

## Methods in Comparative Effectiveness Research

Katrina Armstrong

From the Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Submitted February 7, 2012; accepted June 22, 2012; published online ahead of print at www.jco.org on October 15, 2012.

Supported by Award No. UC2CA148310 from the National Cancer Institute.

The content is solely the responsibility of the author and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

Author's disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Katrina Armstrong, MD, MSCE, 1204 Blockley Hall, University of Pennsylvania, Philadelphia, PA 19104; e-mail: karmstro@mail.med.upenn.edu.

© 2012 by American Society of Clinical Oncology

0732-183X/12/3034-4208/\$20.00

DOI: 10.1200/JCO.2012.42.2659

### A B S T R A C T

Comparative effectiveness research (CER) seeks to assist consumers, clinicians, purchasers, and policy makers to make informed decisions to improve health care at both the individual and population levels. CER includes evidence generation and evidence synthesis. Randomized controlled trials are central to CER because of the lack of selection bias, with the recent development of adaptive and pragmatic trials increasing their relevance to real-world decision making. Observational studies comprise a growing proportion of CER because of their efficiency, generalizability to clinical practice, and ability to examine differences in effectiveness across patient subgroups. Concerns about selection bias in observational studies can be mitigated by measuring potential confounders and analytic approaches, including multivariable regression, propensity score analysis, and instrumental variable analysis. Evidence synthesis methods include systematic reviews and decision models. Systematic reviews are a major component of evidence-based medicine and can be adapted to CER by broadening the types of studies included and examining the full range of benefits and harms of alternative interventions. Decision models are particularly suited to CER, because they make quantitative estimates of expected outcomes based on data from a range of sources. These estimates can be tailored to patient characteristics and can include economic outcomes to assess cost effectiveness. The choice of method for CER is driven by the relative weight placed on concerns about selection bias and generalizability, as well as pragmatic concerns related to data availability and timing. Value of information methods can identify priority areas for investigation and inform research methods.

*J Clin Oncol* 30:4208-4214. © 2012 by American Society of Clinical Oncology

### INTRODUCTION

The desire to determine the best treatment for a patient is as old as the medical field itself. However, the methods used to make this determination have changed substantially over time, progressing from the humoral model of disease through the Oslerian application of clinical observation to the paradigm of experimental, evidence-based medicine of the last 40 years. Most recently, the field of comparative effectiveness research (CER) has taken center stage<sup>1</sup> in this arena, driven, at least in part, by the belief that better information about which treatment a patient should receive is part of the answer to addressing the unsustainable growth in health care costs in the United States.<sup>2,3</sup>

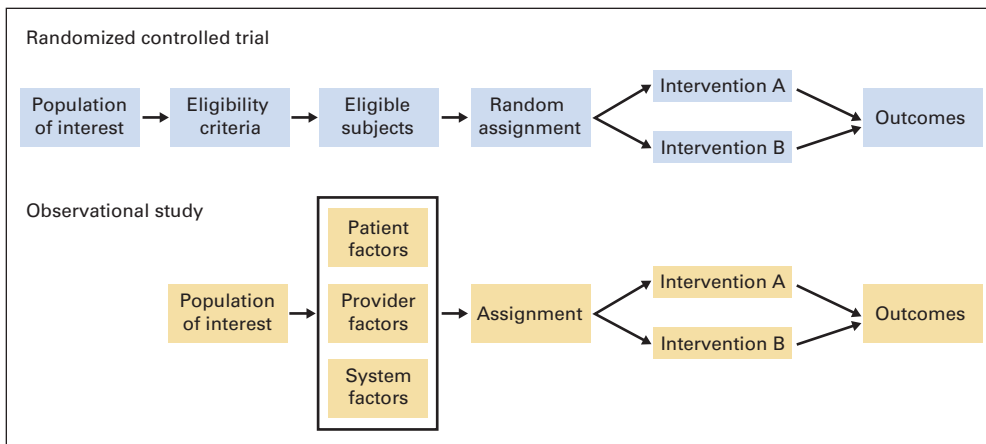
The emergence of CER has galvanized a re-examination of clinical effectiveness research methods, both among researchers and policy organizations. New definitions have been created that emphasize the necessity of answering real-world questions, where patients and their clinicians have to pick from a range of possible options, recognizing that the best choice may vary across patients, settings, and even time periods.<sup>4</sup> The long-standing

emphasis on double-blinded, randomized controlled trials (RCTs) is increasingly seen as impractical and irrelevant to many of the questions facing clinicians and policy makers today. The importance of generating information that will “assist consumers, clinicians, purchasers, and policy makers to make informed decisions”<sup>1(p29)</sup> is certainly not a new tenet of clinical effectiveness research, but its primacy in CER definitions has important implications for research methods in this area.

