
Approaches to Answering Critical CER Questions

Christine V. Kinnier, Jeanette W. Chung and Karl Y. Bilimoria

Abstract

While randomized controlled trials (RCTs) are the gold standard for research, many research questions cannot be ethically and practically answered using an RCT. Comparative effectiveness research (CER) techniques are often better suited than RCTs to address the effects of an intervention under routine care conditions, an outcome otherwise known as effectiveness. CER research techniques covered in this section include: effectiveness-oriented experimental studies such as pragmatic trials and cluster randomized trials, treatment response heterogeneity, observational and database studies including adjustment techniques such as sensitivity analysis and propensity score analysis, systematic reviews and meta-analysis, decision analysis, and cost effectiveness analysis. Each section describes the technique and covers the strengths and weaknesses of the approach.

Keywords

Comparative effectiveness research • Surgical oncology • Observational and database studies • Pragmatic trials

C.V. Kinnier

Department of Surgery, Massachusetts General Hospital, Boston, MA, USA

J.W. Chung

Department of Surgery, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

K.Y. Bilimoria (✉)

Surgical Outcomes and Quality Improvement Center, Department of Surgery, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
e-mail: k-bilimoria@northwestern.edu

Contents

1	Introduction	2
2	Randomized Controlled Trials: Limitations and Alternatives	3
3	Experimental Studies	3
	3.1 Pragmatic Trials	4
	3.2 Cluster Randomized Trials	4
	3.3 Adaptive Trials	5
4	Treatment-Response Heterogeneity	6
5	Observational and Database Studies	7
	5.1 Commonly Used Datasets	7
	5.2 Sensitivity Analysis	8
	5.3 Propensity Score Analysis	9
	5.4 Subgroup Analysis	9
	5.5 Instrumental Variable Analysis	10
6	Systematic Reviews and Meta-Analysis	11
	6.1 Systematic Reviews	11
	6.2 Meta-Analysis	12
7	Decision Analysis	12
8	Cost Effectiveness Analysis	13
9	Conclusion	13
	References	14

1 Introduction

Significant advances in evidence-based medicine have occurred over the past two decades, but segments of medical care are still practiced without underlying scientific evidence. Many practice patterns are so firmly established as the standard of care that a randomized controlled trial (RCT) would be unethical. Where evidence from RCTs exist, the study population is narrow and not easily applicable to most patients. Finally, multiple treatments are firmly engrained in clinical practice and have never been rigorously questioned.

Answering most of these knowledge gaps with an RCT, however, would be unethical, impractical, or dissimilar to routine care. In the last case, this is because an RCT is concerned with measuring efficacy: the effect of an intervention as compared to placebo when all other variables are held constant. In other words, an RCT creates a study environment in which outcome differences are most attributable to the intervention. While RCTs accurately identify treatment effects under ideal conditions, patients do not receive their routine care under those conditions. When it comes to routine patient care, physicians are less interested in efficacy than effectiveness: the effect of an intervention under routine care conditions. In order to study effectiveness, health care investigators have developed a toolbox of alternative techniques, known collectively as comparative effectiveness research (CER).

Both researchers and policy agencies have recognized the power of studying effectiveness. The 2009 government stimulus package allocated \$1.1 billion to CER [1]. The following year, the Affordable Care Act proposed multiple health care reforms to improve the value of the United States health care system. These reforms included the foundation of the Patient-Centered Outcomes Research Institute (PCORI), a publically funded, non-governmental institute charged with conducting and funding CER projects [2]. As a result, funding for CER has grown substantially in the past 5 years.

2 Randomized Controlled Trials: Limitations and Alternatives

RCTs are the gold standard of medical research because they measure treatment efficacy, but they remain dissimilar to routine care and are impractical under many circumstances. First, RCTs are often prohibitively expensive and time consuming. One study reported that Phase III, National Institutes of Health-funded RCTs cost an average of \$12 million per trial [3]. RCTs also take years to organize, run, and publish. Consequently, results may be outdated by publication. Second, some events or complications are so infrequent that enrolling a sufficiently large study population would be impractical. Third, clinical experts working from high-volume hospitals follow RCT participants closely in order to improve study follow-up and treatment adherence. Following study conclusion, however, routine patients receiving routine care may not achieve the same level of treatment adherence. Therefore study outcomes and routine care outcomes for the same intervention may differ substantially. Fourth, and perhaps most importantly, RCTs are restricted to a narrow patient population and a limited number of study interventions and outcomes. These necessary restrictions also restrict the broad applicability of RCT results.

