



# Comparative Effectiveness Research

# 2

Nader N. Massarweh and George J. Chang

## Abstract

Comparative effectiveness research (CER) involves human subjects or the use of secondary data to compare the effectiveness of one preventive, diagnostic, therapeutic, or care delivery modality to another using real-world approaches and under real-world conditions. CER utilizes a variety of data sources and methods to generate pragmatic data that can be used to improve health outcomes and value for health care systems by developing and disseminating evidence-based information for patients, clinicians, and other decision-makers. The intent of this type of research is to help stakeholders better delineate which type of treatments are best for which specific patients and under what conditions.

## 2.1 Introduction

New research findings are used to drive practice, to better understand the role for novel technologies and therapeutics in the care of patients, and to provide health care practitioners with information at the point of care regarding the management of

---

N. N. Massarweh (✉)

Center for Innovations in Quality, Effectiveness and Safety, Michael E DeBakey VA Medical Center, Houston, TX, USA

Michael E DeBakey Department of Surgery, Baylor College of Medicine, Houston, TX, USA  
e-mail: [massarwe@bcm.edu](mailto:massarwe@bcm.edu)

G. J. Chang

Department of Surgical Oncology, University of Texas, MD Anderson Cancer Center, Houston, TX, USA

Department of Health Services Research, University of Texas, MD Anderson Cancer Center, Houston, TX, USA

© Springer Nature Switzerland AG 2020

J. B. Dimick, C. C. Lubitz (eds.), *Health Services Research*, Success in Academic Surgery, [https://doi.org/10.1007/978-3-030-28357-5\\_2](https://doi.org/10.1007/978-3-030-28357-5_2)

patients with rare or uncommon conditions. While the US healthcare system supports a robust research enterprise, an important shortcoming of many contemporaneously published studies is that they don't address a fundamental question relevant to patients and providers—what is the best treatment for this specific patient in this specific clinical context? Given an ever-increasing emphasis on shared decision-making and value in US healthcare, the importance and timeliness of Comparative Effectiveness Research (CER), which is intended to address this specific question, cannot be understated.

Fundamentally, clinical research is intended to compare the safety, benefits, and/or limitations of two or more treatments. In this regard, one might assume that any study in which two interventions are compared is CER. However, CER is specifically intended to provide data comparing the effectiveness of two interventions when applied under *real-world conditions*. Furthermore, CER encompasses research derived from different data sources, and utilizes a variety of study designs and analytic methodologies. These varied sources are assimilated to ascertain which specific intervention(s) will work best for which specific patient (or population of patients) while at the same time balancing the relative benefits and harms in order to inform patient care or health policy decisions. This is one of the main reasons for the proliferation of CER over the past decade, why new opportunities for funding CER have emerged, and why there is an ongoing need for CER to inform contemporary health care reform efforts and the transition from volume-based to value-based care models.

---

## 2.2 Efficacy vs Effectiveness

### 2.2.1 Efficacy

Randomized clinical trials (RCTs) remain the benchmark for data used to either change clinical practice or to drive evidence-based care. However, most RCTs typically address a very specific question when comparing interventions—what is the efficacy of one intervention over another? Simply defined, *efficacy* is a measure of the effect of a particular intervention under idealized circumstances. In addition to the cost and time needed for study initiation and completion (which are acknowledged limitations of RCTs), the external validity (i.e.: generalizability) of findings from RCTs frequently create important challenges to the integration of data into practice. More specifically, because the patient populations included in RCTs are typically strictly defined with numerous inclusion and exclusion criteria (which may not reflect the population of patients providers see in actual practice) and because of the intensity of the care enrolled patients receive (which may not reflect the level of care or the type of practice where a patient is being treated), it is not clear that the findings from a given RCT will directly translate into real-world conditions.

### 2.2.2 Effectiveness

By comparison, *effectiveness* is defined as the effect of an intervention under real-world conditions and includes an evaluation of not only the benefits, but also the

harms. CER is important because what may demonstrate efficacy in the strictly controlled context of a clinical trial may not yield the same outcomes in everyday practice. In many respects, this type of data is much closer to what health care providers and patients need at the point of care when choosing between two different interventions. CER studies attempt to make comparisons between two or more treatment strategies within populations reflective of the types of patients a provider might see in his or her practice and, as importantly, to ensure the conditions under which the comparison is made reflect the varied practice environments in which care is delivered in the general community.

---

### 2.3 The Evolution of CER

Distilled to its primary goals, CER compares two or more treatment modalities or approaches to the care of patients or populations. Thus, CER is research comparing the effectiveness of two or more preventive, diagnostic, therapeutic, or care delivery strategies using real-world approaches and under real-world conditions. The interventions compared in CER studies can be anything used in the care of patients or populations of patients including, health care interventions, treatment protocols, care delivery models, invasive procedures, medical devices, diagnostic tools, pharmaceuticals therapeutics, and any other strategies used for treatment, diagnosis, or prevention of illness or injury.