CER encompasses both evidence generation and evidence synthesis.<sup>5</sup> Generation of comparative effectiveness evidence uses experimental and observational methods. Synthesis of evidence uses systematic reviews and decision and cost-effectiveness modeling. Across these methods, CER examines a broad range of interventions to “prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care.”<sup>1(p29)</sup>

### EXPERIMENTAL METHODS

RCTs became the gold standard for clinical effectiveness research soon after publication of the first RCT in 1948.<sup>6</sup> An RCT compares outcomes across groups



**Fig 1.** Experimental and observational study designs. In a randomized controlled trial, a population of interest is screened for eligibility, randomly assigned to alternative interventions, and observed for outcomes of interest. In an observational study, the population of interest is assigned to alternative interventions based on patient, provider, and system factors and observed for outcomes of interest.

of participants who are randomly assigned to different interventions, often including a placebo or control arm (Fig 1). RCTs are widely revered for their ability to address selection bias, the correlation between the type of intervention received and other factors associated with the outcome of interest. RCTs are fundamental to the evaluation of new therapeutic agents that are not available outside of a trial setting, and phase III RCT evidence is required for US Food and Drug Administration approval. RCTs are also important for evaluating new technology, including imaging and devices. Increasingly, RCTs are also used to shed light on biology through correlative mechanistic studies, particularly in oncology.

However, traditional approaches to RCTs are increasingly seen as impractical and irrelevant to many of the questions facing clinicians and policy makers today. RCTs have long been recognized as having important limitations in real-world decision making,<sup>7</sup> including: one, RCTs often have restrictive enrollment criteria so that the participants do not resemble patients in practice, particularly in clinical characteristics such as comorbidity, age, and medications or in sociodemographic characteristics such as race, ethnicity, and socioeconomic status; two, RCTs are often not feasible, either because of expense, ethical concerns, or patient acceptance; and three, given their expense and enrollment restrictions, RCTs are rarely able to answer questions about how the effect of the intervention may vary across patients or settings.

Despite these limitations, there is little doubt that RCTs will be a major component of CER.<sup>8</sup> Furthermore, their role is likely to grow with new approaches that increase their relevance in clinical practice.<sup>9</sup> Adaptive trials use accumulating evidence from the trials to modify trial design of the trial to increase efficiency and the probability that trial participants benefit from participation.<sup>10</sup> These adaptations can include changing the end of the trial, changing the interventions or intervention doses, changing the accrual rate, or changing the probability of being randomly assigned to the different arms. One example of an adaptive clinical trial in oncology is the multiarm I-Spy2 trial, which is evaluating multiple agents for neoadjuvant breast cancer treatment.<sup>11</sup> The I-Spy2 trial uses an adaptive approach to assigning patients to treatment arms (where patients with a tumor profile are more likely to be assigned to the arm with the best outcomes for that profile), and data safety monitoring board decisions are guided by Bayesian predicted probabilities of pathologic complete response.<sup>12,13</sup> Other examples of adaptive clinical trials in oncology include a ran-

domized trial of four regimens in metastatic prostate cancer, where patients who did not respond to their initial regimen (selected based on randomization) were then randomly assigned to the remaining three regimens,<sup>14</sup> and the CALGB (Cancer and Leukemia Group B) 49907 trial, which used Bayesian predictive probabilities of inferiority to determine the final sample size needed for the comparison of capecitabine and standard chemotherapy in elderly women with early-stage breast cancer.<sup>15</sup> Pragmatic trials relax some of the traditional rules of RCTs to maximize the relevance of the results for clinicians and policy makers. These changes may include expansion of eligibility criteria, flexibility in the application of the intervention and in the management of the control group, and reduction in the intensity of follow-up or procedures for assessing outcomes.<sup>16</sup>

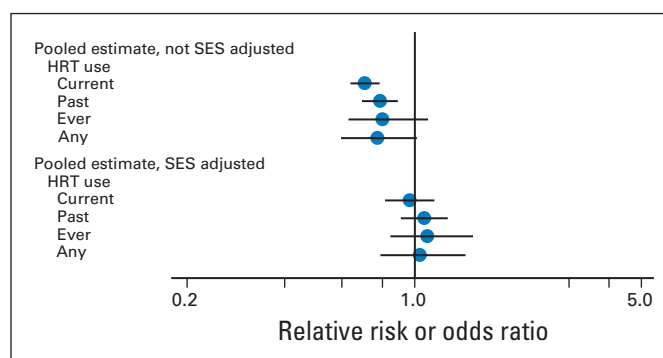
## OBSERVATIONAL METHODS

The emergence of comparative effectiveness has led to a renewed interest in the role of observational studies for assessing the benefits and harms of alternative interventions. Observational studies compare outcomes between patients who receive different interventions through some process other than investigator randomization. Most commonly, this process is the natural variation in clinical care, although observational studies also can take advantage of natural experiments, where higher-level changes in care delivery (eg, changes in state policy or changes in hospital unit structure) lead to changes in intervention exposure between groups. Observational studies can enroll patients by exposure (eg, type of intervention) using a cohort design or outcome using a case-control design. Cohort studies can be performed prospectively, where participants are recruited at the time of exposure, or retrospectively, where the exposure occurred before participants are identified.

