Oncology Comparative Effectiveness Research: A Multistakeholder Perspective on Principles for Conduct and Reporting

SCOTT D. RAMSEY,^a SEAN D. SULLIVAN,^b SHELBY D. REED,^c YA-CHEN TINA SHIH,^d KEN SCHAECHER,^e RAHUL DHANDA,^f DEBRA PATT,^{f,g} KELLY PENDERGRASS,^h MARK WALKER,ⁱ JENNIFER MALIN,^j LEE SCHWARTZBERG,^k KURT NEUMANN,^l ELAINE YU,^m ARLIENE RAVELO,^m ART SMALL^m

^aFred Hutchinson Cancer Research Center, Seattle, Washington, USA; ^bDepartment of Pharmacy, University of Washington, Seattle, Washington, USA; ^cDuke Clinical Research Institute, Duke University, Durham, North Carolina, USA; ^dDepartment of Medicine, University of Chicago, Chicago, Illinois, USA; ^eSelectHealth, Salt Lake City, Utah, USA; ^fMcKesson Speciality Health, Woodlands, Texas, USA; ^gU.S. Oncology Network, Austin, Texas, USA; ^hKansas City Cancer Center, Kansas City, Missouri, USA; ⁱAcorn Research, Memphis, Tennessee, USA; ^jWellPoint, Thousand Oaks, California, USA; ^kThe West Clinic, Memphis, Tennessee, USA; ^lION Solutions, Frisco, Texas, USA; ^mGenentech, San Francisco, California, USA

Disclosures of potential conflicts of interest may be found at the end of this article.

Key Words. Comparative effectiveness • Oncology • Costs

Learning Objectives

Compare well-conducted CER and phase I-III clinical trials.

Describe barriers to the acceptance of CER studies by the oncology community.

Demonstrate the use of CER for decision-making in oncology.

ABSTRACT

Comparative effectiveness research (CER) can assist patients, clinicians, purchasers, and policy makers in making more informed decisions that will improve cancer care and outcomes. Despite its promise, the factors that distinguish CER from other types of evidence remain mysterious to many oncologists. One concern is whether CER studies will improve decision making in oncology or only add to the massive amount of research information that decision makers must sift through as part of their professional responsibilities. In this report, we highlight several issues that distinguish CER from the most common way evidence is generated for cancer therapy—

phase I-III clinical trials. To identify the issues that are most relevant to busy decision makers, we assembled a panel of active professionals with a wide range of roles in cancer care delivery. This panel identified five themes that they considered most important for CER in oncology, as well as fundamental threats to the validity of individual CER studies—threats they termed the “kiss of death” for their applicability to practice. In discussing these concepts, we also touched upon the notion of whether cancer is special among health issues with regard to how evidence is generated and used. *The Oncologist* 2013;18:760–767

Implications for Practice: Comparative effectiveness research can inform clinical practice when studies directly compare competing therapies, include patient populations that are typical of clinical practice, and include outcomes that are meaningful to patients as well as clinicians. Busy clinicians should not waste time reading studies that are technically dense, compare outmoded treatments, have very small sample sizes, use data of questionable quality, or use dense statistical methods to address serious limitations of retrospective data.

INTRODUCTION

In response to government and private sector initiatives, a tremendous amount of research in oncology will be labeled (and re-labeled) as “comparative effectiveness research” (CER). Given the wide variety of study designs that could potentially be considered as CER, it is reasonable to consider whether it is desirable and feasible to have a set of principles for the conduct and reporting of CER that are transparent, straightforward, and useful to busy decision makers. A related issue is whether standards for

evaluating CER would differ from existing criteria for evaluating evidence. Finally, many view cancer as a unique area in health care, and this may imply that “special” principles should be applied when evaluating CER in oncology versus other medical conditions.

In this article, we review the definition and purpose of CER and commonly agreed-upon principles for the conduct, reporting, and assessment of CER in oncology, highlighting differences from standard approaches to evidence development for cancer

Correspondence: Scott D. Ramsey, Ph.D., Member, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, M3-B232, Seattle, Washington 98109, USA. Telephone: 206-667-7846; Fax: 206-667-5977; E-Mail: sramsey@fhcrc.org Received September 21, 2012; accepted for publication February 1, 2013; first published online in *The Oncologist Express* on May 6, 2013. ©AlphaMed Press 1083-7159/2013/\$20.00/0 <http://dx.doi.org/10.1634/theoncologist.2012-0386>

treatments. To identify the challenges and opportunities that CER represents for decision makers, a panel of decision makers in oncology (including community practitioners and representatives from the health insurance, academics, policy, and pharmaceutical industries) was assembled. The panel provided their views on the opportunities and challenges that CER represents from their perspectives, specifically in the context of demanding professions for which evidence appraisal is just one of many job responsibilities. The panel was asked to identify specific issues that prevent or inhibit the use of CER in practice—a concept they defined as “kisses of death” for published CER research studies. Finally, we asked whether cancer itself warrants special consideration for evidence generators, evidence appraisers, and oncology practitioners and policy makers.