While the principles underlying CER have been around for a number of years, it is the recent emphasis on value in US healthcare and the transition from volume-based to value-based care that has brought the attention and support of policy makers for this type of research. Over the past decade, two important pieces of legislation have contributed to the growth of CER. The American Recovery and Reinvestment Act of 2009 allocated \$1.1 billion to the Department of Health and Human Services, the National Institutes of Health, and the Agency for Healthcare Research and Quality stipulating that this funding should be used for the dual purpose of supporting research intended to compare the outcomes, effectiveness, and appropriateness of interventions for the prevention, diagnosis, or treatment of patients *AND* to encourage the development and use of more robust clinical data sources. This legislation also established the Federal Coordinating Council for Comparative Effectiveness Research whose charge was “to foster optimum coordination of CER conducted or supported by Federal departments and agencies”.

The second piece of legislation was The Patient Protection and Affordable Care Act passed by Congress and signed into law by President Obama in 2010 which established and funded the Patient-Centered Outcomes Research Institute (PCORI). Prior to the establishment of PCORI, there had been numerous efforts in both the private and public sectors to conduct CER studies and to generate comparative effectiveness data, but these efforts were limited by the lack of a unified definition for CER, variable funding priorities, and they lacked a robust means of tracking the types of studies being performed and on which topics. To fill these gaps, PCORI was created to become the primary funding agency for investigators performing CER. Since its inception, PCORI has:

- Provided \$2.3 billion to help fund a broad portfolio of CER studies, develop research infrastructure, and disseminate and promote findings into actual practice.

- Established a policy for funded researchers to share their data, documentation, and statistical programming to encourage data sharing through open science.
- Developed methodologic standards (through the Methodology Committee) for performing CER and patient-centered outcomes research.
- Created a national data platform to support and improve the efficiency of conducting CER (i.e.: PCORnet).

PCORI has established National Priorities for Research in the following domains:

- Comparing the effectiveness and safety of alternative prevention, diagnosis, and treatment options to see which one works best for different people with a particular problem.
- Comparing health system-level approaches to improving access, supporting patient self-care, innovative use of health information technology, coordinating care for complex conditions, and deploying workforce effectively.
- Comparing approaches to providing comparative effectiveness research information, empowering people to ask for and use the information, and supporting shared decision-making between patients and their providers.
- Identifying potential differences in prevention, diagnosis, or treatment effectiveness, or preferred clinical outcomes across patient populations and the healthcare required to achieve best outcomes in each population.
- Improving the nation's capacity to conduct patient-centered outcomes research by building data infrastructure, improving analytic methods, and training researchers, patients, and other stakeholders to participate in this research.

A major criticism of contemporary clinical research is that the findings from very few studies actually fill a practical knowledge gap that can impact everyday clinical practice. Because a principal goal of CER is to improve individuals' ability to make informed healthcare decisions through the generation of data that can help patients, providers, and policy makers understand what treatment(s) will work best and for whom, a unique aspect of PCORI is the engagement of stakeholders such as patients, providers, and other decision-makers throughout the CER process. By involving stakeholders in the research process, the hope is that the most relevant questions and priorities can be identified, knowledge gaps can be better addressed, and approaches for dissemination and implementation of study findings can be optimized. It is this engagement that has led CER to be referred to at times as 'patient-centered outcomes research' and is believed to be a previously under-appreciated avenue for enhancing dissemination of data and translation into practice.

---

## 2.4 Conducting CER

CER is intended to impact the care of either individual patients or patient populations and can be conducted from various stakeholder perspectives. It can also affect health policy decisions as well as how or why care is delivered, organized, and paid for by health care entities. As such, a key component of CER is the external validity

of the data or the ability to generalize the results to patients and clinical settings outside of the study population. Given the breadth of topics that can be addressed by CER, a variety of study designs and analytic methods are employed. However, prior to initiating a CER study, an understanding of the limitations of a given research question and specific study design are equally critical to the successful execution of a CER study with internal validity. In this regard, several important questions must be addressed during the study conception and design phase to ensure the right data source is selected, an appropriate study design is chosen, and appropriate statistical methods are employed.

- *Is the intent of the study to compare the effect of an intervention at the individual patient-level or at the population-level?*
  - Certain data allow for the analysis of patients clustered within hospitals, health systems, or geographic regions while others do not.
- *Is the research question appropriate for CER methods?*
  - The available data (or data that can be readily generated) must be able to answer the research question through the application of appropriate statistical methods.
- *Is the data source appropriate to address the chosen research question?*
  - Observational data sources used for CER often have important, unique, and inherent limitations that can create relevant sources of bias that must be considered and addressed either through the study design, the selection of the study population, and/or the methodology employed. In addition, for studies that truly seek to address a CER question, the data source should support the external validity of the findings.
- *Will the chosen study design and/or analytic methods minimize bias and enhance the internal validity of the findings?*
  - Investigators must have a working knowledge of available statistical tools and analytic approaches and understand the extent to which conclusions may (or may not) be supported by the data.