The strengths and limitations of observational studies for clinical effectiveness research have been debated for decades.<sup>7,17</sup> Because the incremental cost of including an additional participant is generally low, observational studies often have relatively large numbers of participants who are more representative of the general population. Large, diverse study populations make the results more generalizable to real-world practice and enable the examination of variation in effect across patient subgroups. This advantage is particularly important for understanding effectiveness among vulnerable populations, such as

racial minorities, who are often underrepresented in RCT participants. Observational studies that take advantage of existing data sets are able to provide results quickly and efficiently, a critical need for most CER. Currently, observational data already play an important role in influencing guidelines in many areas of oncology, particularly around prevention (eg, nutritional guidelines, management of *BRCA1/2* mutation carriers)<sup>18,19</sup> and the use of diagnostic tests (eg, use of gene expression profiling in women with node-negative, estrogen receptor-positive breast cancer).<sup>20</sup> However, observational studies also have important limitations. Observational studies are only feasible if the intervention of interest is already being used in clinical practice; they are not possible for evaluation of new drugs or devices. Observational studies are subject to bias, including performance bias, detection bias, and selection bias.<sup>17,21</sup> Performance bias occurs when the delivery of one type of intervention is associated with generally higher levels of performance by the health care unit (ie, health care quality) than the delivery of a different type of intervention, making it difficult to determine if better outcomes are the result of the intervention or the accompanying higher-quality health care. Detection bias occurs when the outcomes of interest are more easily detected in one group than another, generally because of differential contact with the health care system between groups. Selection bias is the most important concern in the validity of observational studies and occurs when intervention groups differ in characteristics that are associated with the outcome of interest. These differences can occur because a characteristic is part of the decision about which treatment to recommend (ie, disease severity), which is often termed confounding by indication, or because it is correlated with both intervention and outcome for another reason. A particular concern for CER of therapies is that some new agents may be more likely to be used in patients for whom established therapies have failed and who are less likely to be responsive to any therapy.

There are two main approaches for addressing bias in observational studies. First, important potential confounders must be identified and included in the data collection. Measured confounders can be addressed through multivariate and propensity score analysis. A telling example of the importance of adequate assessment of potential confounders was found through examination of the observational studies of hormone replacement therapy (HRT) and coronary heart disease (CHD). Meta-analyses of observational studies had long estimated a substantial reduction in CHD risk with the use of postmenopausal HRT. However, the WHI (Women's Health Initiative) trial, a large, double-blind RCT of postmenopausal HRT, found no difference in CHD risk between women assigned to HRT or placebo. Although this apparent contradiction is often used as general evidence against the validity of observational studies, a re-examination of the observational studies demonstrated that studies that adjusted for measures of socioeconomic status (a clear confounder between HRT use and better health outcomes) had results similar to those of the WHI, whereas studies that did not adjust for socioeconomic status found a protective effect with HRT<sup>22</sup> (Fig 2). The use of administrative data sets for observational studies of comparative effectiveness is likely to become increasingly common as health information technology spreads, and data become more accessible; however, these data sets may be particularly limiting in their ability to include data on potential confounders. In some cases, the characteristics that influence the treatment decision may not be available in the data (eg, performance status, tumor gene expression), making concerns about confounding



**Fig 2.** Meta-analysis of observational studies of hormone replacement therapy (HRT) and coronary artery disease incidence comparing studies that did and did not adjust for socioeconomic status (SES). Data adapted.<sup>22</sup>

by indication too high to proceed without adjusting data collection or considering a different question.

Second, several analytic approaches can be used to address differences between groups in observational studies. The standard analytic approach involves the use of multivariable adjustment through regression models. Regression allows the estimation of the change in the outcome of interest from the difference in intervention, holding the other variables in the model (covariates) constant. Although regression remains the standard approach to analysis of observational data, regression can be misleading if there is insufficient overlap in the covariates between groups or if the functional forms of the variables are incorrectly specified.<sup>23</sup> Furthermore, the number of covariates that can be included is limited by the number of participants with the outcome of interest in the data set.

