A new federal initiative has allocated $1.1 billion to comparative effectiveness research, and many have emphasized the importance of including observational studies in this effort. The rationale for using observational studies to assess comparative effectiveness is based on concerns that randomized controlled trials (RCTs) are not “real world” because they enroll homogeneous patient populations, measure study outcomes that are not important to patients, use protocols that are overly complex, are conducted in specialized centers, and use study treatments that are not consistent with usual care, and that RCTs are not always feasible because of a lack of equipoise, the need to assess delayed endpoints, and concerns that they take years to complete and are expensive. This essay questions the validity of each of these proposed limitations, summarizes concerns raised about the accuracy of results generated by observational studies, provides some examples of discrepancies between results of observational studies and RCTs that pertain to pulmonary and critical care, and suggests that using observational studies for comparative effectiveness research may increase rather than decrease the cost of health care and may harm patients.

Keywords: observational studies; randomized controlled trials; comparative effectiveness research

In an attempt to improve patient care and reduce health care expenditures, a new federal initiative has allocated $1.1 billion to comparative effectiveness research (CER). The Institute of Medicine has emphasized the need to include observational studies in this research effort (1), and the Federal Coordinating Council for CER recommended that large electronic data networks and patient registries be developed to facilitate conducting observational studies (2). Although observational studies contribute important knowledge about the causes and pathogenesis of disease and are useful for generating hypotheses (3), with respect to CER, the GRADE Working Group concluded that they represent “low grade evidence” because “further research is very likely to have an important impact on our confidence in the estimate of effect, and is likely to change the estimate” (author’s emphasis) (4).

Numerous publications, including one from the CER in Lung Diseases Workshop Panel (5), base the rationale for including observational studies in CER on concerns that randomized controlled trials (RCTs) are not “real world” because they enroll homogeneous patient populations, include study outcomes that are not important to patients, use protocols that are overly complex, are conducted in highly specialized centers, and use study treatments that may not be consistent with usual care and that they are not always feasible because of a lack of equipoise on the part of physicians or patients, the need to assess delayed endpoints, and because they frequently require years to complete and are expensive (5, 6).

The purpose of this essay is to question the validity of each of these proposed limitations of RCTs, to summarize concerns raised about the accuracy of results generated by observational studies, to provide examples of discrepancies between results of observational studies and RCTs that pertain to pulmonary and critical care, and to suggest that using observational studies to compare the effectiveness of various interventions may increase rather than decrease the cost of health care and, more importantly, may harm patients.

RCTs Are NOT “REAL WORLD”

Homogeneous Patient Populations

Because most RCTs enroll subjects meeting numerous inclusion and exclusion criteria, the ability to extrapolate the results to less rigorously defined patients is limited. Consider a RCT designed to assess the effects of two different treatments of pneumonia in outpatients. The inclusion criteria could require that patients have one or more clinical symptoms and signs consistent with pneumonia and a chest X-ray confirming the presence of a parenchymal infiltrate. All patients with diabetes, chronic obstructive pulmonary disease (COPD), or lung cancer could be excluded because these comorbidities could alter the response treatment. Such a trial could be criticized as not being “real world” because of the excluded patients and because taking chest X-rays might not represent standard practice in the outpatient setting.

Alternatively, the inclusion criteria for this trial could define patients as having pneumonia on the basis of a physician’s subjective assessment, eliminating the need for specific symptoms or signs or for a confirming chest X-ray. The trial could also include all patients regardless of comorbidities. RCTs do not require that patient populations be homogeneous. Rather, investigators can define enrollment criteria as narrowly or broadly as they wish while preserving the ability to randomly assign the various treatment options.

Outcomes That Are Not Important to Patients

The same response pertains to this criticism of RCTs. In the above example, an appropriate primary endpoint might be the extent to which the infiltrate on the chest X-ray resolved after treatment, but it could just as easily be the patient’s degree of subjective improvement. As with inclusion and exclusion criteria, outcome variables in RCTs can be anything one wishes. Using outcomes that are important to patients does not mandate the need for observational study designs.

RCTs Are Too Complex

RCTs are frequently very complex but not because they are RCTs. One recently published RCT designed to determine if azithromycin decreased acute exacerbations of COPD used time-to-first exacerbation as the primary endpoint but also assessed 10
secondary endpoints (one of which was the St. George Respiratory Questionnaire, an outcome that would be considered important to patients) (7). Had the sole endpoint in this study been the St. George Respiratory Questionnaire, it would have been much less complex and much less costly (see below), but it also would have yielded much less information.

Treatment compliance is almost always assessed in RCTs despite the fact that doing so may increase the complexity and cost of the study. However, there is nothing about RCTs that requires assessment of compliance. Although investigators involved with designing and conducting RCTs generally want to get as much out of the study as possible, protocols could be as simple as identifying a patient (“Doc, I think I have pneumonia”), randomly assigning treatment, and assessing a single subjective endpoint (i.e., “calling to ask if the patient is better”). Reducing the complexity of RCTs does not lead to a need for observational study designs.