The EQUATOR (Enhancing the QUALity and Transparency Of health Research) network is an organization that has developed a standardized set of reporting guidelines for many of the typical types of CER studies like RCTs and cohort studies. These guidelines were developed in a collaborative and multi-disciplinary fashion with input from funding agencies, journal editors, peer reviewers, and researchers with the primary goal being to elevate and improve the overall quality of contemporary published research. These guidelines can be helpful to ensure the rigor and clarity of presentation for CER studies.

---

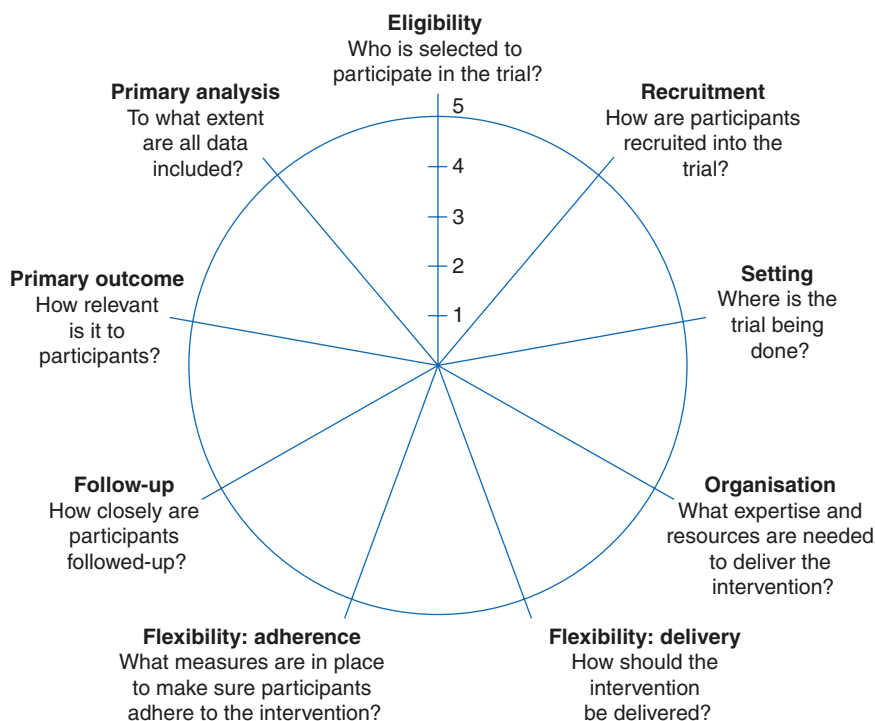
## 2.5 Types of CER Study Designs

There are four principal, broad categories of study design used to conduct CER, each with their own advantages and limitations. These can be applied to generate new data to fill knowledge gaps in current clinical practice or to evaluate the

existing evidence regarding benefits and harms of specific treatments when applied to different patients or populations. CER studies can either be prospective or retrospective and can be based on primary data collection or secondary analysis of existing data.

### 2.5.1 Randomized Clinical Trial

Data derived from RCTs remain the benchmark against which all other sources of data driving changes in clinical practice are compared. RCTs can span a spectrum from **explanatory** to **pragmatic** trials (Fig. 2.1). The majority of trials conducted and published are explanatory in nature and designed to address the issue of efficacy. As such, most explanatory trials will have study protocols with stringent inclusion and exclusion criteria. Not only are enrolled patients frequently extremely healthy, which may not reflect the real-world population of patients with a given condition, but the trial protocols also generally involve rigorous patient follow-up and monitoring, which also may not be indicative of typical day-to-day practice for



**Fig. 2.1** Tool for determining where a given RCT protocol falls in the explanatory to pragmatic continuum. Each of 9 domains for a given trail are scored from 1 to 5 (1 = very explanatory and 5 = very pragmatic) and then used to gauge where on that continuum it falls (taken from Loudon K, et al. BMJ. 2015)

providers in most practice settings. These are important drawbacks that can have important ramifications for the external validity of these types of studies.

By comparison, and as the name would suggest, pragmatic trials are intended to define the effectiveness of a given intervention and are more in line with the goals of CER. Whereas a strict study protocol and numerous inclusion and exclusion criteria are important for evaluating efficacy in an explanatory trial (as these features help to minimize any possible impact of confounding on study findings) this creates a critical blind spot for patients and practitioners—namely, how will this therapy work in routine clinical practice? In line with the goals of CER, pragmatic trials are intended to compare the effectiveness of varying treatments or management strategies with findings that can readily be generalized to most patients being treated in most clinical contexts or settings. To this end, the inclusion and exclusion criteria for such trials are typically more inclusive with study protocols that may even be flexible. In addition, study outcomes frequently only represent the most pertinent information required to address the research question and/or are the most easily assessed or adjudicated. A limitation of these studies is that the more parsimonious approach to data collection can limit the ability to conduct subgroup analyses or to perform post-hoc secondary data analyses addressing related questions.