Clearly investigators cannot rely solely on RCTs to address the unanswered questions in surgical oncology. CER techniques offer multiple alternatives. We will introduce these approaches here and then describe each technique in more detail throughout the series.

3 Experimental Studies

The term “clinical trials” evokes images of blinded, randomized patients receiving treatment from blinded professionals in a highly specialized setting. These RCTs are highly sensitive to the efficacy of the intervention under investigation. In other words, an intervention is most likely to demonstrate benefit in a setting where patients have few confounding medical diagnoses, every dose or interaction is monitored, and patients are followed closely over the study period. Unfortunately, RCTs are often prohibitively expensive and require many years to plan and complete. Furthermore, a medication that is efficacious during a highly-monitored RCT may prove ineffective during routine care where patients more frequently

self-discontinue treatment due to unpleasant side effects or inconvenience. Clinical trials performed with the comparative effectiveness mindset aim to address some of these limitations.

3.1 Pragmatic Trials

Due to the constraints of an RCT, results may not be valid outside the trial framework. Pragmatic trials attempt to address this limitation in external validity by testing an intervention under routine conditions in a heterogeneous patient population. These routine conditions may include a broad range of adjustments. First, pragmatic trials may have broad inclusion criteria; ideally trial patients are only excluded if they are not intervention candidates outside of the study. Second, the intervention may be compared to routine practice, and clinicians and patients may not be blinded. This approach accepts that placebo effect may augment intervention outcomes when used in routine practice. Third, pragmatic trials may use routine clinic staff rather than topic experts, and staff may be encouraged to adjust the medication or intervention as they would in routine practice. Fourth, patient-reported outcomes may be measured in addition to—or instead of—traditional outcomes. Finally, patients are usually analyzed according to their assigned intervention arm; this is also known as intention-to-treat analysis. Pragmatic trials may range anywhere along this spectrum: on one end, an otherwise traditional RCT may use an intention-to-treat analysis; on the other, investigators may aim to conduct the study under completely routine circumstances with the exception of intervention randomization. The investigators must determine what level of pragmatism is appropriate for their particular research question.

Pragmatic trials help determine medication or intervention effectiveness in a more realistic clinical setting. The adjustments that make pragmatic trials more realistic, however, also create limitations. Pragmatic studies are conducted under routine clinical circumstances, so an intervention that is effective in a large, well-funded private clinic may not be equally effective in a safety-net clinic. Therefore, clinicians must consider the study setting before instituting similar changes in their own practice. In addition, pragmatic trials include a broad range of eligible patients and consequently contain significant patient heterogeneity. This heterogeneity may dilute the treatment effect and necessitate large sample sizes and extended follow-up periods to achieve adequate statistical power. This may then inflate study cost and counterbalance any money saved by conducting the trial in a routine clinic with routine staff.

3.2 Cluster Randomized Trials

Cluster randomized trials (CRTs) are defined by the randomization of patients by group or practice site rather than by individual. Beyond group randomization, CRTs may use either traditional RCT techniques or pragmatic trial techniques. Group

level randomization has multiple effects. Group contamination is uncommon since participants are less likely to know study participants from other sites. This makes CRTs ideally suited for interventions that are organizational, policy-based, or provider-directed. With these education or resource-based interventions, well-intentioned participants or physicians may distribute information to control-arm participants without the investigator's knowledge. Risks of cross contamination are significantly lowered when participants from different study arms are separated by site and less likely to know one another. Group randomization also better simulates real-world practice since a single practice usually follows the same treatment protocol for most of its patients. Finally, physicians and clinic staff can be educated on the site-specific intervention and then care for patients under relatively routine circumstances. In some circumstances this may help to coordinate blinding, minimize paperwork, and reduce infrastructure and personnel demands.

