

12-2012

Getting Answers We Can Believe In: Methodological Considerations When Using Electronic Clinical Data for Research

EDM Forum

AcademyHealth, edmforum@academyhealth.org

Follow this and additional works at: http://repository.academyhealth.org/edm_briefs



Part of the [Health Services Research Commons](#)

Recommended Citation

AcademyHealth. "Getting Answers We Can Believe In: Methodological Considerations When Using Electronic Clinical Data for Research," EDM Forum, December 2012.

This Original Article is brought to you for free and open access by the EDM Forum Products and Events at EDM Forum Community. It has been accepted for inclusion in Issue Briefs and Reports by an authorized administrator of EDM Forum Community.



Issue Brief

Getting Answers We Can Believe In: Methodological Considerations When Using Electronic Clinical Data for CER and PCOR

About the EDM Forum

The Electronic Data Methods (EDM) Forum is a three-year grant from the Agency for Healthcare Research and Quality (AHRQ) to advance the national dialogue on the use of electronic clinical data for the conduct of comparative effectiveness research (CER), patient-centered outcomes research (PCOR), and quality improvement (QI). The EDM Forum facilitates exchange and collaboration between eleven AHRQ-funded projects, including: the Prospective Outcome Systems using Patient-specific Electronic data to Compare Tests and therapies (PROSPECT) studies; the Scalable Distributed Research Networks for CER, the Enhanced Registries for QI and CER, as well as other relevant healthIT initiatives. The EDM Forum and the research projects connected to the Forum are funded through the American Recovery and Reinvestment Act of 2009 (ARRA).

Abstract

The increasing availability of electronic clinical data (ECD) may present opportunities to answer research questions that can drive improvement in patient outcomes. This brief outlines important research design considerations when using ECD for comparative effectiveness research (CER) and patient-centered outcomes research (PCOR). A practical orientation to study designs is provided to encourage consideration of which methods are best suited to answer particular research and quality improvement (QI) questions using ECD, and ways of making information obtained from ECD more useful in the clinical setting are suggested. Differences between observational studies and randomized controlled trials (RCTs) are outlined, and specific challenges of each are discussed. Promising study design and analytic approaches that can be used with observational ECD for CER - and considerations for each - are addressed.

I. Introduction and Background

Today, health information technology (healthIT) is facilitating the collection, aggregation and efficient analyses of large quantities of electronic health data. Data warehouses, distributed data networks, registries, and novel point-of-care data collection technologies are only a few of the many approaches that are exponentially increasing the amount of new and newly accessible data for research. Health care research is joining other fields in exploring numerous uses for “big data”¹, particularly the use of data from electronic health records (EHRs) and other forms of electronic clinical data (ECD).

This data explosion offers many opportunities for comparative effectiveness research (CER) and patient-centered outcomes research (PCOR). CER and PCOR represent efforts to re-orient clinical research so that it produces information that is useful to patients, families, and clinicians when they face decisions about medical treatment. Researchers, health care



AcademyHealth

providers, insurers, payers and patient and consumer advocates are all excited by the potential of new technology and data systems to facilitate discovery and improve care. Ultimately, new data streams and new ways of structuring studies with this data present opportunities to address questions that in the past were difficult, if not impossible, to answer.

While this excitement is warranted, it has to be tempered by an understanding of appropriate strategies that maximize the utility of these new data for research. Access to large amounts of data does not in itself guarantee correct or *useful* answers to CER questions.² Close attention to research design and data analyses is necessary in order to re-use data originally generated for patient care (or other) purposes in clinical research. As a result, it is important for the research community and research users (policymakers, providers, patients, etc.) to be both critical and flexible in thinking about which combinations of study designs and ECD are appropriate to answer particular questions. For example, some types of evidence are most helpful when comparing the benefits of interventions, however, no one study design, neither randomized controlled trials (RCT) nor observational studies, is “best” or even appropriate for all questions.

