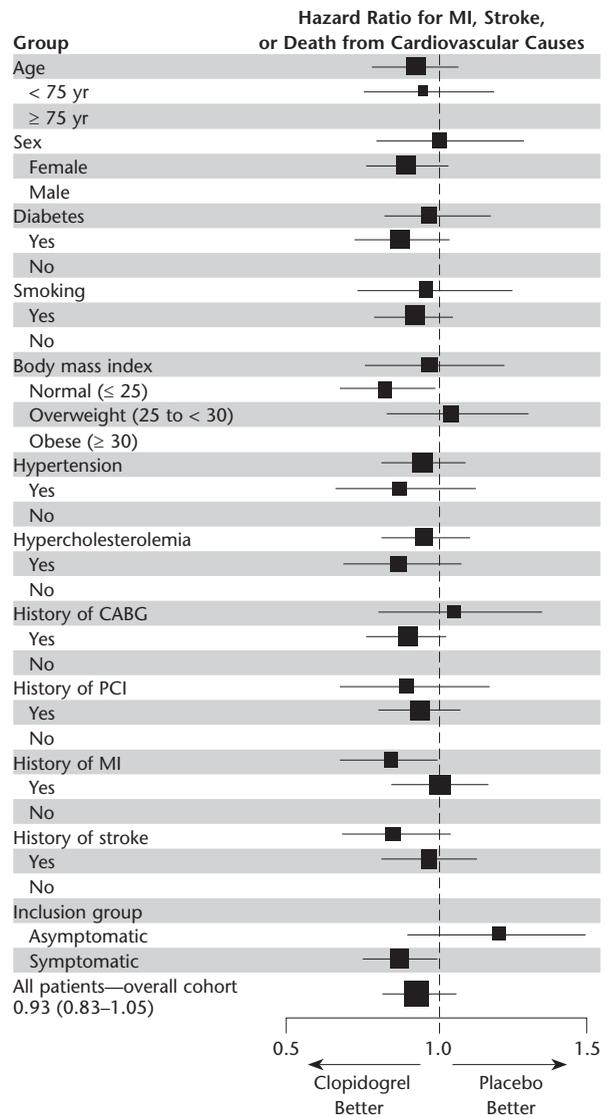
Figure 2. Results of subgroup analysis in the CHARISMA trial. CABG = coronary artery bypass grafting; MI = myocardial infarction; PCI = percutaneous coronary intervention. (Adapted with permission from Bhatt DL, Fox KA, Hacke W, et al. Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. CHARISMA Investigators. N Engl J Med 2006;354:1714. Copyright © 2006 Massachusetts Medical Society. All rights reserved.)

clopidogrel in different categories of patients [5]. In a figure showing the results of the subgroup analyses (**Figure 2**), you notice there are no significant differences in the effect of clopidogrel plus aspirin in any of the subgroups except the group labeled as "symptomatic." Reading further in the results section, you note that in this subgroup, defined as those patients enrolled on the basis of established

cardiovascular disease, clopidogrel plus aspirin resulted in a statistically significant reduction in the primary efficacy outcome compared with aspirin alone, with an RR of 0.88 (95% CI, 0.77–0.998). However, for asymptomatic patients, defined as those enrolled on the basis of multiple atherothrombotic risk factors, clopidogrel plus aspirin resulted in a nonsignificant increase in the primary outcome compared with aspirin alone, with a RR of 1.2 (no 95% CI was supplied).

In the discussion section of the article, the authors conclude that "clopidogrel was beneficial with respect to the primary efficacy end point in patients who were classified as symptomatic for the purposes of the trial (ie, who were enrolled because of a documented history of established vascular disease)" [5]. Based on this conclusion, you decide that Mrs. Girouard, having had a previous MI, would benefit from the addition of clopidogrel to her cardiovascular treatment regimen. Confident that you fully understand the findings of the CHARISMA trial and how they apply to Mrs. Girouard, you go to the ED to admit your next patient.