Propensity score analysis is another approach to the estimation of an intervention effect in observational data that enables the inclusion of a large number of covariates and a transparent assessment of the balance of covariates after adjustment.<sup>23-26</sup> Propensity score analysis uses a two-step process, first estimating the probability of receiving a particular intervention based on the observed covariates (the propensity score) and estimating the effect of the intervention within groups of patients who had a similar probability of receiving the intervention (often grouped as quintiles of propensity score). The degree to which the propensity score is able to represent the differences in covariates between intervention groups is assessed by examining the balance in covariates across propensity score categories. In an ideal situation, after participants are grouped by their propensity for being treated, those who receive different interventions have similar clinical and sociodemographic characteristics—at least for the characteristics that are measured (Table 1). Rates of the outcomes of interest are then compared between intervention groups within each propensity score category, paying attention to whether the intervention effect differs across patients with a different propensity for receiving the intervention. In addition, the propensity score itself can be included in a regression model estimating the effect of the intervention on the outcome, a method that also allows for additional adjustment for covariates that were not sufficiently balanced across intervention groups within propensity score categories.

The use of propensity scores for oncology clinical effectiveness research has become increasingly popular over the last decade, with six articles published in *Journal of Clinical Oncology* in 2011 alone.<sup>27-32</sup>

**Table 1.** Hypothetic Example of Propensity Score Analysis Comparing Two Intervention Groups, A and B

Characteristic	Quintiles of Propensity Score*											
	Overall Sample		1*		2		3		4		5*	
	A	B	A	B	A	B	A	B	A	B	A	B
Mean age, years	45.3	56.9	58.9	59.0	56.2	56.1	50.4	50.4	46.9	46.7	43.0	43.2
No. of comorbidities												
0	54.0	26.5	60.8	60.4	51.7	51.8	43.6	43.4	38.9	39	24.3	24.5
1-2	34.7	28.8	36.8	36.9	34.4	34.4	32	32.1	29.7	29.5	26.4	26.5
> 3	11.3	44.7	2.4	2.7	13.9	13.8	24.4	24.5	31.4	31.5	49.3	49

\*Quintile 1 is the lowest probability of receiving intervention A, and quintile 5 is the highest probability of receiving intervention A.

However, propensity score analysis has limitations, the most important of which is that it can only include the variables that are in the available data. If a factor that influences the intervention assignment is not included or measured accurately in the data, it cannot be adequately addressed by a propensity score. For example, in a prior propensity score analysis of the association between active treatment and prostate cancer mortality among elderly men, we were able to include only the variables available in Surveillance, Epidemiology, and End Results–Medicare linked data in our propensity score.<sup>33</sup> The data included some of the factors that influence treatment decisions (eg, age, comorbidities, tumor grade, and size) but not others (eg, functional status, prostate-specific antigen score). Furthermore, the measurement of some of the available factors was imperfect—for example, assessment of comorbidities was based on billing codes, which can underestimate actual comorbidity burden and provide no information about the severity of the comorbidity. Thus, although the final result demonstrating a fairly strong association between active treatment and reduced mortality was quite robust based on the data that were available, it is still possible that the association represents unaddressed selection factors where healthier men underwent active treatment.<sup>34</sup>

Instrumental variable methods are a third analytic approach that estimate the effect of an intervention in observational data without requiring the factors that differ between the intervention groups to be available in the data, thereby addressing both measured and unmeasured confounders.<sup>35</sup> The goal underlying instrumental variable analysis is to identify a characteristic (called the instrument) that strongly influences the assignment of patients to intervention but is not associated with the outcomes of interest (except through the intervention). In essence, an instrumental variable approach is an attempt to replicate an RCT, where the instrument is randomization.<sup>36</sup> Common instruments include the patterns of treatment across geographic areas or health care providers, the distance to a health care facility able to provide the intervention of interest, or structural characteristics of the health care system that influence what interventions are used, such as the density of certain types of providers or facilities. The analysis involves two stages: first, the probability of receiving the intervention of interest is estimated as a function of the instrument variable and other covariates; second, a model is built predicting the outcome of interest based on the instrument-based intervention probability and the residual from the first model.

Instrumental variable analysis is commonly used in economics<sup>37</sup> and has increasingly been applied to health and health care. In oncology, instrumental variable approaches have been used to examine the

effectiveness of treatments for lung, prostate, bladder, and breast cancers, with the most common instruments being area-level treatment patterns.<sup>38–42</sup> One recent analysis of prostate cancer treatment found that multivariable regression and propensity score methods resulted in essentially the same estimate of effect for radical prostatectomy, but an instrumental variable based on the treatment pattern of the previous year found no benefit from radical prostatectomy, similar to the estimate from a recently published trial.<sup>41,43</sup> However, concerns also exist about the validity of instrumental variable results, particularly if the instrument is not strongly associated with the intervention, or if there are other potential pathways by which the instrument may influence the outcome. Although the strength of the association between the instrument and the intervention assignment can be tested in the analysis, alternative pathways by which the instrument may be associated with the outcome are often not identified until after publication. A recent instrumental variable analysis used annual rainfall as the instrument to demonstrate an association between television watching and autism, arguing that annual rainfall is associated with the amount of time children watch television but is not otherwise associated with the risk of autism.<sup>44</sup> The findings generated considerable controversy after publication, with the identification of several other potential links between rainfall and autism.<sup>45</sup> Instrumental variable methods have traditionally been unable to examine differences in effect between patient subgroups, but new approaches may improve their utility in this important component of CER.<sup>46,47</sup>