While RCTs are often considered a gold standard for comparing the benefits of interventions, there are several circumstances when they not feasible or even possible (e.g. when comparing long-term or rare events). For this reason it is helpful to identify key features of RCTs and consider ways of conducting rigorous research that can answer useful questions without always requiring randomization. Likewise, access to large datasets with millions of records and hundreds of variables do not necessarily guarantee these data can provide answers to key CER questions. Given the need to think more flexibly about how to develop useful, rigorous studies that make the most of new data resources, this brief aims to address key questions the research community should consider with respect to understanding *what data and methods are most useful to answer specific CER and PCOR questions*.

Three key issues need to be addressed:

- methodological challenges of traditional research designs for effectiveness studies;
- methodological challenges of using ECD for effectiveness studies; and
- promising study design and analysis approaches that can be used with observational ECD for CER, and considerations for each.

Defining Key Terms in CER and PCOR

As in all scientific disciplines, there is a wealth of methodological terminology in CER, PCOR, and quality improvement (QI). Understanding the range of major methodologies (and related terms) offers important insights into which methods are useful for which purposes. This brief is meant to be accessible to a range of potential audiences, and for this reason key concepts are described in brief. For more complete definitions of terms and more discussion of these methods please visit the EDM Forum website, www.edm-forum.org. Definitions are provided in the wiki glossary on the site. Additional methodological resources for CER are available at www.hsrmethods.org, and through the DEClIDE methods center at www.drugapi.org.

Comparative effectiveness research (CER) is the conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in “real world” settings. The purpose of this research is to improve health outcomes by developing and disseminating evidence-based information to patients, clinicians, and other decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances.³

Patient-Centered Outcomes Research (PCOR) helps people and their caregivers communicate and make informed health care decisions, allowing their voice to be heard in assessing the value of health care options.” This research answers patient-centered questions such as:

- “Given my personal characteristics, conditions and preferences, what should I expect will happen to me?”
- “What are my options and what are the benefits and harms of those options?”
- “What can I do to improve the outcomes that are most important to me?”
- “How can clinicians and the health care delivery system they work in help me make the best decisions about my health and healthcare?”⁴

In addition, it is important to at least briefly discuss the ways in which decision-makers may use information gained from a particular study, as well as how specific studies may be evaluated when assessing the net benefit based on the available body of evidence. Understanding this context is critical to assessing which methods are most appropriate to answer questions of interest.

Two examples of useful CER and PCOR studies that use observational designs are described on page 3.

Examples of Observational CER Studies

Bariatric surgery

There is a sizeable literature on treatments for obesity, including lifestyle changes, medication, and bariatric surgery. While the literature is sparse on long-term randomized trials that measure the effectiveness of bariatric surgery on weight loss, diabetes, and cardiac risk factors⁵, a 2004 AHRQ report looked at over 1,000 studies to examine the effectiveness and safety of the pharmacological and surgical interventions for obese individuals.⁶ Among the studies reviewed was the large and highly influential Swedish Obese Subjects Study (SOS) a matched cohort design that compared bariatric surgery to medical interventions⁷ as well as evaluating the impact of different treatments for obesity on weight loss, comorbidities, and quality of life.

In the studies that were reviewed, morbidly obese individuals (BMI>40) were shown to achieve greater weight loss following bariatric surgery compared to pharmacological treatments. Subjects undergoing surgery lost on average 20-30 kg of weight; maintained the lost weight for 8 years; and saw improvements in comorbidities associated with obesity, such as sleep apnea and diabetes. Altogether, the authors found that “quality of life in the severely obese” was “improved by substantial weight loss.” The researchers also found that patients with a BMI between 35 and 40 also benefitted from bariatric surgery, though the findings were not conclusive. Altogether the study concluded that the comprehensive data from observational studies demonstrates that bariatric surgery is the most effective option for morbidly obese patients, and is relatively safe (with mortality rates below 1 percent in some settings).