BRIEF SYNOPSIS OF COMPARATIVE EFFECTIVENESS RESEARCH

CER has been defined as “the generation and synthesis of evidence that compares the effectiveness of alternative methods to prevent, diagnose, treat, monitor, and improve delivery of care for a clinical condition” [1]. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels. Certain aspects of CER make it distinct from other types of clinical research. First, CER is focused on direct, head-to-head comparisons of competing treatments, with topics directly oriented to clinical decision making. Second, the study design and endpoints for CER should be relevant to a wide group of stakeholders, such as patients, clinicians, purchasers, and policy makers. Finally, study populations for CER should be generally representative of those in clinical practice.

FUNDING AND PRIORITIZING TOPICS FOR COMPARATIVE EFFECTIVENESS RESEARCH IN CANCER

As one of the pillars of the U.S. health care reform law, CER was funded with the hope that bringing better evidence to bear on “real-world” patients and problems would reduce practice variations, improve health outcomes, and reduce costs of care. Although the Institute of Medicine identified several areas in oncology among its list of 100 priority areas for CER, the research agenda for CER in oncology has not been established, making it important for the oncology community to develop its own guiding principles for prioritizing CER studies [2]. For CER to be successful in changing oncology practice, many stakeholders groups who have been traditionally left out of the research enterprise must be engaged in the priority setting process. In addition, although study design will remain in the purview of methodologists, it will be vital to consider the information needs and time constraints of clinical and managerial professionals in oncology, particularly community oncologists and decision makers in health insurance, given the influence these individuals have on oncology care and patients with cancer.

ONCOLOGY DRUG DEVELOPMENT TRIALS AND COMPARATIVE EFFECTIVENESS RESEARCH

The clinical development process for a new oncology drug is typically guided by the objective of attaining regulatory approval for the product through Phase I–III trials designed to evaluate safety and efficacy [3]. Drug development trials are

comparative in that they evaluate the performance of a new drug relative to standard care or as an addition to standard care. Eligible patients typically have a specific stage of disease (e.g., late-stage disease vs. adjuvant therapies for early stage), have failed other therapies, and are highly selected (e.g., older patients and those with comorbidities or poor performance status are excluded). Studies are conducted in specialized treatment facilities and often limited to narrowly defined clinical endpoints. Because endpoint selection drives sample size calculations, secondary endpoints such as quality of life, survival, medical resource use and costs, and cost-effectiveness are often inadequately powered for detecting statistically significant differences.

Phase IV studies, or “post-marketing studies to delineate additional information including the drug’s risks, benefits, and optimal use” [4] could be construed as CER, particularly if the studies pair a new treatment with commonly used treatments. Nevertheless, like phase III trials, phase IV studies are often focused on a narrow group of stakeholder interests, primarily (a) meeting the U.S. Food and Drug Administration (FDA) requirements for monitoring safety, (b) evaluating in other populations, and (c) providing data to promote adoption of a new treatment. These studies represent a limited step toward addressing the goals of CER.

Thus, the phase I–IV trial process leaves many unanswered questions, such as how a drug will perform in practice in broader patient populations, longer term safety and effectiveness, and its broader impact on patient care and outcomes. These issues are exacerbated by off-label use of oncology products after FDA approval. Filling this broad evidence gap represents a tremendous opportunity for comparative effectiveness researchers (Table 1).

STAKEHOLDER PERSPECTIVES ON COMPARATIVE EFFECTIVENESS RESEARCH: NEEDS, STRENGTHS, AND CONCERNS

To begin the process of developing consensus on guiding principles regarding the priorities and methods of CER for oncology, we convened a multistakeholder consensus panel, taking into special consideration the day-to-day needs of community oncology professionals and health care payers. Fifteen stakeholders representing a broad array of perspectives (supplemental online data) met over 2 days to discuss these issues. Participants were asked to specifically consider the daily demands on time-constrained decision makers involved in the provision and payment for oncology care. The stakeholder discussions centered on five themes discussed in the following sections.