Clustered patient randomization, however, introduces analytical barriers. Patients often choose clinics for a specific reason, so patient populations may differ more among clinics than within clinics. Furthermore, differences may not be easily measurable (e.g., patients may differ significantly in how much education they expect and receive prior to starting a new medication), so adjusting for these differences may be difficult during analysis. Consequently, cluster randomization requires hierarchical modeling to account for similarities within groups and differences between groups, but hierarchical modeling produces wider confidence intervals. As with pragmatic trials, this may require increases in subject number and follow-up time in order to detect clinically significant outcome differences. Unfortunately, individual participant enrollment is usually limited within each cluster, so increasing a trial's statistical power usually requires enrollment of additional clusters.

3.3 Adaptive Trials

Due to a history of unethical research, like the Tuskegee Syphilis Study, RCTs now undergo multiple interim analyses [4]. These routine evaluations check for interim results that may make trial continuation unethical, such as changes in routine care, early and robust outcome differences, or failure to see outcome differences where expected. Early termination prevents the inferior outcome group from suffering further harm.

Simply initiating an RCT, however, requires significant time, funding, and energy. Rather than terminating a trial, adaptive trials take advantage of interim analyses to adjust the trial conditions or outcomes and address further treatment questions. Changes to adaptive trials may include adjustments to eligibility criteria, randomization procedure, treatment dose or duration, or the number of interim analyses. They may also incorporate the addition of concomitant treatments or

secondary endpoints. To prevent the introduction of bias, both the adjustments and the circumstances under which they are introduced must be clearly delineated prior to trial initiation.

Unblinding the data for interim analysis may introduce bias, so data and resulting analyses must be sequestered from clinicians and patients still participating in the trial. Changing a study's outcomes may also cloak long term results. A medication that provides significant benefit after one year may have dangerous long term side effects that will remain unknown if a study is adapted. The investigator must therefore remember that any planned adaptations may affect important secondary outcomes.

4 Treatment-Response Heterogeneity

RCTs study the efficacy of a drug at a specific dose and frequency, but patients may vary widely in their ability to metabolize a medication or their response to a standard serum level. Furthermore, patient comorbidities may variably affect their susceptibility to drug side effects. These variations in effectiveness are collectively known as treatment-response heterogeneity.

There are three major techniques for addressing treatment-response heterogeneity. If the affected population and heterogeneity are already known, then a new trial may stratify patients according to the groups that require investigation and evaluate for outcome differences. If the affected population and heterogeneity are unknown, then data from a previous trial may be divided into subgroups and reanalyzed. This raises a number of analytical issues of which the investigator must be aware. The original trial may be insufficiently powered to detect differences in subgroups, especially when those subgroups are small segments of the larger study population. If the study is sufficiently powered, the investigator must analyze the subgroups appropriately. Evaluating the treatment effect within a subgroup is not sufficient to draw conclusions about treatment-response heterogeneity within that subgroup. Treatment effects within the subgroup must also be compared to remaining subgroups in order to determine if treatment-response heterogeneity truly exists within the subgroup of interest as compared to the general population. If subgroups are too small for formal subgroup analysis, the investigator may simply check for correlation between the concerning subgroup variable and the treatment. High correlation between the two suggests treatment-response heterogeneity.

Health care studies have recently employed a fourth technique known as finite mixture models [5]. These models allow certain covariate coefficients to vary for patient variables containing treatment-response heterogeneity. While new, this technique has been used with increasing frequency in health care cost modeling and is likely to find other applications in coming years.

5 Observational and Database Studies

In contrast to the experimental studies described above, observational and database studies examine patients and outcomes in the setting of routine care. Observational studies—which include cohort studies, case-control studies, and cross-sectional studies—and database studies help generate hypotheses when insufficient evidence exists to justify a randomized trial or when a randomized trial would be unethical.

5.1 Commonly Used Datasets

Databases are an excellent data repository for large numbers of patients. Data may be collected retrospectively or prospectively. While many institutions maintain their own databases to facilitate single institution studies, we will focus here on some commonly used and readily available national databases that are relevant to surgical oncology.

