Evidence and Education

Elliott M. Antman, MD

Given the illustrious list of prior awardees, I feel deeply honored to have been the James B. Herrick Lecturer for 2010. In planning my Herrick lecture, I thought it would be instructive to use the legacy of Dr Herrick as a lens to assess the evolution of how we acquire evidence and organize our educational mission. This article is based on the 2010 Herrick Lecture but also includes some additional material derived from my own experience in clinical research and writing guidelines documents.

Ninety-eight years ago, Dr Herrick published his classic article on the sudden obstruction of a coronary artery, emphasizing that this event was not uniformly fatal. He hypothesized that when a coronary was occluded, left ventricular function could be preserved by enhancing collateral flow. Herrick was clearly a master clinician, a keen observer of his patients, and, I would argue, a nascent clinical investigator. However, unlike today’s clinical investigators, Herrick was not able to test hypotheses, such as the one about improving collateral flow, because conspicuously absent from his black bag were modern day methods for acquiring the evidence to make an informed therapeutic decision.

Methods for Acquiring Evidence: Clinical Trials

Our evidence toolbox grew with a watershed event in 1948 when Doll, D’Arcy Hart, and Avery Jones working in the United Kingdom introduced the concept of randomization of individual subjects. These early randomized controlled trials (RCTs) were not in cardiology but involved pertussis vaccine and streptomycin for tuberculosis. Major research concepts that influence our contemporary research strategies including informed consent, the ethics of RCTs, and blinding grew out of this early experience.

Another major advance in clinical trial methodology came 10 years later in 1958. A mathematician working at Bell Labs wondered how long vacuum tubes would last in the repeaters of undersea cables before they had to be replaced. At the same time, a biostatistician at Johns Hopkins was working on a way to describe how long a person with cancer would be expected to live with and without treatment. The common problem these 2 investigators faced was individual variation in follow-up, resulting in incomplete observations at any moment in time. Edward Kaplan, the mathematician, teamed up with Paul Meier, the biostatistician, to describe a method for calculating, at any time T, the probability of surviving free of an event, thereby permitting comparison of treatment A versus treatment B by Kaplan–Meier curves.

Now cardiovascular investigators were off and running. A rich array of clinical trials were performed, a robust evidence base was developed, and clinical practice guidelines were generated starting in 1984. Derivative products and performance measures are now produced to facilitate translation of the recommendations from guidelines into routine practice.

The writing committees of the American College of Cardiology/American Heart Association (AHA) guidelines turn not only to RCTs but to registries and in many cases expert consensus. However, despite their high degree of precision, clinical trials are limited in scope, with evidence from RCTs often insufficient to inform general clinical practice. In the future, we will also need to include information gleaned from comparative effectiveness research (CER) and patient preferences. Despite the economic challenges facing contemporary healthcare systems, it remains in dispute as to whether cost should drive those recommendations as well. An important task is to learn how to incorporate data from CER and patient preferences into a larger evidence base when formulating future practice recommendations.

Let us first reflect on the current state of clinical trials in cardiovascular medicine through a prototypical structure of a randomized double-blind controlled trial. Today, the questions we ask are quite complex. Whereas surrogate end points may be helpful for early-phase proof-of-concept research, they can be misleading for designing large-scale RCTs. Valid surrogate end points must be tightly linked to the relevant clinical outcome so that changes in the surrogate are predictive of outcomes. The surrogate end point must also be in the causal pathway of the disease and reliably capture the impact of the treatment being tested on clinical outcomes. Thus, investigators are increasingly focusing on clinical outcomes as the primary end point for an RCT. However, event rates are dropping and we see a greater use of a composite clinical end point. The greater availability of effective treatments makes it harder to show superiority, but important clinically meaningful advantages may still exist with a new treatment. This leads to a noninferiority design.
Because blinding of some of the treatment arms in an RCT can be challenging, investigators may use a prospective, randomized, open-label, blinded end point (or PROBE) design. But a PROBE design may still introduce bias. Therefore, blinding of investigators and patients remains preferable, given the public health impact of application of evidence from RCTs. To increase efficiency in clinical trials, after periodic review of blinded aggregate data investigators may use an adaptive design that could lead to modification of the enrollment criteria, of treatments being tested, or the sample size.

Heightened attention to the safety of marketed drugs has led the US Food and Drug Administration to consider postapproval clinical trials. New drug applications based on RCTs may be required to include a risk evaluation and management strategy (or REMS) if the Agency believes that such a plan is important to ensure that the benefits of the drug outweigh the risks. In an effort to accelerate the development of new medical products, the Food and Drug Administration introduced a critical path initiative that promotes tools for acquiring evidence. These include new assays, standards, computer modeling techniques, biomarkers, and end points of clinical trials. The goal is to reduce the time for delivering promising new treatments to patients by identifying products that do not hold clinical promise earlier in the development pathway and facilitating testing of those products that are more likely to improve the health of patients.

The final point to consider is a direct outgrowth of information delivery in a high-technology age. All too often, a filtered sound bite of the findings of a study in a headline or blog is all that reaches a busy clinician or a patient. Such a simplistic summary fails to capture the rich array of information contained in a clinical trial. We need to think much more comprehensively about a set of issues when interpreting evidence. This is a 6-step process (Figure 1).