On rounds the next day, Dr. Mainieri asks what you decided to do about the addition of clopidogrel to Mrs. Girouard's treatment regimen. You summarize the results of the CHARISMA trial for the team, demonstrating how to calculate RR, RRR, NNT, and NNH. You conclude based on the subgroup analyses performed by the study authors that Mrs. Girouard would benefit from the addition of clopidogrel. Dr. Mainieri congratulates you on your EBM skills but suggests you may have been misled by the authors' conclusions on the benefits of clopidogrel in "symptomatic" patients. He assigns an article [9] for the team to read and asks that, after you read the article, you again present the results of the CHARISMA trial.

Strategies to Guard Against Misleading Interpretations

The *Users' Guides* [6] were developed to detect methodologic weaknesses that can introduce bias into a study. Unfortunately, these guides cannot help us detect misleading interpretations of methodologically sound studies. Montori and colleagues [9] have proposed 6 strategies we can apply to help detect biased or misleading inferences when reading original research reports. As we review each of these strategies, we will consider how the CHARISMA trial measures up.

- **Read only the methods and results sections.**

In the discussion section of research reports, the authors summarize the results of the study and try to place them into the context of other research in the field. In this section, study authors offer explanatory mechanisms and identify remaining unresolved issues. Unfortunately, authors often use the discussion section to opine about the importance of their findings and to temper any unexpected findings. Their assertions about the study findings might be influenced by faulty interpretations of point estimates and CIs, *a priori* beliefs, variable cost considerations, and conflicts of interest. Financial relationships among industry, academia, and researchers are widespread and carry the risk of conflicts of interest [10]. Overall, studies funded by the pharmaceutical industry have been found to be 4 times more likely to report results favorable to the company than studies funded by others [11]. Thus, it is easy to see that science is often not objective [12]. Evidence also indicates that it is 5 times more likely that the discussion section will conclude with enthusiasm toward the experimental intervention when for-profit agents fund the trial versus not-for-profit organizations [13].

To guard against being misled, we should read only the methods and results sections and come to our own dispassionate assessment of the findings. The authors of the CHARISMA trial assert in the discussion that "clopidogrel was beneficial with respect to the primary efficacy end point in patients who were classified as symptomatic for the purposes of the trial (ie, who were enrolled because of a documented history of established vascular disease)" [5]. This assertion is based on a subgroup analysis that is flawed, as will be discussed below.

- **Read the abstract reported in preappraised evidence resources.**

Secondary, or preappraised, evidence resources (eg, *ACP Journal Club*, *Evidence-Based Medicine*, *Evidence-Based Mental Health*, and *Evidence-Based Cardiovascular Medicine,* among others) publish structured summaries of studies published elsewhere. A key advantage of these publications is that clinicians and methodologic experts objectively do the work of critical appraisal to ensure that all included articles meet strict methodologic criteria. Thus, we can feel secure that articles included in these resources contain no serious methodologic flaws. Of further value is the commentary that typically accompanies the structured summary. Here, an expert clinician without financial or personal interests objectively places the study into the context of what is already known, discusses any methodologic problems that might affect the interpretation of

the study, and comments on how the results of the study might be applied to clinical care.

The titles of the structured summaries give us a quick glimpse of the overall outcome of the study. For example, the title of the CHARISMA trial summary in *ACP Journal Club*—"Clopidogrel plus aspirin did not differ from aspirin alone for reducing MI, stroke, and CV death in high-risk atherothrombosis" [3]—highlights the lack of effectiveness of clopidogrel. The accompanying commentary notes that the subgroup of patients with established vascular disease had a small reduction in risk of the primary end point but cautions readers about the limitations of subgroup analyses [4].

- **Beware of faulty comparators.**

When making a comparison between a treatment group and a control group, we want a fair comparison. If an effective treatment agent exists for a condition, that drug (not a placebo) should be used as the comparator in a study of the efficacy or effectiveness of a new drug. For example, it would be unreasonable (and unethical) to do a placebo-controlled trial of a new antihypertensive agent, because effective therapy exists for hypertension (eg, diuretics). This approach was taken by the RENAAL (Reduction of Endpoints in NIDDM with the Angiotensin II Antagonist Losartan) study investigators [14]. In RENAAL, patients with diabetic nephropathy were randomly assigned to losartan or placebo in addition to conventional antihypertensive treatment (calcium channel antagonists, diuretics, α blockers, β blockers, and centrally acting agents). Not surprisingly, losartan reduced the incidence of doubling of the baseline serum creatinine concentration, end-stage renal disease, and proteinuria compared with placebo. The authors concluded that losartan conferred significant renal benefits in patients with type 2 diabetes and nephropathy. Although this conclusion is accurate, it is misleading because angiotensin-converting enzyme inhibitors have been shown to effectively reduce these same outcomes in patients with diabetic nephropathy; thus, angiotensin-converting enzyme inhibitors should have been used as the comparator in the RENAAL trial.