**Cluster RCTs** are an example of a pragmatic trial design. In cluster trials, randomization is not performed at the individual level, but rather as a group, a practice, a clinic, or any other specified population. Within each cluster, patients receive usual care in addition to the experimental intervention and may not be aware they are participating in an RCT. This approach can markedly improve the external validity of study findings. However, a drawback to cluster trials is that because the unit of analysis is the cluster rather than the individual patient, the required sample size to ensure adequate statistical power may be larger and statistical methods, such as hierarchical models, must be used to address the within cluster correlation of the data (i.e.: patients treated within a given cluster are likely receiving similar care and thus are likely to have similar outcomes).

**Stepped-wedge RCTs** are another unique type of trial that can be considered a subtype of the cluster design. Whereas in cluster RCTs, each cluster is assigned to either the control or the intervention, in a stepped-wedge design all clusters initially start not being exposed to the intervention and will eventually receive the study intervention by the end of the trial, but the timing with which the intervention is administered is random. One of the benefits of these types of trials is that all study participants will receive the intervention. So, in cases where the intervention seems likely to be beneficial, this could enhance willingness for trial participation. Another benefit is the efficiency of the this design because the nature of the randomization process allows for each cluster to act as its own control. This also provides data that allows for both between and within cluster comparisons.

**Adaptive RCTs** are designed to allow changes to the protocol or the statistical analysis after trial initiation. Such changes are based on Bayesian analytic approaches as compared to the frequentist approaches typically employed in more traditional RCTs. This provides adaptive RCTs with a number of advantages. For example, protocol and/or procedural changes have already been approved as part of

the trial design and, as such, can be implemented more efficiently. Total accrual and even study outcomes can change during the conduct of the trial as data accumulate. In this regard, adaptive RCTs can actually allow for more rapid study completion. However, by their nature adaptive RCT designs are more complex and as the trial protocol changes Bayesian analytic approaches become compulsory. As such, investigators should be well-versed in Bayesian statistics and ensure they have biostatistical support to ensure the integrity of trial results.

### 2.5.2 Observational Studies

Observational studies constitute the majority of contemporary HSR and outcomes research. The availability of numerous data sources, the efficiency with which data can be obtained and analyzed, and the relatively low associated costs for conducting this type of research are all reasons why these also represent a very common form of CER. In comparison to the rigorous protocols often used in controlled trials, an important feature of observational studies, in particular those based on the secondary use of local, regional, or national data sources (e.g.: administrative claims, registry data, or electronic health record data), is that they frequently reflect the actual management patients received during a given episode of care. Whereas, the emphasis in RCTs is frequently on internal validity sometimes at the expense of external validity, observational studies often implicitly emphasize external validity at the expense of internal validity. Specifically, although the data may reflect the type of care patients actually receive in real-world clinical practice settings and contexts, because of the non-controlled nature of observational studies, numerous sources of bias and confounding must be considered and either addressed through the study design, the selection of the study population, or through the application of various analytic and statistical approaches. Issues such as selection bias, confounding by indication, and missing data are all potential barriers to the internal validity of the findings from observational CER studies that must be considered and addressed. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for reporting of observational studies can provide investigators with a useful list of considerations when preparing a manuscript of their study results.

### 2.5.3 Research Synthesis

In certain situations, there may be a body of literature regarding a given intervention or treatment approach, but the data from individual studies could either be discordant or the sample size may not be large enough to clearly and definitively support a specific conclusion. In this context, **systematic reviews**, **meta-analyses**, and **health technology assessments** are valuable tools that can be used to synthesize the existing data. The goals of each of these three types of studies are distinct. In a systematic review, the goal is to provide an unbiased, comprehensive, clear summary



of the body of data on a given topic. In meta-analysis, the goal is to combine the results of available studies on a given topic through quantitative techniques in order to create a collective data set that is better powered than its component data sources. In a health technology assessment, the goal is to generate data (specifically regarding a health technology) that can be used to inform clinical and policy-level decision making directed at the introduction and diffusion into practice of a given innovation. In all three types of research syntheses, the data used can be based on RCTs or on observational studies.

#### 2.5.4 Decision Analysis

Decision analyses are informed by two types of data. The first is the probability of an outcome given a particular treatment or management pathway. The second is the patient's current and future health status, which inherently considers both the benefits and harms attributable to that treatment or pathway. These two components are applied and used to perform model-based quantitative evaluations of the outcomes associated with specific management strategies in specific situations. These are central study designs for CER because the underlying goal is to help patients and providers derive the best treatment decision for a specific patient in a specified clinical context or from a specific health-care perspective. Cost-effectiveness analyses also integrate aspects of decision analytic techniques to incorporate cost and quality of life inputs to assess the comparative value attributable to a given intervention or treatment approach. Through simulation modeling using best available data and assessing which parameters impact the outcomes most, future areas of needed research (i.e. RCTs or other prospective designs) can be prioritized.