The Nationwide Inpatient Sample (NIS) [6] is the largest all-payer, inpatient database in the United States and is sponsored by the Agency for Healthcare Research and Quality. Data elements are collected retrospectively from administrative billing data and include primary and secondary diagnosis codes, procedure codes, total charges, primary payer, and length of stay. Data collection does not extend beyond discharge, however, and patient data cannot be linked across inpatient episodes. This means that even short-term variables like readmission and 30-day mortality cannot be measured. Limited information on patient demographics and hospital characteristics may be obtained by linking to other databases where permitted, but additional clinical details are unavailable. Consequently, patient-level risk adjustment and investigation of clinical complications or intermediate outcomes are largely impossible. Even with these limitations, the NIS is an all-payer, inpatient database. This makes the NIS a valuable resource for investigators interested in care and hospital cost differences associated with insurance coverage or changes in procedure use over time. Furthermore, NIS data is relatively inexpensive and available to any investigator that completes the online training.

The American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) [7] maintains a database of 30-day outcomes for all qualifying operations at ACS NSQIP hospitals. In contrast to the NIS, data elements in ACS NSQIP are collected prospectively by a trained nurse registrar. This significantly improves data quality by minimizing missing data elements and standardizing data entry. Data elements include patient demographics and comorbidities, operative and anesthesia details, preoperative laboratory values and 30-day postoperative morbidity and mortality. Hospital participation in NSQIP is voluntary, so unlike the NIS, ACS NSQIP data skews toward large, well-funded hospitals. Datasets from the ACS NSQIP are free but only available to ACS NSQIP participants.

The reach of ACS NSQIP is expanded significantly by linking to the National Cancer Data Base (NCDB) [8]. The NCDB, is a joint program of the Commission on Cancer (CoC) and the American Cancer Society. Like ACS NSQIP, the NCDB is a prospective database collected by trained registrars that provides high quality, reliable data. The NCDB collects data elements on about 70 % of cancer patients who receive treatment at a CoC-accredited cancer program. In addition to patient demographics and tumor characteristics, the database also contains data elements regarding chemotherapy, radiation, and tumor-specific surgical outcomes that are not available through the standard ACS NSQIP database. Perhaps most powerfully, registrars enter 5-year mortality data for all patients in the database. The NCDB is moderately priced but is publically available once an investigator has completed the mandatory training. Data linking to the ACS NSQIP database, however, is only available to investigators from ACS NSQIP hospitals.

The NCDB only collects data from cancer patients who receive care at CoC-accredited hospitals. Consequently, many minorities are underrepresented within the NCDB. In order to encourage high quality research on cancer epidemiology, the National Cancer Institute developed the Surveillance, Epidemiology, and End Results Program (SEER) [9]. SEER is a prospective database that over-represents minority cancer patients. For example, SEER samples all cancer patients in Connecticut but only collects data on Arizona cancer patients who are Native Americans. Like the NCDB, SEER data is also compiled by trained registrars thus making the data high quality and reliable. SEER elements include patient demographics, cancer stage, first treatment course, and survival. SEER is publically available following mandatory training but does have a modest data compilation fee. SEER may also be linked to cost and payer datapoints through the Medicare database for an additional price.

While research using each of these databases is associated with its own challenges, databases remain an excellent resource for large sample, retrospective research and offer a good starting point for many effectiveness research questions.

5.2 Sensitivity Analysis

Third-party, observational databases like those described above rarely contain the precise data elements with the precise coding desired by the investigator. Investigators may need to merge data elements to create a new study variable or define parameters to categorize a continuous data element. These decisions may unintentionally affect analytic results. Investigators can determine whether study results are robust using sensitivity analysis. An analysis is first run using parameters and specifications identified by the investigator, and the results are noted. Slight changes to variable definitions are then made and the analysis is rerun. Substantial changes in results suggest the results are largely dependent on investigator-selected parameters and specifications. Conversely, unchanged or similar results suggest the results are robust.

5.3 Propensity Score Analysis

Observational studies are primarily limited by the absence of randomization. In routine care, clinicians guide patients towards certain treatments based on the clinician's experience and the patient's primary condition and comorbidities. In order to control for these non-random choices in clinical care, propensity score analysis (PSA) attempts to match treatment subjects to control subjects with similar traits that may influence treatment choice and outcome. The investigator first identifies covariates that are likely to predict which patients received treatment, such as patient comorbidities or underlying disease severity. Using logistic regression, a propensity score is then calculated for each subject. Subjects who received treatment are matched to subjects with similar propensity scores who did not receive treatment, and outcomes for the two subject groups are then compared using multivariable analysis.