If an active comparator (drug with proven effectiveness) is used in a study, we should make sure that the dose, formulation, and administration are optimal. For example, in the Treat-to-Target (3T) trial, patients were randomly assigned to atorvastatin 20 mg or simvastatin 20 mg; the outcome of the study (percent of patients reaching low-density lipoprotein goal at 8 weeks) was achieved in 45% of patients assigned to atorvastatin compared with only 24% of those assigned to simvastatin [15]. The authors concluded that atorvastatin 20 mg was significantly more effective than simvastatin 20 mg. Although this conclusion is accurate, it is misleading because the dose of simvastatin that is equivalent to 20 mg of atorvastatin is 40 mg, not 20 mg (the dose used in the study). Had the 3T investigators used 40 mg of simvastatin, they may have seen similar reductions in low-density lipoprotein. Use of unfair comparators is a major reason why studies have shown that industry-sponsored studies typically yield larger treatment effects (favoring the sponsor's product) than studies funded by not-for-profit agencies [11].

In the CHARISMA trial, patients were randomized to clopidogrel 75 mg plus aspirin 75 to 165 mg/day or to matching placebo plus aspirin [5]. Aspirin is a standard, proven therapy for both primary and secondary prevention of cardiovascular disease, and the doses used are those recommended by evidence-based guidelines [16]. Thus, there is no evidence that the CHARISMA investigators used a faulty comparator.

- **Beware of composite end points.**

The outcomes, or end points, measured in a trial have important implications for the design of a trial. Investigators typically designate one of the end points as the *primary outcome*—the outcome that will be used to determine if the treatment or intervention is effective. During the design phase of a trial, investigators determine how many people will need to be included in the trial, the sample size, based in part on the effect they anticipate the intervention or treatment will have on the primary outcome. Outcome events that occur infrequently require larger sample sizes to detect, whereas frequent events do not require as many patients. Investigators will often use composite outcomes to help lower the sample size needed for the trial. This is beneficial to investigators because the more patients who need to be followed in a trial, the more difficult and costly the trial will be to conduct.

A composite end point is an outcome of a study in which multiple end points are combined into a single outcome. For example, in the CHARISMA trial, the primary efficacy end point was a composite end point that included the first occurrence of MI, stroke, or death from a cardiovascular cause [5]. If a patient enrolled in the CHARISMA trial experienced any 1 of these (or even more than 1), that patient counted as having the primary end point. Table 2 shows the number of patients in each

arm of the CHARISMA trial who had the primary end point and each of its component end points. The primary efficacy end point occurred in 534 (6.8%) of the patients assigned to clopidogrel plus aspirin compared with 573 (7.3%) of patients assigned to aspirin alone, resulting in a nonsignificant RR of 0.93 (95% CI, 0.83–1.05). Because the table also shows the number of patients who experienced each component of the primary efficacy end point, we are able to determine which components were affected. Although the composite end point was not significantly reduced, nonfatal stroke was significantly reduced by 21% in patients assigned to clopidogrel plus aspirin. The other components of the primary efficacy end point, death from cardiovascular causes and MI, were not significantly affected.

As noted, investigators use composite end points to reduce the sample size required for the study. Furthermore, the use of composite end points is usually justified based on the assumption that the intervention will have equal effects on each of the individual components of the composite end point and that each of the components has equal importance [17]. Unfortunately, this is often not the case. When encountering a composite end point in a trial, we should ask ourselves 3 questions [18]:

- *Are the component end points of similar importance to patients?* If patients consider all the components to be equally important, it does not matter how the benefit is distributed among them.