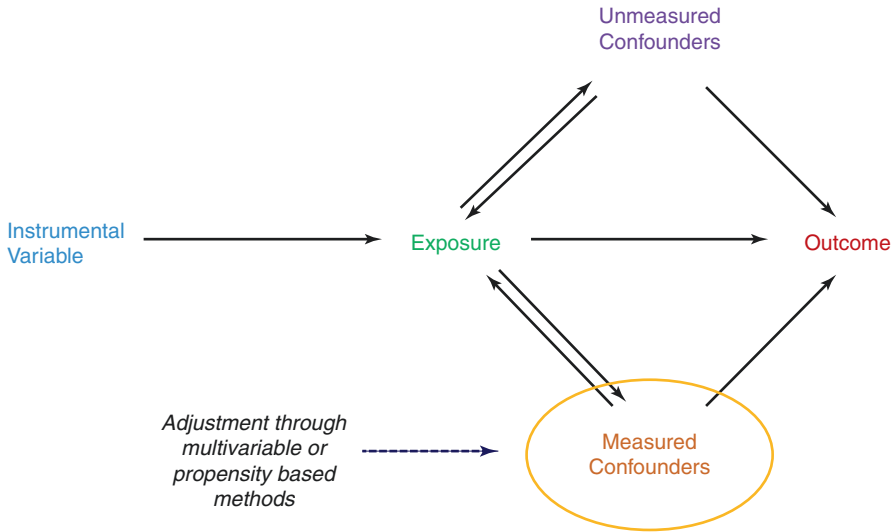
---

## 2.6 Commonly Used Statistical Methodology

For observational CER, the appropriate selection and use of statistical methodology is critical for ensuring the internal validity of the study and for addressing sources of bias and/or confounding. While different statistical approaches might be appropriate for a given study, often times the 'correct' choice is predicated on the data source, the nature of the research question, and the experience or expertise of the investigative team. Additionally, using a combination of these statistical approaches can be helpful to evaluate the robustness of study findings in the context of varying assumptions about the data. Similarly, carefully planned subgroup and sensitivity analyses can also help to bolster the robustness of study results to varying assumptions.

### 2.6.1 Methods to Address Confounding

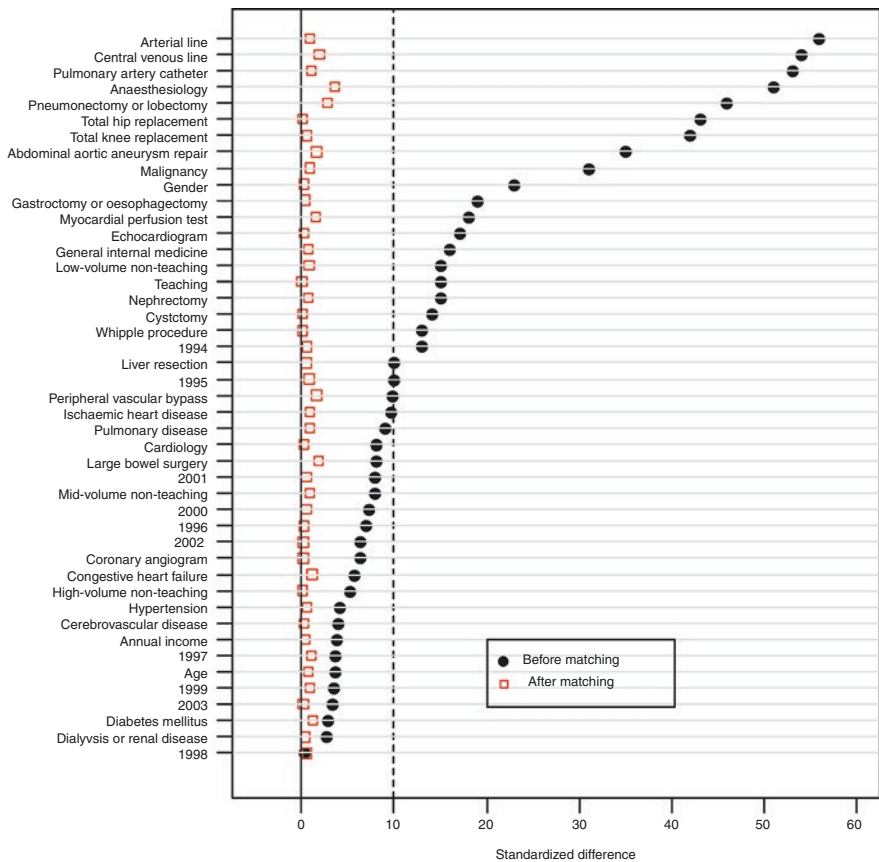
One of the most common approaches for addressing confounding (Fig. 2.2) is the use of a **multivariable model**. Models are used to estimate the effect of a given