**RCTs Are Conducted in Specialized Centers**

Many RCTs are conducted by investigators working in academic health centers, so it is not surprising that many are performed in these specialized institutions. Although the implication is that patients seeking care at these specialized centers are different from those treated in the “real world,” study populations are defined by inclusion and exclusion criteria, not by where the study is conducted. RCTs could be designed to require that all patients be recruited and followed exclusively by physicians in private practice. Assuring “real world” patients treated in “real world settings” does not require observational study designs because inclusion and exclusion criteria can be applied anywhere, by anyone, and qualifying patients can still be randomized.

**Management Does Not Represent Usual Care**

Many RCTs compare the effects of a new intervention with the effects of a placebo. Although placebo-controlled trials may satisfy regulatory requirements, they can rightly be criticized for not providing data that enable practitioners to compare the effects of the new treatment with usual care. RCTs comparing one treatment with another can similarly be criticized when neither intervention represents usual care, as was thought to be the case with the study of low tidal volumes in patients with acute respiratory distress syndrome (8). This problem does not require an observational study design to solve, however, because RCTs can test new treatments against usual care, even when “usual care” varies from provider to provider.

RCTs that enroll heterogeneous patient populations, study outcomes that are important to patients, and compare clinically important interventions against usual care have been termed “pragmatic” or “practical” clinical trials (9). The designs of such studies address the “real world” concerns noted above while preserving the critically important feature of being able to randomize patients.

**RCTs ARE NOT FEASIBLE**

**Lack of Equipoise**

Physicians may believe it is unethical to randomize patients in a RCT if they are convinced that the treatment being studied is effective or ineffective or if they believe that the side effects of treatment are too frequent or too severe. If large numbers of patients are excluded from RCTs, generalizing the results becomes more difficult, and it is harder to achieve the targeted enrollment. This latter problem has occurred in numerous RCTs in critical care, resulting in studies that are underpowered for their primary endpoint. This same lack of equipoise would be an even greater limitation to any observational study of the same question, however, because it would result in substantial compromise of the results by sampling bias. Although lack of equipoise limits both RCTs and observational studies, RCTs can quantify the extent of this limitation because the patient flow diagrams in these studies should specifically identify the number of patients excluded by physician preference.

**Delayed Endpoints**

If the proposed benefit of an intervention is not realized until well into the future, studying that intervention with a RCT would not seem feasible. Whether this concern represents a true limitation depends on the importance of the question being asked. Observational data collected over decades indicated that hormonal replacement in postmenopausal women resulted in a 40 to 50% decrease in coronary artery disease. The question was finally addressed by a RCT that enrolled 161,809 subjects over 5 years and followed them for a mean of 5 years before the trial was stopped early on learning that hormone replacement increased the incidence of breast cancer and increased the incidence of coronary disease (10).

**Prolonged RCTs**

RCTs take longer to complete than observational studies, and the National Institutes of Health, the National Cancer Institute, the Food and Drug Administration, and the The Institute of Medicine are actively pursuing ways to shorten the time it takes from conception to approval of an RCT. Even if this part of the process were streamlined, enrollment rates in RCTs are almost always slower than anticipated. Despite the fact that enrollment is a function of the number of participating sites and can be increased by appropriate administrative planning and careful attention to inclusion and exclusion criteria, observational studies will always produce data more rapidly than RCTs. As noted by the GRADE Working Group, however, further research is “very likely to alter the estimate of effect(s)” observed in observational studies and is “likely to change the result” (author’s emphasis) (4). This leads to the question of whether it is better to rapidly acquire data of questionable reliability or to accrue higher quality information more slowly. Given the dictum of *primum non nocere*, the answer should be evident.

**RCTs Are Too Expensive**

Recently published RCTs by the NHLBI-sponsored clinical research networks studying asthma, COPD, or acute respiratory distress syndrome have ranged in cost from $1.4 M to $13 M, with per-patient costs of $8,000 to $30,000. Industry-sponsored trials generally cost far more. Although this clearly represents an enormous expense, there are at least two reasons why this criticism of RCTs is not valid. First, as noted above regarding trial complexity, the cost of RCTs is high in part because investigators are trying to maximize the information collected. If the primary focus were to limit trial expense, RCTs could be designed to address a single endpoint, minimize testing, ignore compliance, reduce patient contact during the trial, and eliminate collecting specimens for future analysis. Regarding the design of RCTs, with some limitations, you get what you pay for. Second, the expense of RCTs should only be one part of the equation. The cost of the lung volume reduction surgery trial was $60 M (i.e., $48,000/patient), but the savings resulting from applying the results of the study to future care was estimated at to be as much as...
$660 M (11). The cost of the Women’s Health Initiative was approximately $800 M ($48,200/patient), but sales of hormone replacement medications decreased by over $2 billion in the first 2 years after the results became available (12). If the intent of CER is to improve the quality of health care and reduce the cost, far more RCTs should be funded because the resulting savings could dwarf the expense. Reducing the cost of research by encouraging observational study designs is penny wise and pound foolish.