- *Did the more and less important end points occur with similar frequency?* If most of the effect on the composite end point comes from the accumulation of events in the least important outcome, we will be left in the dark about the effect of treatment on the most important outcomes.

- *Are the component end points likely to have similar RRRs?* Composite end points should be constructed based on similar biologic effects of treatment on the individual components. If we see similar RRRs, we can feel more comfortable using the composite end point.

If the answer to most of these questions is no, we will not be able to rely on the composite end point and will need to consider each of the individual components separately to determine the effect of therapy.

Turning to the CHARISMA trial, we can postulate that patients would likely view death and stroke (espe-cially if disabling) as being almost equally important, whereas MI may not be viewed as being as important as the other 2 outcomes. In all, the gradient of patient importance across the components seems moderate. Looking at Table 2, we can see that, in patients assigned to clopidogrel plus aspirin, MI and stroke occurred with similar frequency (1.9% and 1.7%, respectively), but death from cardiovascular causes occurred more frequently (3.1%). Because the most important outcome (death from cardiovascular causes) was more common, this gradient in end points is both relatively minor and not too problematic. Finally, the RRRs for the component end points differed greatly (+4% for death from cardiovascular causes, –6% for MI, –21% for stroke), but these estimates lacked precision (ie, had wide CIs), limiting our confidence that indeed there is a gradient effect. Thus, we should feel relatively comfortable using the composite end point of the CHARISMA trial in clinical decision making.

- **Beware of small treatment effects.**

Small treatment effects occur when the point estimate is close to no effect (RR or odds ratio close to 1; RRR or ARR close to 0) or the CI includes values close to no effect. With a large enough study, even a difference between 2 groups that is so small as to be clinically trivial can be highly statistically significant. Thus, we should always ask ourselves whether statistically significant changes seen in a study are clinically important. Examining Table 2, we can see a good example of this. The RR of the secondary efficacy end point is 0.92 (95% CI, 0.86–0.995). The point estimate (0.92) is close to no effect (RR of 1) as is the upper limit of the 95% CI (0.995). Despite this small apparent treatment effect, the *P* value is significant at 0.04.

Small treatment effects can be made to look larger than they are by presenting the result as a RRR instead of an ARR. From Table 2, 1.9% of patients assigned to the clopidogrel plus aspirin group had a nonfatal stroke compared with 2.4% of those assigned to aspirin alone, resulting in a RR of 0.79 (95% CI, 0.64–0.98). The RRR is 21% (100% – RR), which appears much more impressive than the ARR of 0.5% (|1.9% – 2.4%|).

- **Beware of subgroup analyses.**

Subgroup analyses are undertaken to determine whether a treatment had different effects in certain groups of patients enrolled in a study, for example women compared with men or diabetic compare with nondiabetic patients. Seemingly, subgroup analyses would be useful in that they could help us better decide if therapy applies to an individual patient. For

example, for Mrs. Girouard, we would like to know if clopidogrel benefited the subgroup of women with previous MI, even though the CHARISMA trial showed no overall benefit [5]. Unfortunately, subgroup analyses are often misleading or misinterpreted. These analyses typically are undertaken when the study shows no overall benefit. Thus, we are left to wonder if a subgroup effect is real or just a spurious result of a data-dredging exercise.

Common problems with subgroup comparisons. Subgroup comparisons may be fraught with problems. First, these comparisons are usually nonrandomized. Subdividing a randomized population into smaller groups will likely result in imbalances in prognostic factors among the subgroups. This can be avoided if subgroups are planned in the design phase of the study and randomization is stratified based on an important subgroup. Unfortunately, randomization can be stratified on only a couple of factors at most. Second, because subgroups are smaller than the overall study population, they are often underpowered to detect differential treatment effects. Thus, they often demonstrate no beneficial effect in the subgroups of interest (even if a beneficial effect truly exists, a false-negative finding) or they result in wide nonsignificant CIs around a subgroup treatment effect estimate. Third, a fairly large number of subgroup analyses are often undertaken, increasing the probability of false-positive or chance findings in which an ineffective or even harmful treatment is deemed beneficial for some patients. To reduce the risk of these errors, methodologists recommend that the subgroup analyses be limited in number and pre-specified at the start of the trial.

