Right Answers, Wrong Questions in Clinical Research

Ralph I. Horwitz,1,2* Jill E. Abell,1,3 Jennifer B. Christian,1,3 Ashley E. Wivel1

To ensure that clinical research arrives at the “right” answers to the right questions for patients, studies should be designed to more closely approximate real-world use of therapeutics and devices.

The emphasis on the internal validity of clinical research in evidence-based medicine has had the effect of making the quality of the “answer” more important than the appropriateness of the “question.” In the pursuit of a valid answer, randomized controlled trials (RCTs) that emphasize efficacy under near-ideal conditions have become a preferred strategy for both regulators (who need to approve medicines and devices for clinical use) and investigators (who design trials). When “efficacy trials” dominate, and studies that reflect real-world use of the treatment are reduced in importance, a surprising collateral effect is that the value attributed to the patient’s experience with their disease and its treatment is diminished.

Fundamentally, there is no bright line that separates efficacy studies from those that assess effectiveness. With the right methods, observational studies can be used to assess treatment benefits. And RCTs exist on a continuum from those that emphasize efficacy (explanatory trials) to those that emphasize effectiveness (practical or pragmatic trials). In this Focus, we illustrate the central importance of the research question in the design of clinical studies to get the right answer for patients and physicians.

RIGHT QUESTIONS FOR RCTS

Numerous RCTs of the asthma drug montelukast had shown that it was inferior to inhaled corticosteroids as a first-line treatment for asthma control and also inferior to long-acting beta agonists (LABAs) as a second-line “add-on” therapy. Montelukast is an oral medicine with higher adherence than that of inhaled agents when used in real-world settings. However, as often occurs in short-term (~6 months), placebo-controlled trials, patients were able to maintain high levels of adherence to all treatments, therefore obscuring the influence of adherence levels on the effectiveness of treatment. Furthermore, the RCT investigators chose to define the primary study outcomes as changes in lung function, a hard physiological measure.

A study that illustrates the apparent advantage of inhaled steroids and LABAs comes from the results of an efficacy RCT in which montelukast was compared with beclamethasone and placebo (1). In this study, both montelukast and beclamethasone were superior to placebo [in improvements in forced expiratory volume in 1 second (FEV1)], and beclamethasone was superior to montelukast. Despite these findings, montelukast remains popular with patients and physicians. Evidence to support this popularity was generated when montelukast was tested in real-world effectiveness studies by using a pragmatic randomized trial design (2). In these real-world studies, montelukast was just as effective as both inhaled corticosteroids and LABAs. What was different?

The real-world studies enrolled patients with asthma who were more like those seen in actual clinical practice; had lower adherence rates that reflect levels seen commonly in customary care, not those artificially achieved in placebo controlled trials; and measured clinical end points such as symptoms, function, and well-being. When tested rigorously in this setting, the adherence advantage of montelukast as a once-daily oral medication was evident. Thus, what was different between the traditional efficacy RCTs and the pragmatic studies of montelukast is that they asked different questions about the performance of the drug in different settings.

The choice of the question is closely linked to the features that distinguish efficacy from effectiveness trials. Four features distinguish the traditional efficacy RCT: (i) populations are homogeneous, (ii) the spectrum of the medical illness is often narrowly defined, (iii) the new treatment requires strict guidelines for use of the evaluated medicine and concomitant therapies, and (iv) “hard” end points (death and major

Open-ended questions. Asking the wrong questions in clinical research can generate seemingly right answers, but these answers may not be enough to reflect or predict real-life scenarios.
morbidity) are preferred over “softer” clinical outcomes. When studies are designed to test the effectiveness of medicines or devices as they are used in clinical practice, they emphasize the four areas differently: (i) populations are heterogeneous, (ii) medical conditions are broadly defined, (iii) conditions of use reflect real-world settings, and (iv) softer end points (symptoms, function, and patient preference) complement outcomes such as death or major morbidity.