Identifying problems with replacement joints

Successful joint replacement surgery can restore a patient’s mobility, relieve pain and improve quality of life. The safety and efficacy of replacement joints are tested in clinical trials before they are approved for general use by the FDA. However clinical trials cannot test artificial joints in all types of patients likely to use them. Nor can these studies be continued for the decades over which patients with artificial joints are likely to live.

Registries provide a way to collect and link the data needed to track the outcomes of surgery and performance of different types of artificial joints over an extended period. A 2010 analysis of data collected in the National Joint Registry of England and Wales revealed that patients with metal-on-metal hip implants required a second surgery to replace or remove part of the artificial joint (called a revision) more frequently than was expected based on clinical trial results.⁸ Further investigation led to a worldwide recall of one implant by the manufacturer, more intensive monitoring by the FDA⁹ in the United States, and extensive research about metal-on-metal joints using other registries maintained by provider organizations, the United States and other countries.^{10, 11, 12}

Differentiating Efficacy and Effectiveness

A 2006 Agency for Healthcare Research and Quality (AHRQ) report¹³ proposed a set of nine criteria for distinguishing effectiveness trials from efficacy trials. RCTs that adhere to the following criteria are considered to be pragmatic, and to more accurately reflect decisions in the clinical setting on a day to day basis.

1. Using populations from primary care settings (as opposed to efficacy trials, which tend to take place in specialized care facilities).
2. Use less stringent eligibility criteria in order to enroll a more heterogeneous group of participants that better reflects real-world population characteristics.
3. Measure health outcomes rather than surrogate or intermediate outcomes (especially when our understanding of the natural history of a disease is incomplete and the relationship of surrogate outcomes to health outcomes is unclear).
4. Study durations in effectiveness trials should reflect length of treatment in the clinical setting in actual practice.
5. Define compliance as an outcome measure, since adherence is often low in the clinical setting.
6. Use relevant comparators that reflect the treatment choices that are made in the clinical setting.
7. A large enough sample size to ensure that the study is adequately powered to detect effects that may be small but clinically significant.
8. Use intention to treat analysis, which includes participants that break protocol or switch treatments, rather than excluding them.
9. Assessment of adverse events may be limited to important outcomes observed in previous trials, because extensive reporting of adverse events does not always occur in the clinical setting.

II. Understanding Key Characteristics of Randomized Controlled Trials (RCTs) and Observational Studies

In order to delineate important issues related to using ECD for research and QI it is essential to understand key characteristics of RCTs; how observational studies differ; and the strengths and limitations of different study designs.

RCTs are considered the gold standard to compare outcomes of different treatments in biomedical research by many people because—at least theoretically—RCTs allow clearer conclusions about the effect of the treatment being studied. A defining characteristic of RCTs is that patients are randomly assigned to treatment and control groups. Random assignment is used to simulate the counterfactual. The counterfactual is what would happen if we could rewind time and have the same group of patients live twice, once with the treatment and again exactly the same way with no or an alternative treatment. Since this is not possible, randomization is used to create groups that are as similar as possible, or with randomly distributed differences that are less likely to be the cause of differences in the results for the two groups. This approach minimizes problems with selection bias that may invalidate a study's conclusions.

RCTs fulfill the three core requirements for effect identification: 1) positivity (all patients have a probability of being treated); 2) exchangeability (patients can be swapped between treatment and control groups without changing results); and 3) consistency (the treatment is the same across patients). A more in-depth discussion of effect identification in CER, and reasons why large ECD datasets cannot always address the challenges of effect identification is available in Oakes.¹⁴

In practice, patients and physicians select the treatments that are likely to be best for the patient. However, in research, since it is crucial to isolate the effect of a treatment from other characteristics that could impact outcomes, the complexity of real-world treatment decisions creates potential sources of bias that can compromise a study. One is the so-called “healthy user effect.” An example from a 2007 study on adherence to lipid-lowering therapy illustrates how this phenomenon can result in an overestimation of the effectiveness of a treatment.¹⁵