Faulty statistical analysis. The statistical analysis of subgroup treatment effects is often done incorrectly [19]. Detailed discussion of statistical methods of subgroup analyses is beyond the scope of this article, but proving that treatment effects differ among patient subgroups is not possible with separate analyses of treatment effects in each subgroup. This method, while commonplace, is fundamentally misleading because it uses a within-subgroup comparison to draw inferences about between-subgroup differences. Instead, there should be a formal test of interaction between the subgroup characteristic of interest and the treatment effect size. Interactions can be quantitative or qualitative. In quantitative interactions, treatment is beneficial in both subgroups of patients (eg, men and women), but the magnitude of benefit is greater in one subgroup than the other (eg, more effective in men than in women). In qualitative interactions,

treatment is found to be beneficial in one subgroup but ineffective or harmful in the other. Qualitative interactions are often spurious, and interpretation rests as much on biologic and epidemiologic plausibility as on any statistical interaction test. Another common statistical error made in subgroup analyses is the lack of adjustment for multiple statistical comparisons. When a statistical test is conducted, the false-positive (or type I error) rate is usually set by convention at 0.05, meaning we accept a 5% probability that a positive finding will be due to chance alone. From this, we consider a finding statistically significant when the P value (false-positive or type I error rate) is less than 0.05. When multiple interaction tests (or any statistical tests for that matter) are conducted, each has a 5% probability of finding a treatment effect purely due to chance. If we perform multiple tests, the chance of a false-positive finding increases. (This probability can be estimated using the following formula: chance of a false-positive finding = $1 - 0.95^x$, where x is the number of tests being performed.) One way to correct for the higher risk of a false-positive finding is to apply a stricter P value (ie, less than 0.05) criterion to signify statistical significance. An estimate of this stricter P value criterion is given by the formula: $0.05/x$, where x is the number of statistical tests undertaken. For example, if 20 subgroup analyses are performed, the P value signifying statistical significance of any one test of interaction would be 0.05 divided by 20, or 0.0025. Any P value greater than this, even if less than the conventional 0.05, would be considered nonsignificant. Readers are referred to the *Users' Guides* for a more detailed discussion on interpreting the results of subgroup analyses [6].

Subgroup analysis in the CHARISMA trial. The authors of the CHARISMA trial undertook 20 subgroup analyses, the results of which are shown in Figure 2 [5]. Results of subgroup analyses usually are presented graphically. Individual subgroups are listed on the left side of the figure, and the result of each subgroup analysis is represented as a box, with the 95% CI or precision of that result represented by the horizontal lines extending outward from the box. The size of the box is proportional to the number of patients in the subgroup. Examining Figure 2, the only subgroup in which there was a statistically significant reduction in the primary efficacy outcome was the subgroup of patients labeled "symptomatic" (hazard ratio for the primary outcome, 0.88 [95% CI, 0.77–0.998]; $P = 0.046$). These were patients who had clinically evident cardiovascular, cerebrovascular, or peripheral

vascular disease. This finding is somewhat encouraging because Mrs. Girouard had a previous MI and would have been classified as "symptomatic" had she been enrolled in this study. However, considering the previously noted cautions about subgroup analyses, we must ask ourselves if this is a true finding or just a spurious association. What is the chance that the beneficial effect seen in the symptomatic subgroup was due to chance alone? Using the given formula $(1 - 0.95^x)$, the probability of finding at least 1 false-positive or spurious beneficial result out of 20 subgroup analyses is 64%. The authors did not adjust for the multiple statistical comparisons done in the 20 subgroup analyses (and admit this in the statistical analysis section of the methods). Thus, the resulting significant *P* value of 0.045 for the interaction between the effect in symptomatic and asymptomatic patients is misleading. In fact, the subgroup analyses were so misleading that the *New England Journal of Medicine* published an editorial discussing the limitations of the CHARISMA subgroup analyses [20]. What would the *P* value be had it been corrected for multiple comparisons? When *x* independent interaction tests are performed, the corrected *P* value is given by the formula: $1 - (1 - p)x$, where *p* is the uncorrected *P* value. In this case, the corrected *P* value is $1 - (1 - 0.045)20$, which equals 0.60. Thus, the finding of a beneficial effect of clopidogrel in the symptomatic subgroup of patients is not statistically significant and represents a spurious finding.