## SYSTEMATIC REVIEWS

For some decisions faced by clinicians and policy makers, there is insufficient evidence to inform decision making, and new studies to generate evidence are needed. However, for other decisions, evidence exists but is sufficiently complex or controversial that it must be synthesized to inform decision making. Systematic reviews are an important form of evidence synthesis that brings together the available evidence using an organized and evaluative approach.<sup>48</sup> Systematic reviews are frequently used for guideline development and generally include four major steps.<sup>49</sup> First, the clinical decision is identified, and the analytic framework and key questions are determined. Sometimes the decision may be straightforward and involve a single key question (eg, Does drug A reduce the incidence of disease B?), but other times the question may be more complicated (eg, Should gene expression profiling be used in early-stage breast cancer?) and involve multiple

key questions.<sup>50</sup> Second, the literature is searched to identify the relevant studies using inclusion and exclusion criteria that may include the timing of the study, the study design, and the location of the study. Third, the identified studies are graded on quality using established criteria such as the CONSORT criteria for RCTs<sup>51</sup> and the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) criteria for observational studies.<sup>52</sup> Studies that do not meet a minimum quality threshold may be excluded because of concern about the validity of the results. Fourth, the results of all the studies are collated in evidence tables, often including key characteristics of the study design or population that might influence the results. Meta-analytic techniques may be used to combine results across studies when there is sufficient homogeneity to make a single-point estimate statistically valid. Alternatively, models may be used to identify the study or population factors that are associated with different results.

Although systematic reviews are a key component of evidence-based medicine, their role in CER is still uncertain. The traditional approach to systematic reviews has often excluded observational studies because of concerns about internal validity, but such exclusions may greatly limit the evidence available for many important comparative effectiveness questions. CER is designed to inform real-world decisions between available alternatives, which may include multiple tradeoffs. Inclusion of information about harms in comparative effectiveness systematic reviews is desirable but often challenging because of limited data. Finally, systematic reviews are rarely able to examine differences in intervention effects across patient characteristics, another important step for achieving the goals of CER.

## DECISION AND COST-EFFECTIVENESS ANALYSIS

Another evidence synthesis method that is gaining increasing traction in CER is decision modeling. Decision modeling is a quantitative approach to evidence synthesis that brings together data from a range of sources to estimate expected outcomes of different interventions.<sup>53</sup> The first step in a decision model is to lay out the structure of the decision, including the alternative choices and the clinical and economic outcomes of those alternatives.<sup>54</sup> Ensuring that the structure of the model holds true to the clinical scenario of interest without becoming overwhelmed by minor possible variations is critical for the eventual impact of the model.<sup>55</sup> Once the decision structure is determined, a decision tree or simulation model is created that incorporates the probabilities of different outcomes over time and the change in those probabilities from the use of different interventions.<sup>56,57</sup> To calculate the expected outcomes, a hypothetical cohort of patients is run through each of the decision alternatives in the model. Estimated outcomes are generally assessed as a count of events in the cohort (eg, deaths, cancers) or as the mean or median life expectancy among the cohort.<sup>58</sup>

Decision models can also include information about the value placed on each of the outcomes (often referred to as utility) as well as the health care costs incurred by the interventions and the health outcomes. A decision model that includes cost and utility is often referred to as a cost-benefit or cost-effectiveness model and is used in some settings to compare value across interventions. The types of costs that are included depend on the perspective of the model, with a model from the societal perspective including both direct and indirect medical costs (eg, loss of productivity), a model from a payer (ie,

insurer) perspective including only direct medical costs, and a model from a patient perspective including the costs experienced by the patient. Future costs are discounted to address the change in monetary value over time.<sup>59</sup> Sensitivity analyses are used to explore the impact of different assumptions on the model results, a critical step for understanding how the results should be used in clinical and policy decisions and for the development of future evidence-generation research. These sensitivity analyses often use a probabilistic approach, where a distribution is entered for each of the inputs and the computer samples from those distributions across a large number of simulations, thereby creating a confidence interval around the estimated outcomes of the alternative choices.