Matching methods vary widely amongst investigators. One-to-one subject matching, where each subject receiving treatment A is uniquely matched to a single subject receiving treatment B, is the most common, but multiple treatment B subjects may be used to improve statistical power. In addition, matching techniques may either optimize matching for the entire group and thereby minimize the total within-pair differences in propensity scores, or matching may be optimized for each patient as they are encountered in the dataset. In the latter case, the within-pair difference is minimized for the first subject but within-pair differences may increase as control subjects are progressively assigned to patients in the treatment group. In some cases, control subjects are replaced so that they can be used as the matched-pair for multiple treatment subjects, but this must be done uniformly throughout the matching process.

PSA is most limited by the covariates available in the dataset. For example, if tumor size is not recorded consistently then tumor size cannot be used to calculate the propensity score, and treatment differences will not be matched for tumor size. Conversely, including too many irrelevant covariates during propensity score calculation may minimize differences between groups and produce false negative results. The investigator must therefore optimize the number of covariates in the model so that patients are adequately matched but real group differences may be detected.

5.4 Subgroup Analysis

As mentioned above, PSA corrects for measured confounders in a dataset, but there is often reason to suspect that unmeasured confounders may be affecting the analysis. Subgroup analysis is one way to identify whether results have been biased by unmeasured covariates. There are multiple ways to use subgroup analysis to identify bias caused by unmeasured covariates; two are described here.

Observational studies most often compare treatment subjects to control subjects, but control subjects may include multiple subgroups. For example, if the treatment of interest is a medication, investigators typically assume that control subjects are patients who are not currently taking, and have never taken, the medication. But it is possible that control subjects used the medication in the past but stopped many years ago (remote users) or used the medication in the past but stopped recently (recent users). Control subjects may also be actively receiving a similar therapy using a different class of medications (alternative users). These subgroups can be leveraged to determine whether the study results are robust. In a typical observational study interested in complications from a medication, the investigator would compare the treatment subjects to all control subjects. With subgroup analysis, the investigator might compare treatment subjects to each control subgroup. Significantly different results in each subgroup analysis suggest that unmeasured variables have affected treatment choice and outcomes. In contrast, similar results in each subgroup analysis suggest that unmeasured confounders do not exist and the results are robust. The ability to use this technique largely depends on whether relevant subgroups exist and whether those subgroups are large enough to detect statistical differences.

5.5 Instrumental Variable Analysis

Subgroup analysis can detect unmeasured confounders, but in some cases, an analysis can actually correct for unmeasured confounders using instrumental variable analysis. Instrumental variable analysis essentially simulates randomization in observational data through the use of an instrument: an unbiased variable that differentially affects subject exposure or intervention but is uncorrelated with the measured outcome. Subjects are then grouped by treatment adjusted for the instrument, and outcomes are evaluated. If outcomes are similar regardless of instrument group, then treatment does not affect outcomes; if outcomes differ, then differences in outcomes between instrument groups can be attributed to natural bias in group assignment.

A well-known example of instrumental variable analysis is the Oregon Medicaid health experiment. In 2008, Oregon expanded its Medicaid program. Due to funding limitations, Oregon used a lottery to select the additional individuals who would receive Medicaid and those who would not. Investigators took advantage of this natural experiment and used differential access to Medicaid as an instrument. Investigators then evaluated the effects of Medicaid enrollment on health outcomes in low-income Oregon residents [10]. Sudden and random policy changes are rare events, but unbiased instruments can be identified through creative thinking. For example, distance to a hospital offering partial nephrectomy has been used as an instrumental variable when evaluating the effect of surgical technique on kidney cancer outcomes [11].

Given that instrumental variable analysis can mitigate the effects of unmeasured confounders, it can be a powerful tool in observational research. Unfortunately, use of the technique is highly dependent on having a reliable and unbiased instrumental variable.