Right answers, although useful for regulatory approval, are not always what is right for patients who live in the real world. Indeed, typical patients are often not eligible for efficacy trials owing to narrow inclusion and exclusion criteria or the use of cotherapies not permitted in the trial.

DESIGN FEATURES

The only feature of the RCT that cannot be included in observational studies is the use of randomization to assign treatment. Investigators interested in RCTs evaluating clinical effectiveness can modify other design features so as to mimic clinical practice more closely. Increasingly, we are seeing modified RCT designs, such as cluster-based trials and point-of-care randomization. Other modifications to RCTs include distributing medicines to community pharmacies as well as obtaining data from routine clinical care documented in electronic health records. All of these innovations are helping to embed trials more directly into real-world settings.

Not surprisingly, RCTs frequently deviate from this real-world design, so that the intended evaluation of efficacy is often not achieved. When RCTs follow patients for prolonged durations, for example, the investigators often are left to struggle with sources of bias that are commonplace in observational studies. Consider the Women’s Health Trial that randomized 16,000 postmenopausal women to either hormone replacement therapy (HRT) or placebo and planned an 8-year follow-up to test whether HRT would improve certain clinical outcomes. The study was stopped early when an increased risk was observed for coronary heart disease among women assigned to receive HRT (3).

The interpretation of these results was complicated when it was reported that treatment was discontinued, and blinding was broken for nearly half of the HRT users but only a small percentage of the placebo users. The changes in treatment occurred mostly to manage vaginal bleeding that developed as the trial progressed among women assigned to estrogen therapy. But the loss of blinding created unanticipated problems in detection bias and in the adjudication of end points (because patients and physicians were unblinded by a treatment side effect), and the changes in treatment assignment meant that the planned intent-to-treat analysis no longer could answer the question that had been originally proposed (4).

Instead of answering a question of the efficacy of estrogens on cardiovascular disease (CVD), the study was now a test of initiating a treatment that many patients had changed during follow-up.

Although it is possible to debate the interpretation of the HRT trial results, it is clear that RCTs that follow patients for prolonged time periods may lose the benefits that ordinarily accrue to the design of efficacy RCTs and may find that they are answering a question different from the one that motivated the trial. Newer approaches use analytical methods commonly applied to observational studies that account for changes in medical conditions and treatments over time and strengthen the analysis of such long-term RCTs.

RIGHT QUESTIONS FOR OBSERVATIONAL STUDIES

For many years, observational studies had suggested that estrogen users had a decreased risk of cardiovascular mortality and ischemic heart disease. The studies that reported this apparent decreased risk compared women who were currently using estrogens with nonusers and estimated that estrogen users had a 25 to 35% reduction in risk of CVD (5). However, when the effect of estrogen on CVD was evaluated in an RCT, as noted above, estrogen users had a 20% increased risk of CVD (6). What could explain this discrepancy?

The customary explanation is one familiar to all clinical investigators: confounding. In this instance, the bias was postulated because women who chose to use estrogens had a lower risk of CVD at baseline than that of women who were nonusers. Investigators were aware of this potential bias and took steps in design (matching and similar tactics) and analysis (multivariate adjustment) to mitigate these differences. But as often happens, such efforts were considered insufficient, and “residual confounding” was believed to have led to erroneous conclusions. The collateral damage from this study was the conclusion that observational studies are unable to provide reliable and accurate estimates of the effects of treatment on disease outcomes.

Hernan and colleagues challenged the suggestion that discrepancies in results between the RCTs and the observational studies were explained with residual confounding. In their analysis, the authors pointed out that the RCT and the observational study asked different questions. Whereas the observational study had compared current users of estrogens with nonusers, the RCT compared new users of estrogens with nonusers. To prove the point, Hernan and colleagues reanalyzed a large observational cohort (Nurses Health Study) and reported similar results to the RCT when both types of studies defined treatment based on a new user definition (7). The choice of current users in observational studies is not helpful because it does not indicate what would happen if patients started or stopped a treatment. For this reason, focusing on new users, as is done in RCTs, is preferred in the design of observational studies.