Retrospective Evaluations of Real-World Data Versus Prospective Studies as a Basis for Decision Making

In a perfect world, adequate time and resources would be available to execute well-designed, randomized clinical trials to inform clinical, payment, and policy decisions. Reality dictates trade-offs. An important distinction of CER from traditional clinical research and evidence synthesis is its explicit acceptance of both prospective and retrospective studies to compare treatment alternatives. This acceptance, however, presents challenges to those who design and execute studies as well as those who must translate the results into actionable decisions.

Table 1. Contrasting comparative effectiveness research designs with phase III and IV studies for cancer

Content area	Phase III study	Phase IV study	Comparative effectiveness research
Priority setting	Clinical opportunity as dictated by expert assessment of emerging technology	When FDA mandated	Multiple dimensions (e.g., disease burden, cost, information gap, variability in care) representing perspectives of various stakeholders
Study designs	Prospective randomized trials	Prospective trials; not always randomized	Retrospective and prospective studies
Comparisons	Head-to-head comparison of new therapy or regimen with what is considered the one best alternative treatment	Head-to-head comparison of new therapy, in combination with other agents, and/or in new populations, or added to other regimens to assess safety and efficacy; rarely addresses “usual care”	Direct, head-to-head comparisons of new therapy with “usual care” (the latter may be multiple different therapies)
Study questions	Highly focused	Intermediate	Broad
Perspectives	Researcher, clinician experts, regulatory authorities (e.g., FDA) and selected patients	Researcher, clinician experts, practicing physicians	Multiple (patient advocates, clinicians, payers, policy makers)
Study populations and setting	Highly selected, populations prespecified based on mechanistic research hypotheses to maximize likelihood of testing hypothesis of efficacy	Selected populations, but with attention to including, as much as possible, patients underrepresented in pivotal trial	Representative of clinical practice, with efforts to evaluate subpopulations (e.g., women, elderly, genetic subtypes) to maximize ability to assess effectiveness in target population
Funding	Single source: private investor, corporate, or NCI funding mechanisms	Single source: private investor, corporate, or NCI funding mechanisms	Multistakeholder, “coverage with evidence” programs, public-private partnerships
Endpoints	Clinical- or research-driven Efficacy: RR, TTP, PFS, sometimes OS, sometimes QoL Safety: common AEs (due to sample size) and will not capture late	Clinical- or research-driven Efficacy: RR, TTP, PFS, sometimes OS, sometimes QoL Safety: Common AEs (due to sample size) and will not capture late effects unless patient registry added on	Clinical effectiveness Effectiveness: Survival, QoL Safety: Uncommon and long-term AEs Other: Utility/patient preference weighting of outcomes, resource utilization and costs, function/disability, absenteeism/productivity, impact on caregivers, patient satisfaction

Abbreviations: AE, adverse event; FDA, U.S. Food and Drug Administration; NCI, National Cancer Institute; OS, overall survival; PFS, progression-free survival; QoL, quality of life; RR, response rate; TTP, time to progression.

Published Studies

The stakeholder panel agreed that when they want information about the relative effectiveness of treatment alternatives, the search typically begins with the scientific literature, but that the literature typically has serious limitations for decision making. As noted, even well-executed, prospectively designed studies often have limited relevance to a given stakeholder. Many retrospective studies have been published to help fill evidence gaps left by prospective studies. However, these retrospective studies can suffer from many of the same issues as prospective studies (e.g., limited generalizability, limited range of outcomes, and limited analysis of subgroups) in addition to having issues of confounding that adversely affect the integrity of nonrandomized treatment comparisons (Table 2).

Retrospective Studies

The stakeholder group noted that several characteristics of retrospective analyses are key factors in their value to decision makers. Retrospective studies often can be accomplished more quickly than prospective studies and often require fewer resources. An additional factor favoring retrospective studies for cancer is that testing of endpoints that are not always pri-

mary in drug development studies—most notably survival—are often feasible in retrospective studies. For example, a study using the Surveillance, Epidemiology and End Results (SEER)-Medicare database examined survival differences among common chemotherapy regimens used among older patients with advanced-stage non-small cell lung cancer and found inferior survival outcomes for those who did not receive platinum-containing agents, regardless of the nonplatinum agent used in the doublet [5].