In summary, subgroup analyses are commonly undertaken, especially when the overall trial result is not significant. Readers can feel more comfortable believing the results of subgroup analyses when they are planned *a priori*, there is biologic plausibility that the treatment effect might differ in one subgroup compared with others, and only a small number of subgroup analyses were undertaken.

Applying Evidence to Patient Care

The next day on rounds, you again present the results of the CHARISMA trial [5]. You discuss the results, taking into account what you feel are the 2 main limitations in interpreting the results, namely the small beneficial effect of clopidogrel in reducing nonfatal stroke (ARR of 0.5%) and the misleading subgroup analyses that were done. You report that after adjusting for multiple comparisons, there does not seem to be any beneficial effect of clopidogrel in reducing the risk of MI or cardiovascular death in any subgroup of patients.

Dr. Mainieri asks whether or not you will follow the cardiologist's recommendation and prescribe clopidogrel for Mrs. Girouard upon discharge. You feel that, while clopidogrel does reduce the risk of nonfatal stroke, Mrs. Girouard is not at high risk for stroke [21] but is at greatest risk for death or another MI. Unfortunately, clopidogrel did not reduce the risk of these outcomes in patients with multiple cardiovascular risk factors or preexisting cardiovascular disease. Thus, you would not issue a prescription for clopidogrel to Mrs. Girouard upon discharge.

When applying research results to individual patients, especially for important clinical decisions, we must balance the benefits versus harms of the intervention. In doing so, we should be cautious about blindly following the recommendations of study authors and instead should evaluate the evidence ourselves, calculate the more objective NNT and NNH, and not rely solely on comparing relative benefits and harms. In this case, clopidogrel is not likely to help Mrs. Girouard, and the risk of moderate bleeding likely outweighs the small benefit in stroke reduction.

Conclusion

Table 3 summarizes the EBM approach as it applies to the decision whether to add clopidogrel to Mrs. Girouard's cardiovascular treatment regimen. The case illustrates that biased or misleading interpretations of methodologically sound studies can and do occur.

This case also illustrates strategies that can be used to avoid being misled by biased presentation and interpretation of data. Readers of all medical journals should be aware of the use of unfair comparators and composite end points that are not of equal importance, that do not occur with similar frequency, or that do not have similar treatment effects. Furthermore, readers need to recognize that they can be misled when small treatment effects are presented as RRRs. Finally, subgroup analyses can be misleading when they are not planned for appropriately in the design phase of the study and not analyzed appropriately using adjustment for multiple statistical comparisons.

Whenever possible, readers should refer to abstracts published in secondary, preappraised evidence resources (eg, *ACP Journal Club*). The clinicians and methodologic experts who review each article included in these publications do so more objectively than the impassioned authors of the original studies. If no abstract has been published in a preappraised secondary publication, readers would do well to read the methods and results sections of a study and come

Table 3. Summary of the Evidence-Based Medicine Approach to the Decision Whether to Add Clopidogrel to Mrs. Girouard's Cardiovascular Treatment Regimen

Assess	Mrs. Girouard has CAD. From the cardiologist's perspective, the optimal management of the patient's CAD includes the addition of clopidogrel to aspirin. You are concerned that clopidogrel may not be effective for this use.
Ask	In adults with previous MI, does the addition of clopidogrel to aspirin, compared with aspirin alone, reduce the risk of recurrent MI or death?
Acquire	*ACP Journal Club*: 38 hits, with 1 synopsis of an article concluding that clopidogrel plus aspirin did not differ from aspirin alone for reducing MI, stroke, or cardiovascular death in high-risk atherothrombosis.
Appraise	The study meets the validity criteria. However, other secondary guides reveal that the treatment effect is small and the subgroup analyses are misleading. The use of clopidogrel resulted in a small reduction in nonfatal stroke (0.5%; NNT = 200) but had no effect on recurrent MI or cardiovascular death.
Apply	Clopidogrel only reduced the risk of nonfatal stroke. Mrs. Girouard is not at high risk of stroke. Clopidogrel did not reduce the risk of recurrent MI. As a result, you do not feel compelled to prescribe clopidogrel.