ACCURACY OF OBSERVATIONAL STUDIES

Observational studies are uniformly recognized as being limited by indication bias and subject to the potential effects of unmeasured confounders. Although some case-mix adjustment strategies designed to circumvent the effects of unmeasured confounders may be better than others (13, 14) and although some suggest that newer statistical approaches can adjust for these confounders (15, 16), others have found that no strategy adequately adjusts for confounding by indication (13, 17, 18). It is certainly quicker, easier, and less expensive to perform CER by examining preexisting administrative databases, but these have their own intrinsic problems (19).

Several studies have found good correlations between the results generated by observational studies and RCTs asking the same question (20–22), but a panel of experts in CER in lung diseases noted that there are “striking examples wherein multiple observational studies suggested one therapy and subsequent RCTs showed opposite results” (6), and others concur (22, 23) (e.g., Table 1).

Ioannidis (24) identified 45 articles with more than 1,000 citations that were published between 1990 and 2003 in the New England Journal of Medicine, JAMA, Lancet, and specialty journals with an impact factor greater than 7. Six of these studies were observational, and 39 were RCTs. Subsequently published work contradicted five of the six observational studies (83%) but only 9 of the 39 RCTs (23%; \(P < 0.008\)).

TABLE 1. COMPARISON OF RESULTS FROM OBSERVATIONAL STUDIES AND RANDOMIZED CONTROLLED TRIALS PERTINENT TO PULMONARY AND CRITICAL CARE

<table>
<thead>
<tr>
<th>Question</th>
<th>Observational Studies</th>
<th>Randomized Controlled Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is bariatric surgery more effective than conventional therapy for managing obstructive sleep apnea?</td>
<td>Yes, 60–80% remission of OSA symptoms after bariatric surgery; apnea–hypopnea index decreased 33.85/h (multiple observational studies that are summarized in Ref. 30 as part of a systematic review and metaanalysis)</td>
<td>No, no difference in reduction in apnea-hypopnea index despite more weight loss with surgery (31)</td>
</tr>
<tr>
<td>Does an intra-aortic balloon pump decrease the mortality of cardiogenic shock following acute myocardial infarctions?</td>
<td>Yes, 11% decrease in mortality (eight cohort studies that are summarized in Ref. 32 as part of a metaanalysis)</td>
<td>No, no decrease in mortality observed (33)</td>
</tr>
<tr>
<td>Does prolonged storage of red blood cells prior to transfusion have adverse effects?</td>
<td>Yes, increased mortality, prolonged intubation, and increased rates of infection and sepsis (Ref. 34 and others cited in Ref. 35)</td>
<td>No, no adverse effects observed (35)</td>
</tr>
<tr>
<td>Does low caloric intake adversely affect clinical outcomes in medical intensive care unit patients?</td>
<td>Yes, lower likelihood of achieving spontaneous ventilation, bacteremia, and increased hospital mortality in patients with low caloric intake (Ref. 36 and others cited in Ref. 36)</td>
<td>No, no adverse effects of hypocaloric feeding observed (37)</td>
</tr>
<tr>
<td>Does reducing pulmonary arterial wedge pressure improve survival in patients with ARDS?</td>
<td>Yes, survival increased from 29 to 75% in patients whose wedge pressure was reduced (38)</td>
<td>No, no difference in survival in group whose wedge pressure was reduced (39)</td>
</tr>
<tr>
<td>Is infection increased by prolonged use of central venous catheters?</td>
<td>Yes, greater infection rate with prolonged catheter use (Ref. 40 and others cited in Ref. 41)</td>
<td>No, no effect of routine catheter replacement after 3 d (41)</td>
</tr>
<tr>
<td>Do corticosteroids decrease mortality in patients with permissive acute respiratory distress syndrome?</td>
<td>Yes, decreased mortality (Ref. 42 and others cited in Ref. 42)</td>
<td>No, increased mortality (43)</td>
</tr>
<tr>
<td>Do corticosteroids decrease mortality in septic shock?</td>
<td>Yes, 14% in patients receiving steroids and 43% in control subjects (retrospective portion of Ref. 44 and others cited in Ref. 45)</td>
<td>No, 76–77% in patients receiving steroids, 69% in control subjects (45) and 34% in patients receiving steroids, and 25% in control subjects (46)</td>
</tr>
</tbody>
</table>

Definition of abbreviations: ARDS = acute respiratory distress syndrome; OSA = obstructive sleep apnea; Ref. = reference.
Author disclosures are available with the text of this article at www.atjournals.org.

References