Another advantage of retrospective data is that they offer an opportunity to study the effectiveness of new technologies that are not pharmaceuticals (e.g., new modality of radiation therapy or new surgical technique) and thus are not subject to the same requirement of clinical trials as the FDA drug approval process. Often these new technologies are adopted in oncology practice before conclusive evidence is reached in clinical trials. Smith et al. used a nationwide cohort of older patients with breast cancer extracted from national Medicare data to compare the effectiveness of breast brachytherapy, a newer form of radiation treatment, with the standard whole-breast irradiation (WBI). The study measured effectiveness in terms of cumulative incidence of subsequent mastectomy (an

Table 2. Advantages and disadvantages of prospective and retrospective research designs

Research issue	Designing a study	
	Retrospective analysis	Prospective study
Practical considerations		
Feasibility	Advantage	Disadvantage
Timeliness	Advantage	Disadvantage
Cost	Advantage	Disadvantage
Methodological considerations		
Representativeness of patient population	Advantage	Advantage
Treatments selected for comparison	Variable, but limited to prior availability of treatment alternatives	Advantage
Confounding	Disadvantage	Advantage with randomization; advantage without randomization, data collection on potential confounders
Range of outcomes considered	Disadvantage, but limited to existing data	Advantage
Statistical power to detect differences between comparison groups	Variable ^a	Variable ^a
Ability to evaluate subgroups	Advantage, but limited to existing data	Advantage
Ability to evaluate/perform sensitivity analyses	Advantage, but limited to existing data	Advantage

^aStatistical power is a function of the frequency of an outcome, sample size, and effect size. Thus, it will vary across studies and within an individual study across different outcomes.

indication of failure to preserve the breast), 5-year overall survival, and postoperative and postradiation complications. The authors concluded that breast brachytherapy was associated with worse long-term breast preservation and excess risk of complications compared with WBI, and it offered no survival gain. Findings from this study contribute to the existing controversy over the appropriateness of widespread adoption of breast brachytherapy observed in the current practice despite the fact that results of a long-term randomized trial comparing these two radiation treatment modalities will not be completed for years [6].

An important limiting factor for retrospective comparative studies, however, is whether the intervention has been used and whether its use can be identified in existing data systems. For example, when a product is new to the marketplace or has not been used in the target population (e.g., isotretinoin for chemoprevention of nonmelanoma skin cancer), a retrospective comparative study is not possible. Even when feasible, one must consider whether adequate information regarding potential confounding factors (e.g., performance status at diagnosis and comorbidities, disease severity, or histopathologic features that convey variable prognoses) are available in the existing data to compare relevant outcomes across alternative treatments. In many cases, the quality of a retrospective analysis is limited by the lack of detail about patients' disease characteristics, preferences for treatment, and range of outcomes documented in retrospective data sources. Similar concern over the impact of unobserved confounders on outcomes was also expressed in the study by Smith et al. [6]. Currently, few data systems include adequate information on disease severity or patient-reported outcomes, such as quality of life and patient preferences.

Prospective Studies

Even when a retrospective data analysis is feasible, the stakeholder group noted that a prospective study may represent a good investment of resources when the stakes are high, due to the size of the target population affected or high costs associated with a treatment. The resources to be expended on the execution of a prospective study should be proportional to the expected impact that information from the study would have on medical care and outcomes.

As a basis for decision making, executing a prospective study offers several advantages over conducting a retrospective study (Table 2). First, if randomization is an option, concerns about confounding are minimized. Even with non-randomized designs, prospectively collecting data on variables that may be associated with treatment selection and patient outcomes allow researchers better control over potential confounding. Prospective data collection allows for fuller characterization of patients' conditions, increasing flexibility to adjust for characteristics that may differ between patients receiving one treatment versus another, and evaluation of patient subgroups. Prospective designs also allow for the collection of information on patient-reported outcomes such as health-related quality of life, productivity and employment status, and other outcomes not typically available in existing data sets. In addition, prospective data collection offers the opportunity to ascertain other information rarely available in retrospective data, such as physician-reported reasons for treatment selection and discontinuation or change in treatment. Finally, prospectively collected data can be monitored to reduce missing data, and systematic measurement procedures can be implemented.