The results of observational studies can be distorted if the indication for treatment creates baseline inequalities in the compared groups. This potential bias is avoided when randomization is used to assign treatment. Fortunately, newer approaches to the design and analysis of observational research can mitigate this bias. Other potential sources of bias in observational research, such as losses to follow-up, variations in adherence to treatment, or problems in adjudication of outcomes, are shared in common with RCTs and require careful consideration in design to avoid misleading results.

RIGHT QUESTIONS FOR PATIENTS

When investigators design studies, they often focus on whether the medicine improves survival or reduces some “hard” physiological measure, such as lung spirometry. Although survival and other hard clinical outcomes are important to patients, they also want to know whether they will feel better and can do more—an emphasis on patient experience that is central to the mission of Patient-Centered Outcomes Research Institute (PCORI) (www.pcori.org).

Companies responsible for developing new drugs are getting the patient-focused message, as is the U.S. Food and Drug Administration (FDA), which is responsible for approving new drugs. In November 2011, the FDA approved ruxolitinib to treat...
myelofibrosis based on two phase III trials. Reduction in spleen volume was the primary end point in both trials, but demonstrating that the medicines both reduced spleen volume and improved patients’ symptoms was decisive in the FDA’s decision to grant approval to ruxolitinib. An FDA official involved in the regulatory decision stated, “[Patient symptoms] was a secondary end point, but in our mind this is why we gave the application full approval” (8).

The increasing focus on the patient experience will need to be accompanied by an increasing emphasis on the validity and reliability of patient-based end points. The American Heart Association recently issued a Scientific Statement on the Importance of Measuring Patient Reported Health Status (symptom burden, functional status, and health-related quality of life) (9). The American Society of Clinical Oncology recently drafted recommendations to “raise the bar” for clinical trials in an effort to inspire patients and investigators to demand more from clinical trials, to focus on what is clinically meaningful to the patient, and to “vote with your feet” by only participating in trials that evaluate the impact of treatment on the experience of the patient with their illness and its therapy (10). Hardening softer data captured from the patient experience is a fundamental requirement to ensure that patients and physicians are guided by evidence that is both true and apposite for the clinical circumstances.

QUALITY OF THE QUESTION

For too long, the design of clinical research has been driven by the goals of the investigator to get the unbiased right answer, even at the expense of the importance or applicability of the study results. Thanks to advances both in research methods and the increasingly prominent voice of the patient, there is a long-overdue new emphasis on the quality of the question. But much more needs to be done. Randomized trials to assess the efficacy of treatment will continue to be needed to demonstrate whether a medicine or device can work under ideal circumstances. These trials can be strengthened further by more substantive attention to methodological considerations such as more patient-relevant outcomes, more inclusive study populations, and the use of concomitant therapies. But these modifications, as valuable as they would be, will not substitute for the design and conduct of studies that measure the effectiveness of medicines as they are used in real-world, complex settings. Randomized trials have an important role to play in the evaluation of the effectiveness of treatment as used in clinical practice. Innovative approaches, including adaptive trial design and Web-based RCTs, promise to advance the scientific basis and utility of clinical research.

As new medicines are developed with dosing profiles that enable higher levels of adherence than have been possible to date, the right question for patients will be how treatment outcomes vary according to levels of adherence. In the example of montelukast in real-world studies, adherence was less (65%) than we would like to achieve, although better than adherence levels observed for inhaler-administered medicines, such as corticosteroids and LABA (45%) (2). A new focus on adherence will encourage the development of programs that improve both adherence and the effectiveness of new medicines. It will not be sufficient to implement such programs without also conducting studies that assess their impact on patient outcomes. Investigators and patients will need to work together to measure relevant patient outcomes in real-world settings.

REFERENCES AND NOTES


Competing interests: GSK markets versions of montelukast and beclomethasone in some regions (although not in the United States) where local studies with these medicines may be performed, although no extensive clinical development program is ongoing for either medicine at this time. Additionally, GSK has a respiratory portfolio that includes medications in the ICS and LABA classes