The study found that patients who take medications (such as statins) for preventive purposes are more likely to engage in other behaviors that tend to improve health outcomes, including exercise and adhering to a healthy diet. Consequently, these patients are likely to see greater improvements in health than those who eschew the preventive treatment, leading to a greater apparent treatment effect.¹⁶

While studying the “real world” treatments selected by patients and clinicians using observational data rather than RCTs makes it difficult to isolate the effects of a specific treatment, RCTs as they have been traditionally conducted are generally not considered sufficient for CER and PCOR. The reason for this is that these studies usually cannot answer questions such as “what is likely to happen to patients like me?” or “which treatment for my condition is likely to result in the outcomes that are important to me?”

Here it is useful to clarify the purpose for which specific studies are conducted. While efficacy is a reflection of how well the treatment works under ideal conditions (typical of many RCTs), effectiveness addresses how well a treatment will work under the unpredictable conditions of the clinical setting.

By comparison, studies where the research involves watching and analyzing, but not manipulating or interfering with ongoing treatment are referred to as observational studies because the researchers observe, rather than control, what happens. ECD provides promising opportunities to improve the utility of observational studies because it can facilitate studies on more relevant patient populations and treatments and because it can be paired with statistical methods to adjust for potential sources of bias.

Essentially, ECD provides a new opportunity to use methods to address research questions that previously were not feasible or ethical to conduct with RCTs. Answering these questions was challenging given limitations in the availability of linked clinical data. However, as these data are brought together across research networks using ECD it becomes possible to identify enough cases to study rare diseases, less common side effects, and long-term outcomes, typically using observational study designs. ECD also makes it feasible to study treatment effects in several subgroups of patient populations. The following box contains three examples of how studies can use ECD for QI.

Studies Using Electronic Clinical Data for QI

Electronic Patient Reported Outcomes (ePRO)

In order to improve care and quality of life for patients with cancer, researchers at Duke University Medical Center are using a tablet computer-based system known as ePRO (Electronic Patient-Reported Outcomes) to collect patient-reported data on symptoms, side effects, experiences, and satisfaction. They are also creating ways to link ePRO with administrative and clinical data. The combined data have been used to identify issues in patient care, develop interventions, and evaluate their impact. ePRO and the related data systems allow researchers and clinicians to analyze measures important to patients, providers, and institutions and use this information to improve outcomes, safety, and quality.¹⁷

The ePRO studies confirmed that a significant number of patients report psychological distress (anxiety, depression, and mood disturbance) during cancer treatment. Using data from these studies, clinicians and researchers were able to identify patients, track their symptoms, and conduct a pilot test of an intervention designed to relieve their distress. The pilot test achieved positive results, and will serve as the foundation for larger studies that include more patients with different cancers, examine the intervention's impact on cost and use of health services, and compare different interventions.

Using EHRs to identify and track patients with depression

Researchers from the University of Colorado and the Distributed Ambulatory Research in Therapeutics Network (DARTNET) used EHR data, as well as information on drug fulfillment and findings from a public health questionnaire, to identify and follow patients diagnosed with depression. Among these patients, the researchers collected data on symptom severity and suicide risk.

The researchers used the PHQ-9 PRO measure to assess symptoms of depression and observed significant variation in illness severity and suicide risk among the patient population. Integrating EHR data with information on prescription fulfillment and the public health questionnaire allowed physicians to better judge disease severity and suicide risk, and intervene immediately in the course of a visit if necessary. Collecting information on symptoms and integrating this information in the course of a patient visit was assessed by physicians to improve the quality of care and outcomes for depressed patients.¹⁸