For example, on the basis of relatively small but suggestive retrospective studies, the 21-gene assay (OncotypeDX,

Genomic Health, Redwood City, CA) is increasingly being used to guide therapy decisions in women with hormone receptor (HR)-positive, HER2-negative breast cancer [7]. To address this issue, the Southwest Oncology Group is sponsoring a prospective, randomized study to determine the effect of chemotherapy in patients with node-positive breast cancer who do not have high recurrence scores (RS) by OncotypeDX. Outcomes will include disease-free survival, quality of life, and cost-effectiveness [8]. All of these advantages come with two significant drawbacks that are practical in nature: cost and time. The evaluation of OncotypeDX will require testing of 9,400 women, with a goal of recruiting 4,000 to the randomized trial, and will have a follow-up period of 5 years. The value of costly prospective studies must be considered in the context of their potential influence on treatment recommendations or payment decisions. A prospective study comparing a newly available treatment to standard care could be of little value to a health insurer if regulations required that both treatment options be granted the same coverage status. Likewise, there may be little value to conducting a study of a treatment during its early adoption phase. The patients enrolled in the study who receive the new treatment may have more advanced disease, and there may be a meaningful learning curve before physicians acquire sufficient experience to optimally apply the treatment and manage its complications. Thus, neither the patients nor the way in which the treatment is used during its early days may be useful in predicting the relative effectiveness of the treatment in the future.

As CER becomes integrated into the evaluation and choice of treatment alternatives and the information technology infrastructure is strengthened, these developments could lead to a “learning health care system” in which important data elements are prospectively collected and stored to facilitate the conduct of high-quality retrospective analyses [9]. With this as a goal, the distinction between prospective and retrospective analyses may one day become blurred.

The Need To Consider Cost in CER

Although the current political environment discourages the inclusion of costs (in particular, cost-effectiveness) in government-funded CER, all stakeholders emphasized that cost is a vital issue for CER that cannot be ignored because it is central to health care decision making from multiple perspectives. Oncologists on the panel noted that many of their colleagues now take cost into consideration when making decisions about treatments, in part because of mindfulness about the impact of their decisions on their patient’s out-of-pocket expenses. Others commented that specialty societies such as the American Society for Clinical Oncology and patient advocacy groups are acknowledging that patients are becoming increasingly affected by the cost of oncology care [10, 11]. In time, comparing outcomes per cost for relative value will become more important in distinguishing clinically meaningful and economically efficient interventions.

Acknowledging that studies of health care costs have their own methodological issues, stakeholders nonetheless agreed that transparency and uniformity of reporting of costs in CER studies is critical. At a minimum, studies need to consider up-front costs (e.g., direct medical and drug costs) and downstream cost offsets within a time horizon that is meaningful to those affected by costs. Considerations of the broader context

of costs, such as the impact of treatments on worker productivity of patients and their family caregivers, are also important.

The stakeholders were of mixed opinions regarding the value of cost-effectiveness studies where the outcome was expressed as the additional cost to achieve a one-unit gain in quality-adjusted life years (QALYs; i.e., the incremental cost-effectiveness ratio). Some felt that this was an emerging metric in the U.S. that will be useful in the future as cost pressures escalate. Others noted that most payers do not understand QALYs or use them in day-to-day decision making. Some stated that, for payers, the per-member-per-month cost reported from budget impact analyses often provide more practical and actionable information than cost-effectiveness analyses. The lack of benchmarks for conferring good value versus poor value medical interventions in the U.S. makes it difficult to apply results of economic evaluations in decision making.

Identifying “Outcomes That Matter” and Defining Clinically Meaningful Differences in Those Outcomes

Many stakeholders expressed frustration with the proliferation of outcome measures in oncology research. Although all agreed that overall survival was highly valid, some noted that the movement toward progression-free survival (PFS) as a primary outcome is problematic, in part because gains in PFS are not always predictive of improvements in survival. Additionally, although PFS may provide evidence of disease control and, by extension, tumor-related symptoms, many stressed that the benefit of delaying disease progression and the impact of treatment toxicity on patients should be assessed with patient-reported outcomes measures, including health-related quality of life and symptom measures. Acknowledging that different stakeholders weigh the outcomes differently, the group noted that multistakeholder groups could collaborate to identify primary and secondary outcomes for CER studies. The group strongly agreed that this would be preferable to clinician- or manufacturer-directed selection of primary outcomes, as often occurs today.

A second area of discussion was the lack of clarity on what constitutes a clinically meaningful difference in a particular outcome, as a statistically significant difference does not necessarily imply such. Those in decision-making roles noted the need for reference standards against which to compare study results. The group acknowledged the problem that, when determining clinically meaningful differences, “where one stands depends on where one sits”—that is, patients may have different opinions than insurance companies. In addition, a clinically meaningful benefit for one cancer may be different than for another, depending on the lethality of the disease and availability of suitable alternative treatments. Others noted that even among the public there would be differences, particularly between those who are affected with the condition of interest (or have an affected family member) or not. All agreed that consensus on minimal clinically meaningful differences must be developed from multistakeholder discussions.