Decision models have several strengths in CER. They can link multiple sources of information to estimate the effect of different interventions on health outcomes, even when there are no studies that directly assess the effect of interest. Because they can examine the effect of variation in different probability estimates, they are particularly useful for understanding how patient characteristics will affect the expected outcomes of different interventions. Decision models can also estimate the impact of an intervention across a population, including the effect on economic outcomes. Decision and cost-effectiveness analyses have been used frequently in oncology, particularly for decisions with options that include the use of a diagnostic or screening test (eg, bone mineral density testing for management of osteoporosis risk),<sup>60</sup> involve significant tradeoffs (eg, adjuvant chemotherapy),<sup>61</sup> or have only limited empirical evidence (eg, management strategies in *BRCA* mutation carriers).<sup>62</sup>

However, decision models also have several limitations that have limited their impact on clinical and policy decision making in the United States to date and are likely to constrain their role in future CER. Often, model results are highly sensitive to the assumptions of the model, and removing bias from these assumptions is difficult. The potential impact of conflicts of interest is high. Decision models require data inputs. For many decisions, data are insufficient for key inputs, requiring the use of educated guesses (ie, expert opinion). The measurement of utility has proven particularly challenging and can lead to counterintuitive results. In the end, decision analysis is similar to other comparative effectiveness methods—useful for the right question as long as results are interpreted with an understanding of the methodologic limitations.

**Table 2.** Factors That Influence Selection of Study Design for Patient-Centered Outcome Research

Factor
Intrinsic
Internal validity (bias)
External validity (generalizability or applicability to nonstudy settings and populations)
Precision
Heterogeneity in risk or benefit (eg, subgroup or personalized evidence)
Ethical dimensions of the study (including considerations of risk-benefit balance and study burden for study participants)
Extrinsic
Time urgency (eg, rapidly changing technology, policy or public health needs)
Logistical constraints (eg, feasibility of collecting information from participants, number of participants available, study complexity)
Data availability, quality, and completeness

## SELECTION OF CER METHODS

The choice of method for a comparative effectiveness study involves the consideration of multiple factors. The Patient-Centered Outcomes Research Institute Methods Committee has identified five intrinsic and three extrinsic factors (Table 2), including internal validity, generalizability, and variation across patient subgroups as well as the feasibility and time urgency.<sup>63</sup> The importance of these factors will vary across the questions being considered. For some questions, the concern about selection bias will be too great for observational studies, particularly if a strong instrument cannot be identified. Many questions about aggressive versus less aggressive treatments may fall into this category, because the decision is often correlated with patient characteristics that predict survival but are rarely found in observational data sets (eg, functional status, social support). For other questions, concern about selection bias will be less pressing than the need for rapid and efficient results. This scenario may be particularly relevant for the comparison of existing therapies that differ in cost or adverse outcomes, where the use of the therapy is largely driven by practice style. In many cases, the choice will be pragmatic based on what data are available and the feasibility of conducting an RCT. These choices will increasingly be informed by the value of information

methods<sup>64-66</sup> that use economic modeling to provide guidance about where and how investment in CER should be made.

## DISCUSSION

In reality, the questions of CER are not new but are simply more important than ever. Nearly 50 years ago, Sir Austin Bradford Hill spoke about the importance of a broad portfolio of methods in clinical research, saying “To-day . . . there are many drugs that work and work potentially. We want to know whether this one is more potent than that, what dose is right and proper, for what kind of patient.”<sup>7(p109)</sup> This call has expanded beyond drugs to become the charge for CER. To fulfill this charge, investigators will need to use a range of methods, extending the experience in effectiveness research of the last decades “to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels.”<sup>1(p29)</sup>

## AUTHOR'S DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The author(s) indicated no potential conflicts of interest.