6 Systematic Reviews and Meta-Analysis

Large RCTs that produce definitive results are uncommon due to significant funding and coordination barriers. Frequently, multiple small RCTs will address similar research questions but produce conflicting or non-significant results. In this case, the results of these RCTs can be aggregated to produce more definitive answers.

6.1 Systematic Reviews

Multiple RCTs may address the same research topic with results published in widely varying journals over many years. A systematic review disseminates these research results more concisely through the methodical and exhaustive evaluation of current literature on a specific research question. Investigators start with a research question, or collection of related questions, that are not easily answered by a single study or review. Medical databases are then searched using precise terms. High quality systematic reviews search multiple databases covering multiple disciplines and use redundant and related search terms. While the number of search terms must be brief enough to keep the number of results manageable, authors commonly use too few search terms and therefore miss critical articles on the topic of interest.

Once a list of articles is assembled, each article is carefully screened for eligibility and relevance based on predetermined inclusion and exclusion criteria. Results should summarize not only study results but also the quality of the study. In addition to a traditional results section, systematic reviews typically include a table of articles along with the study results and the level of evidence.

In the surgical community, the CHEST guidelines on venous thromboembolism prophylaxis are perhaps the most well-known systematic review [12]. In addition, the Cochrane collaboration is a not-for-profit, independent organization that produces and publishes systematic reviews on a wide range of healthcare topics, many relevant to surgery [13]. Unfortunately, systematic reviews are limited by the current data and they are unable to merge outcomes from multiple small RCTs. Systematic reviews also become outdated quickly since they do not create any new data. The quality of the conclusions is also highly dependent on the quality of the underlying database review. Despite these limitations, systematic reviews are an effective way to consolidate and disseminate information on a research topic and are often able to reveal trends in study results that are not visible when presented across multiple unique publications.

6.2 Meta-Analysis

Meta-analysis is a form of systematic review in which results from the relevant publications are combined and analyzed together in order to improve statistical power and draw new conclusions. To begin, investigators perform a systematic review, being careful to define the inclusion and exclusion so that study populations are similar. An effect size is calculated for each study and then an average of the effect sizes, weighted by sample size, is calculated. If multiple previous but small studies trended towards an outcome but were not statistically significant, a meta-analysis may produce clear and statistically significant results. If previous studies provided conflicting results, a meta-analysis may demonstrate a significant result in one direction or the other.

A meta-analysis, however, will not necessarily predict the results from a single, large RCT. This is because the results of the meta-analysis are largely dependent on the quality of studies. A few, high-quality RCTs will produce more robust results than multiple, small, low-quality RCTs. In addition, meta-analysis is highly subject to publication bias. Studies with strong, significant results are more likely to be published than studies with equivocal or negative outcomes. This can skew meta-analysis toward a false positive result.

7 Decision Analysis

The comparative effectiveness techniques discussed so far compare one or many treatment options to one another, but they are not equipped to address treatment consequences, including the need for additional evaluation and treatment. Decision analysis attempts to estimate the downstream costs and benefits of these treatment choices. First, a decision tree is developed that includes initial treatment options as well as potential outcomes and their impact on future evaluation and treatment needs. The root of the decision tree will center at either the disease or the sign or symptom requiring evaluation. The final branches of the decision tree should terminate with either treatment cessation or cure or with definitive diagnosis of the sign or symptom in question. The investigator then assigns probabilities to each of the branch points as well as costs and benefits associated with each step in the course of treatment or work up. Finally, the maximum expected utility for each possible pathway in the decision tree is calculated. Some have argued that decision analysis makes an intuitive process cumbersome and time consuming, but research repeatedly demonstrates that high quality decision analysis produces more effective decisions than intuition alone.

The investigator should optimize decision tree usefulness rather than maximize potential complexity. Keeping the target decision maker in mind helps to identify which decision points are necessary and which can be collapsed into existing branches. If the investigator is unsure whether a decision point is necessary, then

extreme cost and benefit values can be substituted for a questioned decision point. If the optimal outcome is unaffected, then the decision point is superfluous.

Decision analysis highly depends on assignment of accurate probabilities as well as costs and benefits. This is most challenging with non-monetary, analog outcomes such as quality of life or emotional distress. Without realistic estimates, however, the entire decision analysis will be inaccurate. The burden falls on the analyst to thoroughly research each treatment branch and estimate costs and benefits based on the best available data. If sufficient information is not available to develop an accurate decision analysis, then preliminary studies may be needed before decision analysis is attempted.