Separating Well-Conducted From Poorly Conducted Comparative Effectiveness Research

As methods for conducting CER become more complex, some stakeholders acknowledged that they have difficulty distin-

guishing well-conducted studies from flawed studies [12]. Acknowledging the complexities of designing prospective and retrospective studies and analyzing data collected from these studies, all agreed that the needs of busy decision makers are being ignored when journals publish studies that employ various sophisticated methods, produce conflicting findings, and do not provide clear interpretation of the findings or explain reasons that drove the difference in findings. The needs were felt to be particularly acute for retrospective analyses.

Stakeholder “Kisses of Death” for Journal Articles

Given the current state of the literature, their individual needs, and training, the group considered the following features to be “kisses of death” when reviewing evidence generated from observational studies:

1. Extensive use of technical jargon ending with an ambiguous conclusion
2. A comparison group that consists of outmoded treatment or incomparable patient population
3. Small sample sizes, particularly when studying a highly prevalent condition
4. Questionable data quality
5. Retrospective studies with dense yet poorly described statistics sections

The group acknowledged that knowing the “true” effect of a new therapy is critical for their decision making, but they commented that authors wanting their research to have a “real-world” impact should avoid using technical jargon in their discussions of study findings and include clear “punch lines” for lay readers in the authors’ concluding remarks. If the study findings are dependent on a particular issue, the authors should clearly articulate this using language that is accessible to a wide range of readers. Alternatively, if the study’s limitations were so significant that a specific treatment recommendation cannot be made, authors should unambiguously acknowledge this to steer stakeholders away from using the study as a basis for a decision.

Recognizing that cancer care is evolving rapidly, stakeholders expressed frustration when reviewing retrospective studies of treatments that clinicians view as outmoded or no longer in use. For example, the SEER-Medicare database is featured in over 650 peer-reviewed publications as of December 2011 [13] and has addressed clinical issues that are often difficult to explore in clinical trials, such as adherence to practice guidelines, long-term complications, and variation in practice pattern or diffusion of new technology by geographic, patient, or provider characteristics. Yet, even with these important contributions to cancer research, some stakeholders noted important limitations in SEER-Medicare data, including the lag in updates (SEER-Medicare records are updated every 3 years), and that many publications consider treatments that are no longer used. All noted that the rapid pace of change in oncology poses a major challenge for retrospective CER.

Most stakeholders noted that studies making inferences of effect using very small samples, either as the primary study cohort or in subgroup analysis, were rarely considered valid for decision making. The question of “How small is too small?” inevitably arose but was difficult to answer. In cases of extremely rare conditions, a small sample size may be accept-

able. However, when studying conditions with greater prevalence, samples of less than 50 patients or those with somewhat larger populations but other important methodological limitations were seen as highly suspect. Stakeholders said the problem is compounded when the study applies multiple exclusion criteria such that the study sample becomes unrepresentative of patients in mainstream clinical practice.

The stakeholders noted that they often ignored studies based on databases that have limited validity or serious data flaws. Commonly cited problems included the lack of information on key clinical variables (e.g., a lack of cancer stage in the comparison of survival among patients with colorectal cancer between different chemotherapy regimens) and having a large proportion of missing values on an important factor (e.g., HER2 status in breast cancer). Retrospective studies that focus on data elements known to have serious quality limitations were also considered fatally flawed and not worth consideration.

A central challenge to CER using observational data is the lack of randomization to treatment assignment, which necessitates the use of more advanced methods to test for differences in health outcomes or costs between treatments. Although a number of methods have been developed to mitigate biases that arise when the treatment a patient receives is not randomly assigned (e.g., instrumental variable, difference-in-difference, or propensity score methods), the stakeholders noted that these methods are often not described in ways that can easily be understood by nonstatisticians and, as a result, are often incomprehensible to them.