Using EHRs and PRO to support CER and QI

Researchers from Cincinnati Children's Hospital Medical Center (CCHMC) have established the Enhanced Registry, linking centers across the ImproveCareNow network enhanced registry to study inflammatory bowel disease (IBD) and Crohn's disease in pediatric patients, using an open-access data sharing network.¹⁹ The goal of the network is to achieve quality improvements by increasing care management by providers, and also patient involvement in their own care. A critical dimension of this effort is achieving the goal of "data in once", which means that high quality data ("research grade" data) is collected in the process of care, without duplicate data entry. To achieve this goal, a closely linked project, the pediatric Enhanced Registry, addresses key governance, informatics, and methodological challenges that are critical to improving data and methods for research and QI.²⁰

Specific aims include methods to assess and improve the quality of data collected on chronically ill patients; and using QI methods to improve disease management using registry data to compare different treatment strategies (especially different biologic agents). Underlying these methodological aims is the need for a pragmatic governance model to guide the secure use of these data for both QI and research. In one study being conducted, data from the EHR combined with PRO data from current and future pediatric patients will be used to evaluate alternative adaptive treatment strategies for newly-diagnosed patients with the goal of improving the quality of care for pediatric patients with digestive system diseases.²¹ Other studies are assessing approaches for using and integrating PRO data from mHealth tools that actively or passively collect information on symptoms between visits. These types of studies will contribute to an understanding of ways in which new technology and validated PRO instruments can contribute to understanding patient's experiences in care.

Much of the interest in ECD is based on the idea that access to big datasets with more patients and more types of data may allow researchers to design studies that do not have the common disadvantages of RCTs. However, it is important to consider why, and ways how ECD might facilitate CER and PCOR using a variety of study designs. Such approaches include large

pragmatic trials or observational studies that match hundreds of patient characteristics and compare treatments and outcomes across patient groups.

The following section discusses key challenges of using ECD for research.

III. Key Challenges in using ECD for Research: Data Quality, Risk of Bias, and Confounding

Innovations in data technology and availability of large quantities of data provide unique opportunities to conduct CER and PCOR studies that would otherwise be impossible. However, using ECD for research also presents significant methodological and analytic challenges, including assuring data quality and minimizing specific risks of bias. In addition, though not explicitly discussed in this brief, there are significant challenges related to other aspects of developing a robust data infrastructure such as collection and storage and exchange, particularly among multi-site research projects.²² These issues can significantly impact the data quality and methods appropriate to answer CER questions. Numerous other resources address these and other components of infrastructure.^{23, 24}

Data Quality

Data quality is a critical issue for any research or analytic activity. In order to ensure that data collected within systems is accurate, valid, and reliable, it is necessary to address specific challenges that may be encountered with ECD. These include:

- Incomplete, missing, or inaccurate data.
- Reconciliation of inconsistent definitions for terms, variables, values and time periods/episodes. This can be particularly difficult when the data were originally created for non-research purposes (such as billing) or when data are stored in legacy systems that are difficult to update or do not interface easily with other software.

Understanding the level of data quality in a research project is necessary if data are to be optimally used for research. Currently there is limited literature or guidance available about how to best assure the quality of ECD so that it can be used in single or multisite studies. Increased interest in using the data for research is changing this and recent efforts have included developing a framework that defines the key domains of data quality as accuracy, objectivity, believability, timeliness, and appropriate amount of data. The proposed framework then integrates these domains in to a process model for data quality assessment that includes data quality rules and methods of assessment).^{25, 26}

Further work in the area of data quality assessment is methodologically important since transparent assessment of data quality can help ensure researchers do not use approaches that lead to incorrect conclusions about the effectiveness of treatments.

Risk of Bias

Risk of bias refers to the possibility that clinical study results deviate from truth because they include “bias in study design or execution in addition to the true effect of the intervention.”²⁷ The underlying idea is that systematic differences cause this deviation from the truth.

Risk of bias is often divided into types or sources. A common classification for types of bias is provided in Table 1.