## REFERENCES

- Committee on Comparative Effectiveness Research Prioritization, Institute of Medicine: Summary: Initial National Priorities for Comparative Effectiveness Research. Washington, DC, The National Academies Press, 2009. p 29
- Berwick DM, Nolan TV, Whittington J: The triple aim: Care, health, and cost. *Health Aff (Millwood)* 27:759-769, 2008
- Brook RH: Can the Patient-Centered Outcomes Research Institute become relevant to controlling medical costs and improving value? *JAMA* 306:2020-2021, 2011
- Federal Coordinating Council: Report to the President and the Congress on Comparative Effectiveness Research. <http://www.hhs.gov/recovery/programs/ceer/execsummary.html>
- Sox HC, Goodman SN: The methods of comparative effectiveness research. *Annu Rev Public Health* 33:425-445, 2012
- Council MR: Streptomycin treatment of pulmonary tuberculosis: A report of the Streptomycin in Tuberculosis Trials Committee. *BMJ* 2:769-782, 1948
- Hill AB: Heberden oration: Reflections on the controlled trial. *Ann Rheum Dis* 25:107-113, 1966
- Tepper JE, Blackstock AW: Randomized trials and technology assessment. *Ann Intern Med* 151:583-584, 2009
- Luce BR, Kramer JM, Goodman SN, et al: Rethinking randomized clinical trials for comparative effectiveness research: The need for transformational change. *Ann Intern Med* 151:206-209, 2009
- Hutson S: A change is in the wind as “adaptive” clinical trials catch on. *Nat Med* 15:977, 2009
- Berry DA: Adaptive clinical trials in oncology. *Nat Rev Clin Oncol* 9:199-207, 2011
- Berry DA: Adaptive clinical trials: The promise and the caution. *J Clin Oncol* 29:606-609, 2011
- Barker AD, Sigman CC, Kelloff GJ, et al: I-SPY 2: An adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther* 86:97-100, 2009
- Thall PF, Logothetis C, Pagliaro LC, et al: Adaptive therapy for androgen-independent prostate cancer: A randomized selection trial of four regimens. *J Natl Cancer Inst* 99:1613-1622, 2007
- Muss HB, Berry DA, Cirincione CT, et al: Adjuvant chemotherapy in older women with early-stage breast cancer. *N Engl J Med* 360:2055-2065, 2009
- Thorpe KE, Zwarenstein M, Oxman AD, et al: A pragmatic-explanatory continuum indicator summary (PRECIS): A tool to help trial designers. *J Clin Epidemiol* 62:464-475, 2009
- Concato J, Lawler EV, Lew RA, et al: Observational methods in comparative effectiveness research. *Am J Med* 123:e16-e23, 2010 (suppl 1)
- Kushi LH, Doyle C, McCullough M, et al: American Cancer Society guidelines on nutrition and physical activity for cancer prevention: Reducing the risk of cancer with healthy food choices and physical activity. *CA Cancer J Clin* 62:30-67, 2012
- Domchek SM, Armstrong K, Weber BL: Clinical management of BRCA1 and BRCA2 mutation carriers. *Nat Clin Pract Oncol* 3:2-3, 2006
- Harris L, Fritsche H, Mennel R, et al: American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* 25:5287-5312, 2007
- Norris S, Atkins D, Bruening W: Selecting observational studies for comparing medical interventions, in *Methods Guide for Comparative Effectiveness Reviews*. Rockville, MD, Agency for Healthcare Research and Quality, 2010, pp 4-6
- Humphrey LL, Chan BK, Sox HC: Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease. *Ann Intern Med* 137:273-284, 2002
- Rubin DB: Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 127:757-763, 1997
- Rosenbaum P, Rubin D: The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41-55, 1983
- Rosenbaum PR, Rubin DB: Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc* 79:516-524, 1984
- D'Agostino RB Jr: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 17:2265-2281, 1998
- Ganz PA, Kwan L, Stanton AL, et al: Physical and psychosocial recovery in the year after primary treatment of breast cancer. *J Clin Oncol* 29:1101-1109, 2011
- Chen AB, Neville BA, Sher DJ, et al: Survival outcomes after radiation therapy for stage III non-small-cell lung cancer after adoption of computed tomography-based simulation. *J Clin Oncol* 29:2305-2311, 2011
- Cleeland CS, Mendoza TR, Wang XS, et al: Levels of symptom burden during chemotherapy for advanced lung cancer: Differences between public hospitals and a tertiary cancer center. *J Clin Oncol* 29:2859-2865, 2011
- O'Connor ES, Greenblatt DY, LoConte NK, et al: Adjuvant chemotherapy for stage II colon cancer with poor prognostic features. *J Clin Oncol* 29:3381-3388, 2011
- Punnen S, Cooperberg MR, Sadetsky N, et al: Androgen deprivation therapy and cardiovascular risk. *J Clin Oncol* 29:3510-3516, 2011
- Wang SJ, Lemieux A, Kalpathy-Cramer J, et al: Nomogram for predicting the benefit of adjuvant chemoradiotherapy for resected gallbladder cancer. *J Clin Oncol* 29:4627-4632, 2011
- Wong YN, Mitra N, Hudes G, et al: Survival associated with treatment vs observation of localized prostate cancer in elderly men. *JAMA* 296:2683-2693, 2006
- Litwin MS, Miller DC: Treating older men with prostate cancer: Survival (or selection) of the fittest? *JAMA* 296:2733-2734, 2006
- Angrist JD, Imbens GW, Rubin DB: Identification of causal effects using instrumental variables. *J Am Stat Assoc* 91:444-455, 1996