REPORTING CER TO MEET THE NEEDS OF BUSY DECISION MAKERS

Transparency and Clarity

All stakeholders agreed that potential biases in CER studies are not always obvious and that transparency regarding study methods was critical to evaluation of findings. Discussion highlighted the increasing complexity of CER methods and the importance of making methods understandable for decision makers. The group identified consistency in the reporting of study methods, with definition of terms and other efforts to educate readers, as processes that would make CER reporting more meaningful. Several stakeholders suggested methods for standardizing the reporting of CER studies, including a checklist verifying the methodological issues that were addressed in a report of a CER study, statements of potential author bias (similar to conflict of interest disclosure), and establishment of a prospective registry of CER studies that would require completion of the aforementioned checklist. These methods would establish a threshold for consideration of self-reported CER for policy makers.

Timeliness

Many stakeholders emphasized that, for CER to be most impactful, studies must be timely and relevant to practice. Prospective and retrospective studies must be published quickly, particularly when they concern new treatments that are emerging because health plans, patients, and physicians have not yet established policies that “set the course” for these interventions. Retrospective studies that rely on databases that are several years old are viewed as problematic, even when published quickly. It was noted that studies based on the SEER-

Table 3. Top recommendations from the stakeholder panel for comparative effectiveness research studies

Design
A diverse group of stakeholders should be engaged to inform the research question and study design
The research design should be fit for purpose. Not all important CER questions require a randomized trial
The most commonly used clinically relevant comparators should be included in the study
Priority outcomes noted by the stakeholders should be included as endpoints in the study
Analysis
Appropriate analytic methods should be rigorously applied, particularly for observational research where confounding can be problematic
Heterogeneity and uncertainty should be identified and assessed
Benefits and harms of all comparators should be reported
Reporting
Investigators must be transparent in reporting conflicts and potential sources of bias
Peer review is an important part of reporting and dissemination

Medicare database, which is updated every 3 years, suffer from this problem for many applications.

Publication and Peer Review

Although timeliness is important, the stakeholders noted that they rely on the peer review process to weed out seriously flawed studies. They noted that peer review may not limit publications that have little importance for decision making, while acknowledging that some studies have more significance in academics than in practice. Table 3 lists the top recommendations for comparative effectiveness research studies from the stakeholder panel.

CONCLUSIONS

CER offers the promise of providing a framework for evidence generation that is both complementary and challenging to the clinical trial-driven model of evaluating new and existing treatments in oncology. Endpoints that are not often primary endpoints in trials, such as long-term survival and patient-reported outcomes, are central aspects of CER. At the same time, by moving away from the highly selected populations of clinical trials in favor of patients that are more representative of clinical practice, CER may reveal that outcomes shown in trials cannot be dupli-

cated. In addition, retrospective CER studies will allow for more timely evaluation of existing treatments as they are given to “real-world” populations at a fraction of the cost of carrying out a prospective clinical trial. The key issue—and challenge—for CER in oncology will be the extent to which the decision makers accept the methods and outcomes of CER, relative to phase II, III and IV studies. In this paper, we have outlined many of the issues that researchers must address when designing CER studies. We have highlighted several problems with studies that are the “kiss of death” in the minds of decision makers. Improved breadth and quality of retrospective databases, as well as advances that increase the transparency of methods that are necessary to address potential biases inherent in observational research, will do much to increase the acceptability of CER to those for whom it is designed.

ACKNOWLEDGMENTS

This study was funded by Genentech. We thank Teah Hoopes for manuscript preparation.

AUTHOR CONTRIBUTIONS

Conception/Design: Scott D. Ramsey, Sean D. Sullivan, Shelby D. Reed, Ya-Chen Tina Shih, Ken Schaecher, Rahul Dhanda, Debra Patt, Kelly Pendergrass, Mark Walker, Jennifer Malin, Lee Schwartzberg, Kurt Neumann, Elaine Yu, Arlene Ravelo, Art Small

Collection and/or assembly of data: Scott D. Ramsey, Sean D. Sullivan, Shelby D. Reed, Ya-Chen Tina Shih, Ken Schaecher, Rahul Dhanda, Debra Patt, Kelly Pendergrass, Mark Walker, Jennifer Malin, Lee Schwartzberg, Kurt Neumann, Elaine Yu, Arlene Ravelo, Art Small

Data analysis and interpretation: Scott D. Ramsey, Sean D. Sullivan, Shelby D. Reed, Ya-Chen Tina Shih, Ken Schaecher, Rahul Dhanda, Debra Patt, Kelly Pendergrass, Mark Walker, Jennifer Malin, Lee Schwartzberg, Kurt Neumann, Elaine Yu, Arlene Ravelo, Art Small