Table 1: Common Classification of Types of Bias²⁸

Type	Systematic differences between:
Selection bias	Baseline characteristics of the groups that are compared
Performance bias	Groups in the care that is provided, or in exposure to factors other than the interventions of interest
Attrition bias	Groups in withdrawals from a study
Detection bias	Groups in outcomes are determined
Reporting bias	Reported and unreported findings

Biases are possible in RCTs *and* in observational studies, and often involve unmeasured confounders. The difference is that some types of bias may be more likely in specific types of studies. For example, selection bias is rare in RCTs when sample size is large enough and randomization is done correctly. Similarly, attrition bias is less likely in observational studies where people may not even know they are included in a study (though it can occur for example when patients dis-enroll from the health plans that maintain the electronic clinic database, resulting in missing data).

Minimizing the risk of biases and their potential impact is at the heart of many study design elements and analysis techniques that are being developed and refined for use with ECD. Adjusting for unmeasured confounders (confounding is discussed below) is also possible by external or internal validation studies.²⁹

Confounding

Confounding is of particular concern in observational studies and other non-randomized study designs. Confounding occurs when factors other than the exposure of interest influence the outcome of interest, leading to a misinterpretation of the effect size.³⁰ Smoking is a good example of a powerful confounder in a study exploring the relationship between another exposure and lung cancer. RCTs use randomization to balance confounders and report the characteristics of the study population in both the treatment and comparison group to demonstrate whether randomization successfully balanced known confounders.

In contrast to bias, it is possible to adjust for confounders using methods such as stratification; creating a pooled estimate using the Mantel-Haenszel method; or a multiple regression model.³¹

IV. Matching Questions to the Right Study Design, Data, and Analytic Methods

Finding a good match between CER and PCOR research questions and ECD and then producing valid results that address these questions requires researchers to adapt traditional study designs and develop new analytic methods. Researchers who want to conduct CER and PCOR using ECD—as well as the decision-makers who use the results—need to consider the strengths and limitations of specific study designs and analytic methods as they relate to the characteristics of ECD.

Developing a clear conceptual model of the ideal study design and data requirements is a critical first step.³² With more and different types of ECD available, investigators have the ability to use different kinds of data, and think through which analytic techniques are most appropriate. Ultimately it is the methods (both study design and analysis) of studies using ECD that will determine whether the level of rigor and relevance of these findings is better able to answer CER and PCOR questions to meet decision-makers' needs.

For example, each study should consider the defining characteristics of CER and PCOR (e.g. comparisons of two or more treatments that are viable options for patients and clinicians; that “real world” conditions that are subject to no or only modest manipulation; that important outcomes are evaluated; diverse patient populations are included, etc.) and evaluate which study designs are most effective to meet particular research goals. Tools such as PRECIS can be used to help formalize this type of assessment on a continuum from exploratory to pragmatic, where pragmatic corresponds closely to the defining characteristics of CER and PCOR.³³

With respect to observational studies conducted with ECD, the number of patients available in ECD may be an advantage for either observational studies or trials designed for CER or PCOR. For example, in any study where two treatments are compared and both are expected to lead to some improvement, a larger sample size will be needed to detect a meaningful difference between the two compared to the sample size needed for a comparison between one treatment and a placebo. ECD may provide the sample sizes needed for studies that could not have been conducted without large datasets.

One of the most exciting developments related to using EHRs for research is that the infrastructure being developed to collect ECD allows for innovation in trial design and conduct, such as pragmatic trials³⁴ and point-of-care clinical trials.³⁵ These permit changes such as adjustments to treatment protocols during the course of the study, inclusion of more diverse patients, and different approaches to randomization. Additionally these studies may be mapped to, or complemented by rich prospective observational data, including ePRO. Since many of these adaptations are data driven, or at least data dependent, such approaches have the potential to make research findings more relevant for decision making while still minimizing risk of bias.

The next section provides examples of research designs and analytic methods that are being used or have been proposed as ways to take advantage of the potential of new and increasing amounts of ECD. A very brief overview of selected methods is provided in order to illustrate their advantages and limitation as well as why ECD is relevant to their use.