36. Rassen JA, Brookhart MA, Glynn RJ, et al: Instrumental variables I: Instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *J Clin Epidemiol* 62:1226-1232, 2009
37. Newhouse JP, McClellan M: Econometrics in outcomes research: The use of instrumental variables. *Annu Rev Public Health* 19:17-34, 1998
38. Lu-Yao GL, Albertsen PC, Moore DF, et al: Survival following primary androgen deprivation therapy among men with localized prostate cancer. *JAMA* 300:173-181, 2008
39. Wisnivesky JP, Halm E, Bonomi M, et al: Effectiveness of radiation therapy for elderly patients with unresected stage I and II non-small cell lung cancer. *Am J Respir Crit Care Med* 181:264-269, 2010
40. Gore JL, Litwin MS, Lai J, et al: Use of radical cystectomy for patients with invasive bladder cancer. *J Natl Cancer Inst* 102:802-811, 2010
41. Hadley J, Yabroff KR, Barrett MJ, et al: Comparative effectiveness of prostate cancer treatments: Evaluating statistical adjustments for confounding in observational data. *J Natl Cancer Inst* 102:1780-1793, 2010
42. Saito AM, Landrum MB, Neville BA, et al: Hospice care and survival among elderly patients with lung cancer. *J Palliat Med* 14:929-939, 2011
43. Bill-Axelson A, Holmberg L, Ruutu M, et al: Radical prostatectomy versus watchful waiting in early prostate cancer. *N Engl J Med* 364:1708-1717, 2011
44. Waldman M, Nicholson S, Adilov N: Does Television Cause Autism? Cambridge, MA, National Bureau of Economic Research, 2006
45. Whitehouse M: Is an economist qualified to solve puzzle of autism? *Wall Street Journal*, February 27, 2007:A1
46. Basu A, Heckman JJ, Navarro-Lozano S, et al: Use of instrumental variables in the presence of heterogeneity and self-selection: An application to treatments of breast cancer patients. *Health Econ* 16:1133-1157, 2007
47. Brooks JM, Chrischilles EA: Heterogeneity and the interpretation of treatment effect estimates from risk adjustment and instrumental variable methods. *Med Care* 45:S123-S130, 2007
48. Chang SM: The Agency for Healthcare Research and Quality (AHRQ) effective health care (EHC) program methods guide for comparative effectiveness reviews: Keeping up-to-date in a rapidly evolving field. *J Clin Epidemiol* 64:1166-1167, 2011
49. Harris RP, Helfand M, Woolf SH, et al: Current methods of the US Preventive Services Task Force: A review of the process. *Am J Prev Med* 20:21-35, 2001
50. Teutsch SM, Bradley LA, Palomaki GE, et al: The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: Methods of the EGAPP Working Group. *Genet Med* 11:3-14, 2009
51. Altman DG, Schulz KF, Moher D, et al: The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 134:663-694, 2001
52. Vandembroucke JP, von Elm E, Altman DG, et al: Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *Ann Intern Med* 147:W163-W194, 2007
53. Pauker SG, Kassirer JP: Decision analysis. *N Engl J Med* 316:250-258, 1987
54. Detsky AS, Naglie G, Krahn MD, et al: Primer on medical decision analysis: Part 1. Getting started. *Med Decis Making* 17:123-125, 1997
55. Weinstein MC, Siegel JE, Gold MR, et al: Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA* 276:1253-1258, 1996
56. Detsky AS, Naglie G, Krahn MD, et al: Primer on medical decision analysis: Part 2. Building a tree. *Med Decis Making* 17:126-135, 1997
57. Sonnenberg FA, Beck JR: Markov models in medical decision making: A practical guide. *Med Decis Making* 13:322-338, 1993
58. Krahn MD, Naglie G, Naimark D, et al: Primer on medical decision analysis: Part 4. Analyzing the model and interpreting the results. *Med Decis Making* 17:142-151, 1997
59. Garber AM, Phelps CE: Economic foundations of cost-effectiveness analysis. *J Health Econ* 16:1-31, 1997
60. Ito K, Elkin EB, Girotra M, et al: Cost-effectiveness of fracture prevention in men who receive androgen deprivation therapy for localized prostate cancer. *Ann Intern Med* 152:621-629, 2010
61. Hillner BE, Smith TJ: Efficacy and cost effectiveness of adjuvant chemotherapy in women with node-negative breast cancer: A decision-analysis model. *N Engl J Med* 324:160-168, 1991
62. Grann VR, Jacobson JS, Thomason D, et al: Effect of prevention strategies on survival and quality-adjusted survival of women with *BRCA1/2* mutations: An updated decision analysis. *J Clin Oncol* 20:2520-2529, 2002
63. Patient-Centered Outcomes Research Institute: Research methodology. <http://www.pcori.org/what-we-do/methodology/>
64. Hoomans T, Fenwick EA, Palmer S, et al: Value of information and value of implementation: Application of an analytic framework to inform resource allocation decisions in metastatic hormone-refractory prostate cancer. *Value Health* 12:315-324, 2009
65. Claxton K: Bayesian approaches to the value of information: Implications for the regulation of new pharmaceuticals. *Health Econ* 8:269-274, 1999
66. Basu A, Meltzer D: Modeling comparative effectiveness and the value of research. *Ann Intern Med* 151:210-211, 2009