1. What was the scientific hypothesis and was it appropriately framed?
2. How effectively did the study plan test the hypothesis?
3. How well was the study implemented?
4. We then examine the actual findings in the study, typically the treatment effect.
5. From the report of the treatment effect, confidence intervals, and measures of statistical significance, we make an inference about “truth” in the study.
6. Importantly for clinical practice, we then draw an inference about the applicability of the findings to the population of patients with the target disease under investigation.

These last 2 inferences allow us to make judgments about the internal and external validity of a clinical trial. Well-designed and implemented clinical trials get high marks for internal validity, but we may still have questions about the generalizability or external validity of the findings.

Methods for Acquiring Evidence: Comparative Effectiveness

Enter the world of CER, a new method of obtaining evidence. We need this type of research because there are gaps and variations in the quality of care delivered and clinical outcomes. We need to know the return on investment of healthcare dollars, and we need to know what works in “the real world.” The toolbox for CER consists of:

1. Clinical trials: These are usually simpler in design and referred to as pragmatic trials (examples include the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial [ALLHAT] trial and the Acute Study of Clinical Effectiveness of Nesiritide in Decompensated Heart Failure Trial [ASCEND-HF]).
2. Systematic overviews (meta-analyses).
3. Decision models.
4. Observational studies (often the most problematic).

Comparative effectiveness studies hold great promise, but we need much more experience with the methodology and statistical tools used in CER in order to apply the evidence from such investigations appropriately in formulating recommendations for clinical care. This point is aptly illustrated by 2 studies investigating personalized medicine using genotype-guided warfarin therapy. Consider the ongoing Clarification of Optimal Anticoagulation Through Genetics (COAG) study sponsored by the National Institutes of Health (http://www.ClinicalTrials.gov, unique identifier 00839657). Patients with an indication for warfarin are randomized and, in a double-blind fashion, have their anticoagulation managed either with a standard clinically guided warfarin regimen or with genotyping added to the standard approach. Experts in this area felt that the benefits of genotyping would be seen early and would not influence late events. The primary end point is, therefore, the time in therapeutic range with warfarin over the first 4 weeks.

Contrast this with the Warfarin Effectiveness Study reported in March 2010. The investigators used a quasi-experimental approach and offered genotyping to participants in 29 prescription drug benefit plans. Eight hundred ninety-six patients accepted that offer, and their genotype results were supplied to their physician a median of 32 days after starting warfarin. The primary end point was the rate of all-cause hospitalization through 6 months. The genotyping group was
comparing with a nonrandomized, matched, historical control group from a year earlier in the same prescription drug plans. Using claims data, the investigators observed a 31%-lower hospitalization rate in the genotyping group. Despite propensity score adjustment, we need to be concerned about selection bias of potentially health-conscious individuals who agreed to be genotyped. Their health behaviors and lifestyles may have had more to do with their rate of hospitalization than knowledge of their genotype for warfarin. Unmeasured confounders might also have impacted hospitalization rates.

The Warfarin Effectiveness Study also raised classic epidemiological concerns. The reduced rate of hospitalization in the genotype group appears to be a statistical association, not a causal relationship, using the 9 criteria introduced in 1965 by Bradford-Hill. For example, the reduction in hospitalizations is inconsistent with other studies; the temporal relationship is unsupported by the evidence to date; and there is no scientific plausibility supporting the observation. Although the evidence from the COAG RCT is likely to be considered by guidelines-writing committee members and perhaps third-party payors, I doubt that the Warfarin Effectiveness Study would be convincing for either group.

In November 2010, a new oral anticoagulant, dabigatran, became available in pharmacies around this country as an alternative to warfarin for preventing stroke and systemic embolic events in patients with atrial fibrillation. Approval of dabigatran was based on the phase-3 trial Randomized Evaluation of Long-Term Anticoagulation Therapy (RE-LY). The results of the RE-LY trial were analyzed by subgroups according to percentage of time the international normalized ratio (INR) was in the therapeutic range at the centers randomizing patients around the world. When the INR was poorly controlled, high-dose dabigatran appeared to show superiority over warfarin; when the INR was well-controlled, both high- and low-dose dabigatran appeared to show noninferiority to warfarin. Imagine if this were a comparative effectiveness study. What degree of INR control would represent the “real world”? The answer is clearly a very complex one.

We must also avoid silo thinking and simplistic blind cost containment coverage decisions to withhold payment for a new anticoagulant such as dabigatran, arguing instead for “using warfarin better.” With >60 years of warfarin experience, we still have problems achieving high rates of INR control uniformly. One needs to factor in the cost of INR monitoring with warfarin and healthcare system savings associated with earlier discharge and prevention of events.

We should also give weight to the voices of patients and clinicians who are seeking a treatment option free of anticoagulation monitoring, and food/drug interactions. Many of these considerations were taken into account by investigators who estimated the incremental cost-effectiveness ratio for quality-adjusted life years saved and found that dabigatran appeared cost effective.