**Fig. 2.2** Conceptual diagram of the association between exposure, outcome, confounders and potential instrumental variable in observational studies. Multivariable modeling and propensity score based adjustment can adjust for measured confounder, but neither approach is able to adjust for the effect of unmeasured confounders. Instrumental variables adjust for the effect of both measured and unmeasured confounders because they are only related to the outcome indirectly through the exposure and thus are not subject to confounders that may affect the association between exposure and outcome

exposure (e.g.: treatment) on a specified outcome while adjusting this estimate for the effect of factors that can potentially confound (i.e.: obscure) this relationship. The type of model used in a given study depends largely on the nature of the outcome of interest. For continuous outcomes (e.g.: post-operative length of stay), linear regression is most commonly applied. For binary outcomes (e.g.: perioperative mortality), logistic regression is frequently used. For time-to-event outcomes (e.g.: time from diagnosis to death), Cox proportional hazard regression is used. The benefits of multivariable models are that they are efficient and familiar for most investigators. In addition, there are hierarchical versions of these models that can be used to evaluate correlated data (e.g.: clustering of patients within a provider or hospital), to explore between and within cluster variation, and to address potentially less reliable estimates due to small sample size. In some instances, model performance can be improved by the inclusion of one or more interaction terms between covariates. An interaction occurs when the effect of one variable on the outcome of interest is predicated on the value of a second variable—also known as effect modification. The value of including interaction terms can be assessed by evaluating the model fit both with and without the inclusion of the interaction. An important limitation and consideration when using a model is that the completeness of adjustment is entirely predicated on the availability of data regarding measured confounders as the model cannot adjust for factors which are not measured or observed in the dataset.

The use of **propensity score analyses** has increased dramatically in recent years. In this type of analysis, the estimated probability (i.e.: propensity) for a patient to receive a given exposure relative to another is calculated. Propensity scores are most frequently used to estimate this probability when comparing the effect of two treatments on a given outcome. The score is derived by using a multivariable logistic regression model with the binary outcome of interest being the two possible treatments of interest. Other available factors that can potentially be associated with the receipt of a given treatment, the outcome, or are believed to be confounders of the relationship between the exposure and outcome are included as covariates in the propensity model. This estimated probability of treatment assignment can then be used in several ways to address potential confounding when comparing the effect of the two treatments on the outcome(s) of interest. For example, the propensity scores can be included as a covariate in the model estimating the association between the exposure and outcome which can be an efficient way to address issues related to statistical power for infrequently occurring outcomes. A popular approach is to perform propensity matching (Fig. 2.3). In this case, the



**Fig. 2.3** Graphical representation of standardized differences before and after propensity matching of covariates (taken from Gayat E, et al. Intensive Care Med. 2010)

propensity score is used to identify patients with an identical or very similar propensity for having received a given treatment. After matching, observed covariates are often well-balanced and can appear to simulate what might be observed in the context of randomization. However, there are important limitations of propensity methods. A major limitation is that propensity score methods can only account for measured factors, while there may be a number of unmeasured factors that are important to treatment assignment (a.k.a. hidden bias) and can influence outcomes. While matching on propensity scores can result in a forced balance between observed covariates, it does not address confounding related to unmeasured factors and may actually exacerbate imbalance in such factors. In addition, matching can significantly reduce sample size and statistical power. As such, matching is often best applied in large datasets with numerous potential covariates where sample size is less of a consideration. Finally, it is unclear that propensity scores achieve significantly different estimates as compared to multivariable modeling.

In **instrumental variable (IV) analyses**, a specific variable is chosen to serve as the “instrument” for comparing two interventions. An instrument is chosen such that it is the external cause of the intervention or exposure of interest, but by itself is unrelated to the outcome except through the causal pathway (Fig. 2.2). Randomization of patients in RCTs is an example of an instrument—the treatment a patient receives is entirely predicated on the randomization, but randomization has no effect on the outcome except through the treatment the patient receives. As an example, a CER study might seek to utilize an existing data source to compare the effect of a minimally invasive surgical approach relative to an open approach on a given outcome. However, because a variety of factors play a role in a clinician’s decision regarding whether or not to recommend a minimally invasive approach (e.g.: prior surgery in the chest or abdomen; the patient’s body habitus; other anatomic considerations at the surgical site; concurrent co-morbidities), a simple comparison of patients treated with these two approaches could be biased because of confounding by indication (based on unmeasured factors). An IV for this comparison might be hospital-level or regional rates of minimally invasive surgery. When patients are categorized into groups based on the value of the instrument, the rates of treatment will differ, the probability of treatment is no longer affected by potential confounding characteristics of an individual patient, and the comparison of interest becomes analogous to comparing randomized groups. Relative to multivariable models and propensity scores, an important benefit of IV analyses is that they not only address imbalance in measured confounders, but they are also believed to address imbalance in unmeasured variables as well. In this respect, estimates from IV analyses are believed to be better for addressing residual confounding (i.e.: confounding from unmeasured or unadjusted factors) and more accurately reflect the true association between a given exposure and outcome in observational studies.

