In the past decade, the number of medications to treat osteoporosis has markedly increased (1). These drugs have distinct mechanisms of action. However, information on the efficacy of these compounds relative to one another remains limited, which frustrates physicians who want to practice according to the evidence.

Some may attribute this gap to the U.S. Food and Drug Administration’s regulatory mandate requiring well-controlled studies to assess both efficacy and safety for registration trials (2). Although the use of placebo is not a mandate, studies largely rely on placebo as a comparator. There is no doubt that the application of this regulation since 1962 dramatically and positively changed the way medicines are developed and led to the approval of many new drugs.

We are all aware that a particular study design will be able to answer a set of questions, and that no single study can answer all of the questions. An extensive literature discusses different approaches for designing clinical trials and the use of placebos and active controls (3, 4). I believe that placebo-controlled trials are an invaluable tool for defining a drug’s efficacy during the drug development process. Nonetheless, some question the ethics and utility of using efficacy relative to placebo as the only hurdle to clear before bringing a drug to the market (5, 6). They ask why we deny a patient a known effective treatment—in the form of an active comparator—while missing a golden opportunity to assess relative efficacy.

A pharmaceutical company takes a substantial risk when it compares a new product with an established drug. Failure to demonstrate superiority over a less expensive agent could be financially ruinous. Providers and patients may not embrace a new therapy when it proves to be inferior to an established product in efficacy, safety, or affordability. In contrast, manufacturers run no such risk when seeking approval based on a comparison with placebo and can then promote their product to physicians who apparently do not question the limitations of this form of evidence.

Under the current incentives for drug development, we have more medicines and more choices, but we often lack the scientific evidence to make choices among them. The use of “placebo-only practice” reduces risk for the pharmaceutical industry, but clearly it does not translate into good medicine or good public policy. In allowing market entry based on a comparison only with placebo—rather than requiring a comparison with established agents—we lose an opportunity to serve the best interest of our patients.

In this issue, Cadarette and colleagues (7) provide information from an observational study conducted to address the knowledge gap created by the lack of head-to-head randomized trials of drugs to prevent and treat osteoporosis. They analyzed a cohort of 43 135 Medicare beneficiaries who were enrolled in 1 of 2 statewide pharmaceutical benefits programs and had received a new prescription for bisphosphonates (alendronate or risedronate), calcitonin, or raloxifene. The study outcome was a fracture of the hip, humerus, radius, or ulna within 12 months of the first dose of drug, as detected by a claim for Medicare reimbursement for services rendered for these diagnoses. By using multivariate statistics to adjust for differences in the frequency of several potential confounders in the populations receiving the 4 drugs, the authors explored the relative efficacy of these drugs. They found that individuals receiving alendronate, risedronate, and raloxifene had similar fracture risk, which was considerably lower than with calcitonin.

The authors provide a balanced discussion of the strengths and weaknesses of their methodological approach. Their tables contained more than 3600 discrete data points, which underscore the complexity of the analysis. This complexity and the tools the authors used may be foreign to many readers who lack the sophisticated statistical training to fully understand the impact of these analyses. I personally belong in this group and have taken the authors’ conclusions at face value.

As with many cohort studies, this report has several limitations (8—10). First, patients were not randomly assigned. Physicians, and therefore clinical circumstances, determined who received which drug. Second, the authors had no access to the factors that led physicians to choose one drug and not another. Third, 1 group—those who received calcitonin—were more ill, based on the number of comorbid conditions, medications listed, and recent hospitalizations. Fourth, fewer patients receiving calcitonin or raloxifene had an osteoporosis diagnosis. All these factors could influence fracture rates independent of the effectiveness of the drug these patients received, and even the most sophisticated multivariate analysis cannot overcome these limitations. To the authors’ credit, they acknowledge several situations in which they cannot exclude confounding by indication—whereby a clinical finding that determines whether a person receives a drug is causally related to the study outcome—as a possible explanation for some results. Randomly assigned treatment would negate most of these concerns, which is why we have come to rely on randomized trials for evidence about drug treatments.

Three points illustrate some of the issues with observational studies based on large data sets. The first is about the clinical importance of a small but statistically significant difference. What is the clinical meaning of a statistically significant difference in mean patient age in the ralox...
ifene group (76.9 years) and the bisphosphonate group (78.7 years)? A 2-year difference could be critical between neonates and 2-year-old children and between 13-year-old prepubertal children and children who are sexually mature at age 15. However, in the older cohort studied by Cadarette and colleagues (7), this “difference” is probably meaningless physiologically. A second issue is multiple comparisons, which raise concern that some of the observed differences are due to chance, even though the authors raised the bar for interpreting a result as statistically significant. Finally, data obtained for administrative purposes, which are the core source for this study, are often suspect because it is difficult to assess the accuracy of the information and what information may be missing. The authors allude to some of these issues without fully explaining how they could affect their conclusions.

In the end, the fracture rates were similar for the 2 bisphosphonates and raloxifene. Will this information be useful to select one drug over another? Will it help physicians decide which of these drugs to use in specific patients? I do not think so. Even if the authors had found a clear, undisputed gradient in efficacy among these drugs, I believe that we would need a prospective randomized study before we could use the findings as the basis for practice guidelines.

These results underscore, once more, the lack of good comparative data for choosing drug therapy. Cadarette and colleagues (7) conclude, in accord with many previous articles and reviews, that more randomized, controlled studies are needed, implicitly acknowledging that their approach has important shortcomings. However, the dearth of head-to-head trials suggests that our system for evaluating health technology has not created the incentives to do these trials. Without incentives, the needed studies may never materialize, and we will continue to cope with uncertainty when choosing a drug. A good example of a creative incentive is the pediatric initiative that allowed manufacturers 6 additional months of exclusivity under patent if they made a commitment to study drug effects in pediatric patients (11). Many manufacturers took advantage of this proposal, and I believe that our knowledge about drug effects in children improved.

I submit that we have potential opportunities to explore the relative efficacy of drugs for the treatment of most conditions. As Cadarette and colleagues (7) show, government agencies now pay for drugs for millions of Americans. Under these programs, the U.S. government has already paid for antosteoporosis drugs for thousands of patients, despite weak evidence about which are the most effective. I propose that we devise an ethical way to prospectively randomly assign patients to different (and apparently equivalent) drug regimens and measure the outcomes of treatment, potential adverse events, drug interactions, and costs. I assert that if the government pays for tests or treatments, it has an obligation to evaluate them relative to one another.

Although this approach alone will not answer every question, it should narrow the gap between what we know and what we need to know, allow us to make better treatment decisions, and help us to better allocate scarce resources.

Saul Malozowski, MD, PhD, MBA
National Institute of Diabetes and Digestive and Kidney Diseases
Bethesda, MD 20892-5460

Disclaimer: This editorial contains the author’s views and does not constitute an official position of the U.S. Department of Health and Human Services.

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Saul Malozowski, MD, PhD, MBA, Division of Diabetes, Endocrinology, and Metabolic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, Democracy Two, 6707 Democracy Boulevard, Room 607, Bethesda, MD 20892-5460; e-mail, smalozow@nih.gov.


References