CAD = coronary artery disease; MI = myocardial infarction; NNT = number needed to treat.

to their own dispassionate conclusions regarding the study results.

Corresponding Author: Terrence Shaneyfelt, MD, MPH, 700 South 19th Street, Medical Service, Birmingham, AL 35233 (email: terry.shaneyfelt@va.gov).

References

1. Haynes RB. Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based health care decisions [editorial]. ACP J Club 2006;145:A8.

2. Yusuf S, Mehta SR, Zhao F, et al. Early and late effects of clopidogrel in patients with acute coronary syndromes. Clopidogrel in Unstable Angina to Prevent Recurrent Events Trial Investigators. Circulation 2003; 107:966–72.

3. Clopidogrel plus aspirin did not differ from aspirin alone for reducing MI, stroke, and CV death in high-risk atherothrombosis [abstract]. ACP J Club 2006;145: 8. Abstract of Bhatt DL, Fox KA, Hacke W, et al. Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. CHARISMA Investigators. N Engl J Med 2006;354:1706–17.

4. Clopidogrel plus aspirin did not differ from aspirin alone for reducing MI, stroke, and CV death in high-risk atherothrombosis [commentary]. ACP J Club 2006;145:8.

5. Bhatt DL, Fox KA, Hacke W, et al. Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. CHARISMA Investigators. N Engl J Med 2006;354:1706–17.

6. Guyatt G, Rennie D, editors. Users' guides to the medical literature: a manual for evidence-based clinical practice. The Evidence-Based Medicine Working Group. Chicago: AMA Press; 2002.

7. Nicholson LJ. Applying evidence to the care of a patient with bacterial meningitis: how to evaluate a therapeutic option. Semin Med Pract 2004;7:12–23.

8. Devereaux PJ, Anderson DR, Gardner MJ, et al. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. BMJ 2001;323:1218–23.

9. Montori VM, Jaeschke R, Schunemann HJ, et al. Users' guide to detecting misleading claims in clinical research reports. BMJ 2004;329:1093–6.

10. Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. JAMA 2003;289:454–65.

11. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 2003;326:1167–70.

12. Horton R. The rhetoric of research. BMJ 1995;310: 985–7.

13. Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? JAMA 2003;290:921–8.

14. Brenner BM, Cooper ME, de Zeeuw D, et al. Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. N Engl J Med 2001;345:861–9.

15. Olsson AG, Eriksson M, Johnson O, et al. A 52-week, multicenter, randomized, parallel-group, double-blind, double-dummy study to assess the efficacy of atorvastatin

and simvastatin in reaching low-density lipoprotein cholesterol and triglyceride targets: the treat-to-target (3T) study. Clin Ther 2003;25:119–38.

16. Harrington RA, Becker RC, Ezekowitz M, et al. Antithrombotic therapy for coronary artery disease: the seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest 2004;126 (3 Suppl):513S–548S.

17. Freemantle N, Calvert M, Wood J, et al. Composite outcomes in randomized trials: greater precision but with greater uncertainty? JAMA 2003;289:2554–9.

18. Montori VM, Permanyer-Miralda G, Ferreira-Gåonzalez I, et al. Validity of composite end points in clinical trials. BMJ 2005;330:594–6.

19. Parker AB, Naylor CD. Subgroups, treatment effects, and baseline risks: some lessons from major cardiovascular trials. Am Heart J 2000;139:952–61.

20. Lagakos SW. The challenge of subgroup analyses—reporting without distorting [published erratum appears in N Engl J Med 2006;355:533]. N Engl J Med 2006;354:1667–9.

21. Witt BJ, Ballman KV, Brown RD Jr, et al. The incidence of stroke after myocardial infarction: a meta-analysis. Am J Med 2006;119:354.e1–9.

How to cite this article:
Shaneyfelt T, Montori VM. Applying evidence to the care of a patient with cardiovascular disease: how to detect misleading claims in clinical research reports. Semin Med Pract 2007;10:18–29.