## 2.6.2 Addressing Sources of Bias

While confounding is related to the effect of one or more measured or unmeasured factors that can obfuscate the association between an intervention and an outcome

of interest, bias is a form of error within the design or analysis of a study that can also distort the estimate of the exposure-outcome relationship. Whereas confounders are typically addressed through model-based adjustment, bias is more effectively dealt with through either study design, selection of the study population, or the use of specific statistical approaches.

**Missing data** is frequently an issue when using observational data sources. There are two main consequences of not adequately addressing missing values. The first, is that the sample size (and thus the power) of a study can be significantly decreased if case-complete (i.e.: analysis of only patients with non-missing data) approaches are selected. Methods such as imputation can be useful to address this issue. However, prior to doing so, it is important to consider the second issue which is the introduction of bias. It is important to consider which variables have missing data and why, how patients with missing data differ from those without missing data, and whether missing values can be predicted based on observed data. Multiple imputation methods are frequently used to address missing values and are believed to provide better powered, unbiased estimates in cases when data are missing completely at random or missing at random. In cases where data are missing not at random (i.e.: missing values are related to unmeasured, non-random, patient-level factors), any methods of addressing missing data will likely result in the introduction of bias.

**Selection bias** occurs when allocation of study subjects to a given intervention does not accurately reflect what happens in actual practice. For example, an observational study might demonstrate a clear benefit associated with the use of adjuvant chemotherapy in patients with colon cancer. However, an important factor that could introduce selection bias into this analysis is whether the data source provides information on postoperative complications (like surgical site infections which are common after colorectal surgery). If patients were simply categorized based on whether or not they received adjuvant therapy without accounting for patients who may not have received adjuvant therapy because they had a postoperative complication, the observed benefit in patients who received adjuvant therapy could be explained by the better postoperative outcome that would be expected when a complication does not occur rather than any effect attributable to the adjuvant therapy itself. Careful selection of the data source and the patients included in this type of study as well as well-selected sensitivity analyses are useful approaches to mitigate, to the extent possible, the effect of selection bias.

**Survivor treatment bias** is a particularly important consideration in oncologic studies evaluating the survival benefit of adjuvant interventions occurring after surgery. In order for a patient to receive a treatment after an operation, they must survive through the post-operative period. Put simply, patients who live longer after an operation have more of an opportunity to receive additional treatment. A *landmark analysis* can be a useful approach to address this issue. In landmark analyses, survival is estimated for groups of patients conditional on the fact that they have survived to at least a specified time point—the landmark (e.g.: all patients in the analysis survived at least 90-days beyond surgery).

**Lead-time** and **length-time bias** are both relevant to studies evaluating screening interventions. *Lead-time bias* occurs when the survival benefit associated with a given intervention is due entirely to the earlier detection of a disease (as opposed to the patient presenting after it has become clinically apparent) rather than any actual effect of the intervention itself. Put differently, it is the time interval between when a disease is detected by an intervention relative to when it would be typically diagnosed. This type of bias can make cancer screening interventions appear to make patients live longer. By comparison, *length-time bias* occurs when slow progressing cases are detected more often and thus patients live longer.

---

## 2.7 Limitations of CER Studies

Conducting research studies that can provide meaningful, generalizable data while at the same time ensuring internal validity by anticipating relevant sources of bias and confounding can be a real challenge even for experienced investigators. This is one reason why RCTs remain the benchmark against which all other types of studies attempting to inform evidence-based practice are judged. However, RCTs can be prohibitively costly and time consuming. When done properly, the estimated effect sizes obtained from well-performed observational CER studies can be quite similar to those obtained from RCTs. There are considerations that both consumers of the peer reviewed literature and investigators should keep in mind when interpreting the findings from a CER study.

It is important to ask if the direction and size of the observed association is believable and consistent with what may already be known. In observational studies, there may be a tendency to believe that estimates derived from the use of advanced statistical techniques by themselves address all sources of bias and thus provide valid estimates of the association—this is simply not the case. Assuming a given statistical approach has been applied correctly and the analysis is sound, frequently the data source used to conduct an observational study may have specific nuances or limitations that are not fully considered during the conduct of the study and can result in biased estimates. In cases where the size of the observed effect is too large relative to what is known from existing RCT data, the results should be viewed with a circumspect eye and consideration should be given to the manner in which relevant sources of bias may have affected the findings. On the other hand, if the findings are corroborated across a variety of data sources, patient populations, and/or using different statistical methodologies, this can lend credence to the study findings.