8 Cost Effectiveness Analysis

As more attention is focused on the disproportionate costs of the United States healthcare system compared to other countries, researchers have attempted to identify treatments that have the most favorable cost-to-benefit ratio. Investigators first focus on the costs and health benefits associated with each treatment option for a particular condition. Costs, at minimum, include the direct costs of the healthcare for which the provider is reimbursed. They may also include measures of indirect costs to the provider or the patient such as time or lost work productivity, intangible costs such as pain or suffering, and future costs. Future costs may relate to additional disease treatment or may be unrelated to the disease and treatment. Health benefits are most often measured as a difference in the quality-adjusted life years attributable to the intervention as compared to an alternative treatment. This measure accounts for both the expected life years remaining following treatment as well as the quality of life of those years. The quality of remaining years is usually discounted over time. For each treatment option, the ratio of total cost to health benefit is calculated and treatments are compared. The lowest ratio is identified as the most cost effective treatment, requiring the lowest cost for each quality-adjusted life year of benefit.

Cost effectiveness analysis is a valuable tool in policy development or insurance policies that are implemented at the population level. It may also provide some benefit for an individual who pays for healthcare out of pocket, but clinicians are unlikely to find cost effectiveness analysis helpful when counseling individual patients.

9 Conclusion

This chapter provided an overview of common CER techniques. Subsequent chapters will describe these techniques in further detail and provide examples of potential applications in surgical oncology. We hope that these examples will stimulate research ideas and encourage surgical oncologists to embrace CER

techniques and use them in their own research fields. Interest in effectiveness research will only increase as policy makers attempt to rein in exponential healthcare costs in the United States. If surgical oncologists begin using CER techniques now, they will be well prepared for the culture shift that is already mounting.

References

1. Pear R (2009) US to compare medical treatments. *The New York Times*, 16 Feb 2009
2. GovTrack HR (2009) Patient protection and affordable care act. In: 3590–111th congress. Accessed 25 Feb 2014. <http://www.govtrack.us/congress/bills/111/hr3590>
3. Johnston S, Rootenberg J, Katrak S, Smith W, Elkins J (2006) Effect of a US national institutes of health programme of clinical trials on public health and costs. *Lancet* 367 (9519):1319–1327
4. National Advisory Heart Council (1967) Organization, review and administration of cooperative studies (Greenberg Report)
5. Hunsberger S, Albert PS, London WB (2009) A finite mixture survival model to characterize risk groups of neuroblastoma. *Stat Med* 28(8):1301–1314
6. Overview of the Nationwide Inpatient Sample (NIS) (2014) <http://www.hcup-us.ahrq.gov/nisoverview.jsp>. Accessed 25 Feb 2014
7. Welcome to ACS NSQIP (2014) <http://site.acsnsqip.org/>. Accessed 25 Feb 2014
8. National Cancer Data Base (2014) <http://www.facs.org/cancer/ncdb/>. Accessed 25 Feb 2014
9. Data and Software for Researchers (2014) Surveillance, epidemiology, and end results program <http://seer.cancer.gov/resources/>. Accessed 25 Feb 2014
10. Taubman SL, Allen HL, Wright BJ, Baicker K, Finkelstein AN (2014) Medicaid increases emergency-department use: evidence from Oregon’s Health Insurance Experiment. *Science* (New York, NY) 343(6168):263–268
11. Tan HJ, Norton EC, Ye Z, Hafez KS, Gore JL, Miller DC (2012) Long-term survival following partial versus radical nephrectomy among older patients with early-stage kidney cancer. *JAMA J Am Med Assoc* 307(15):1629–1635
12. Guyatt GH, Akl EA, Crowther M, Gutterman DD, Schuünemann HJ (2012) Executive summary: antithrombotic therapy and prevention of thrombosis, 9th ed: american college of chest physicians evidence-based clinical practice guidelines. *CHEST J* 141(2_suppl):7S–47S
13. The Cochrane Collaboration (2014) <http://us.cochrane.org/cochrane-collaboration>. Accessed 25 Feb 2014