How might one strengthen the CER toolbox? To improve clinical trials in CER investigators should:

1. Embed randomization into the increasingly available electronic medical records. Imagine how differently we might view the Warfarin Effectiveness Study if the use of genotyping was randomized using the electronic medical record, which was also the case report form for the INR results.
2. Use cluster randomization of practices, hospitals, and when they are defined better, accountable care organizations.
3. Study patient preferences, how their health literacy and the social determinants of health impact their understanding of and compliance with a prescribed regimen.

To improve observational studies, investigators should use more propensity scores adjustments, instrumental variable analyses, sensitivity analyses, and multiple control groups and data sources. These efforts will help increase the degree of confidence in the accuracy of an observation. As emphasized in the AHA position paper on CER, we must be alert to issues of association versus causation.

Some important steps that bring structure and organization to the field of CER have occurred over the last 2 years. In 2009, 1.1 billion dollars of American Recovery and Reinvestment Act grant funds were allocated for CER, to be distributed between the US Department of Health and Human Services, the National Institutes of Health, and the Agency for Healthcare Research and Quality. The Federal Coordinating Council for Comparative Effectiveness submitted a report that defined CER as a comparison of benefit and harm from different interventions in the “real world.”

The Affordable Care Act of 2010 established an independent Patient-Centered Outcomes Research Institute (PCORI). On September 23, 2010, the Government Accountability Office announced the membership of PCORI, charging it with refining the methodologies for CER, setting the research agenda, and overseeing the dissemination of research findings. It remains to be seen what impact, if any, the change in the balance of power in Congress after the midterm elections of 2010 will have on the ability to move forward with the proposals for CER.

Another major development bearing on this discussion is the formation of clinical translational science centers. These are designed to accelerate the transformation of basic biomedical discoveries to improved global health by overcoming a series of translational blocks: T1 = translation to humans (first-in-human and proof-of-concept studies); T2 = translation to patients with disease (RCTs to establish efficacy); T3 = translation to practice (CER plays an important role here); and T4 = translation of research findings on a population scale. The skill sets, collaborations, biostatistical expertise, and educational needs vary considerably along the T1–T4 spectrum.

Our Educational Mission

How should we incorporate information from clinical trials, CER, and the explosion of data from clinical translational science into our thinking, our educational efforts, and our annual scientific meetings? As suggested by the educational
psychologist Benjamin Bloom, we must acquire a core set of information that constitutes our knowledge base (Figure 2).32

To date, in many cases, acquisition of this core set of knowledge is the focus of cardiovascular training programs and is designed to pass a board examination in cardiovascular diseases.

However, given the number of facts involved per decision now and the anticipated exponential increase in the future, we need to move beyond simple information acquisition. We must emphasize to our trainees and to our faculty the importance of striving to comprehend, learn how to apply, analyze, synthesize, and evaluate evidence while integrating our iterative clinical experiences (Figure 2).33 We will then arrive at a state of critical thinking where we have moved beyond recitation of facts from the knowledge base. This taxonomy is a useful construct not only for training cardiovascular investigators and future clinicians caring for patients but also for framing how we communicate with and educate our patients about their healthcare choices.

A Contemporary View of Evidence and Education

How do we put this all together? In Figure 3, the y axis represents the strength of evidence and the x axis “proximity” to the patient; roughly, this translates to generalizability to routine practice. RCTs have the highest strength of evidence but because of their enrollment criteria are the farthest from the patient in routine practice. CER, by design, provides more generalizable information but based on evidence that is not as strong. We have mentioned some ways CER can increase the strength of evidence it provides.

The primary driver for clinical trials is the scientific advance being tested. The primary driver for CER is consumer based, focusing on the interface of the patient and the healthcare system. Our patients move from being research subjects to research partners. Going forward, we need to view RCTs and CER as complementary methods for obtaining the data we need to inform practice guidelines and clinical decisions, as well as how we educate investigators, clinicians, and our patients.

Surely this will be difficult work, but if we get it right, we will have a fair, evidence-based, efficient healthcare system. Importantly, we have an obligation to safeguard our patients against a draconian interpretation of the evidence for blind cost containment and coverage decisions: for example, the denial of coverage for a particular statin, angiotensin-converting enzyme inhibitor, or angiotensin II receptor blocker on the basis of the flimsiest evidence and dominated by a business decision at the most critical point on this pathway of actual delivery of care to the patient.

As members of the Council on Clinical Cardiology, the largest of the 16 councils of the AHA, we are in a position to take a leadership role in maintaining the legacy of Dr Herrick. As reviewers of manuscripts or editorialists for 1 of the 11 AHA journals or as a discussant for a late-breaking clinical trial at one of our meetings, we need to exhibit critical thinking and keep the focus on the patient.

By attending to that responsibility as we evaluate and act on evidence, we will help the AHA reach its bold 2020 impact goal, “… to improve the cardiovascular health of all Americans by 20%, while reducing deaths from cardiovascular disease and stroke by 20%.”34

Disclosures

None.

References

1. American Heart Association. The James B. Herrick Award for Outstanding Achievement in Clinical Cardiology. 2010. Available at: http://