Manuscript writing: Scott D. Ramsey, Sean D. Sullivan, Shelby D. Reed, Ya-Chen Tina Shih, Ken Schaecher, Rahul Dhanda, Debra Patt, Kelly Pendergrass, Mark Walker, Jennifer Malin, Lee Schwartzberg, Kurt Neumann, Elaine Yu, Arlene Ravelo, Art Small

Final approval of manuscript: Scott D. Ramsey, Sean D. Sullivan, Shelby D. Reed, Ya-Chen Tina Shih, Ken Schaecher, Rahul Dhanda, Debra Patt, Kelly Pendergrass, Mark Walker, Jennifer Malin, Lee Schwartzberg, Kurt Neumann, Elaine Yu, Arlene Ravelo, Art Small

DISCLOSURES

Sean D. Sullivan: Genentech (RF); **Debra Patt:** The US Oncology Network (E); **Jennifer Malin:** WellPoint (E); **Kurt Neumann:** ION Solutions (E), Physician Resource Management, Inc. (C/A); **Elaine Yu:** Genentech (E), Roche (OI); **Arlene Ravelo:** Genentech (E), Roche (OI); **Art Small:** Genentech (E, OI); **Ya-Chen Tina Shih (H):** CER Expert Panel. The other authors indicated no financial relationships.

Editor-in-Chief: Bruce Chabner: Sanofi, Epizyme, PharmaMar, GlaxoSmithKline, Pharmacyclics, Pfizer, Ariad (C/A); Eli Lilly (H); Gilead, Epizyme, Celgene, Exelixis (O)

Reviewer “A”: None

C/A: Consulting/advisory relationship; RF: Research funding; E: Employment; H: Honoraria received; OI: Ownership interests; IP: Intellectual property rights/inventor/patent holder; SAB: scientific advisory board

REFERENCES

- Sox HC, Greenfield S. Comparative effectiveness research: A report from the Institute of Medicine. *Ann Int Med* 2009;151:203–205.
- Institute of Medicine. Initial national priorities for comparative effectiveness research. Available at <http://www.iom.edu/Reports/2009/ComparativeEffectivenessResearchPriorities.aspx>. Accessed November 2, 2011.
- U.S. Food and Drug Administration. Development and approval process (drugs). Available at <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/default.htm>. Accessed September 17, 2012.
- National Institutes of Health. Glossary. Office of Extramural Research. Available at <http://grants.nih.gov/grants/policy/hs/glossary.htm>. Accessed February 15, 2013.
- Ramsey SD, Howlader N, Etzioni RD et al. Chemotherapy use, outcomes, and costs for older persons with advanced non-small-cell lung cancer: Evidence from surveillance, epidemiology and end results—Medicare. *J Clin Oncol* 2004;22:4971–4978.
- Smith GL, Xu Y, Buchholz TA et al. Association between treatment with brachytherapy vs whole-breast irradiation and subsequent mastectomy, complications, and survival among older women with invasive breast cancer. *JAMA* 2012;307:1827–1837.
- Albain KS, Barlow WE, Shak S et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: A retrospective analysis of a randomised trial. *Lancet Oncol* 2010;11:55–65.
- Ramsey SD, Barlow WE, Gonzalez-Angulo AM et al. Integrating comparative effectiveness design elements and endpoints into a phase iii, randomized clinical trial (SWOG s1007) evaluating Oncotype

DX-guided management for women with breast cancer involving lymph nodes. *Contemp Clin Trials* 2013;34:1–9.

9. American Society of Clinical Oncology. CancerLinQ: Building a transformation in cancer care. Available at <http://www.asco.org/institute-quality/cancerlinq>. Accessed September 21, 2012.

10. Schnipper LE, Meropol N. ASCO addresses the rising cost of cancer care. *J Oncol Pract* 2009;5:214–215.

11. Meropol NJ, Schrag D, Smith TJ et al. American Society of Clinical Oncology guidance statement: The cost of cancer care. *J Clin Oncol* 2009;27:3868–3874.

12. Luce BR, Drummond MF, Dubois RW et al. Principles for planning and conducting comparative J Comp Eff Res 2012;1:431–440.

13. National Cancer Institute. Surveillance, Epidemiology, and End Results (SEER) publications. Available at <http://seer.cancer.gov/publications/>. Accessed September 17, 2012.

See www.TheOncologist.com for supplemental material available online.

CME

This article is available for continuing medical education credit at CME.TheOncologist.com.

EDITOR'S NOTE: See the accompanying commentary on pages 655–657.