In observational studies, an association should not be immediately interpreted as causality. However, there are established criteria that can support the conclusion of causal inference for an observed association. Of the nine historically described criteria, the following six are the most relevant to observational CER studies: strength; consistency; specificity; temporality; presence of a biological gradient; plausibility. The more of these criteria that are present for the findings of an observational study, the stronger the case that a true association exists.

## **2.8 Barriers to the Conduct and Implementation of Findings from CER Studies**

Although the mission and value of CER are well-established, important and existing barriers within the US healthcare system remain to the generation and implementation of new data. A wealth of data clearly demonstrates that there are national disparities in care for certain populations, that there is ongoing variation in the quality and costs of care, and that health care in the US costs more and is of lower quality than in other comparable industrialized nations. Despite general agreement on the reasons that change is needed and, perhaps to a lesser extent, the manner in which this change should occur, there are numerous legislative obstacles to implementation of findings from CER studies that could inform the transition toward more value-based health care models.

Research supported by PCORI is intended to improve care quality, increase transparency, and increase access to better health care for all patients. However, PCORI is explicitly prohibited from funding research studies that evaluate or apply cost-effectiveness thresholds or utilize quality adjusted life years. There is also specific language in the act that the reports and research findings may not be construed as practice guidelines or policy recommendations and that the Secretary of Health and Human Services may not use the findings to deny coverage. These stipulations are likely based on societal and/or political fears that research findings could lead to the potential for health care rationing.

Current spending on prescription drugs is estimated at between \$400–550 billion. Although many new and novel therapeutic drugs are brought to market, many provide only added costs with minimal clinical benefit. While CER studies (in particular cost-effectiveness analyses) of various types of drugs would be of great value, at present there are statutory limitations in how these data could be used. The Medicare Modernization Act in 2003 established Medicare Part D to provide beneficiaries with prescription coverage. But, it also stipulated that the Secretary of Health and Human Services could neither establish a formulary or negotiate drug prices. The research community will need to work with policy makers and legislators to overcome hurdles such as these to ensure that data from CER studies can fulfill their intended mission and better inform the care of patients within the US healthcare system.

---

## **2.9 Conclusion**

For policy makers, CER has become an important priority in an effort to identify ways to address the rising cost of healthcare and the shift toward more value-based care models. Acknowledged variation in quality and outcomes as well as an ever-increasing number of new therapeutic options creates a need for a steady stream of data that can better inform patients, providers, and other stakeholders as to the incremental value of a given treatment in a real-world context and to identify and promote the most effective interventions. Because the goal of CER is ensure that individual patients, providers, and the US healthcare system as a whole make the

best healthcare decisions, it will be imperative for the health care and research communities to work in tandem to conduct impactful CER studies on relevant topics and, even more importantly, break down barriers to the dissemination and implementation of data from these types of studies.

**Acknowledgements** The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs, Baylor College of Medicine, or MD Anderson Cancer Center.

*Conflict of Interest Disclosures:* None of the authors have any conflicts of interest to report.

*Funding/Support:* This work was supported by the United States Department of Veterans Affairs HSR&D Service of the VA Office of Research and Development Merit Review award I01 HX002447 (NNM) and the Center for Innovations in Quality, Effectiveness and Safety award CIN 13-413 (NNM).

*Role of the Sponsor:* The funding body played no part in the design and/or general conduct of this study, had no access to the data or role in data collection, management, analysis, or interpretation, and had no role in preparation, review, or approval of the manuscript.

---

## Further Reading

1. Conway PH, Clancy C. Comparative-effectiveness research—implications of the Federal Coordinating Council’s report. *N Engl J Med*. 2009;361(4):328–30.
2. Patient-Centered Outcomes Research Institute. National priorities and Research Agenda. Available at <https://www.pcori.org/research-results/about-our-research/research-we-support/national-priorities-and-research-agenda>. Accessed 4 September 2018.
3. EQUATOR network. Reporting guidelines for main study types. Available at <https://www.equator-network.org/>. Accessed on 10 September 2018.
4. Stukel TA, Fisher ES, Wennberg DE, et al. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *JAMA*. 2007;297(3):278–85.
5. Newgard CD, Lewis RJ. Missing data: how to best account for what is not known. *JAMA*. 2015;314(9):940–1.
6. Li P, Stuart EA, Allison DB. Multiple imputation: a flexible tool for handling missing data. *JAMA*. 2015;314(18):1966–7.
7. Dafni U. Landmark analysis at the 25-year landmark point. *Circ Cardiovasc Qual Outcomes*. 2011;4(3):363–71.
8. Sox HC, Goodman SN. The methods of comparative effectiveness research. *Annu Rev Public Health*. 2012;33:425–45.
9. Anglemeyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev*. 2014; <https://doi.org/10.1002/14651858.MR000034.pub2>.