V. Methods for Maximizing the Utility of ECD with Observational Data

Observational designs used for CER are similar in basic structure to observational designs in traditional clinical research. Examples of common observational designs relevant for CER include time series, natural experiments, cohort studies and case-control studies. Examples of data driven approaches that may leverage ECD to reduce the risk of bias in observational studies are described very briefly below. Several experts have outlined the possibilities.^{36, 37}

Matching Methods (MM) and Propensity Scores (PS)

Matching Methods (MM) and Propensity Scores (PS) combine data on observable characteristics to create a score for the probability that a patient received one treatment as opposed to another. These scores can be used to either match patients in each treatment group on their score, or scores can be included as a weight or a control variable in a statistical model of the outcome. Particularly where propensity scores are used to match patients across groups, the objective is to account for observed sources of potential bias in order to produce balanced groups for comparison even when treatments have not been randomly assigned. Studies have documented that analyses of groups created using PS can achieve results similar to those obtained in RCTs.³⁸

However, using MM and PS that includes numerous variables also requires trade-offs. Increasing the number of variables in a propensity score may make it better able to estimate the probability of group assignment, but it may make it harder to find patient matches across groups. Patients without a match are dropped from the analyses, making the results less applicable. Extensions of these methods such as high dimensional propensity scores require large volumes of data across (potentially) hundreds of variables or characteristics. Even ECD may not provide enough information to use these techniques successfully.

Profiling

Profiling patients in terms of risk or the potential to benefit from treatment involves using techniques to model the predictive value of observable patient characteristics or to create composite variables based on patient characteristics. Profiling techniques and composite measures allow the comparison of treatment effects when the data includes heterogeneous patients. This is accomplished by conducting analyses according to the sub categories of patients created by the profiles.³⁹

As is the case with PS, data availability is essential to creating profiles. Having both a large number of variables available to create profiles and large numbers of patients available so that treatments and outcomes can be examined by subcategories is necessary.

Instrumental Variable Analysis

Instrumental Variable Analysis (IV) involves identifying a measured variable that is related to the treatment group patients are in, but not related to the outcome of interest. IVs are used when it is believed that the groups of patients being compared differ on variables that are not observed. In an RCT the random assignment determines the treatment group but is not associated with the outcome, which is similar to an IV, and randomly distributes characteristics that cannot be observed. Finding a variable that behaves like randomization in this way allows groups to be created from the observational data that are not biased by how treatment was determined even when treatment was not randomly assigned.^{40, 41}

Identifying a variable that meets the IV analysis requires meeting stringent assumptions and researchers who attempt to find a valid instrument in available data often fail. Even if a valid instrument can be identified, if it is not strongly associated with treatment the analysis is unlikely to explain variation in outcomes. While it is true that instruments can be very hard to identify, the availability of large quantities and new sources of information in ECD increase the possibility that IVs can be identified and used successfully.

Fixed Effects Models and Regression Discontinuity Methods

Fixed effects models and regression discontinuity methods, like IVs, are approaches designed to address unobservable or unmeasured differences between treatment groups in observational data.⁴² In a fixed effects model it is assumed that differences across the groups are characteristics that do not change with time (or at least in the time frame that involves the study treatment and outcome), such as education level, literacy, or attitudes about cancer prevention or genetic characteristics. It is important to note that whether an individual characteristic is viewed as changing or not over time depends in part on the population and study question, but if unobserved yet important differences are fixed then patients can serve as their own controls.

In essence, the design compares outcomes for patients who experience different treatments or exposures. This can be difficult because it is essential to be sure the important differences really are fixed, and it can be hard to identify eligible patients who have experienced multiple treatments as this is likely to be a small subset of the population of interest. Because of these restrictions the results may not be applicable to the larger population of patients.

Regression discontinuity is similar in that it strives to adjust for differences between treatment groups that cannot be observed or measured. This is accomplished by restricting the study to a subgroup of the population that is more likely to be similar as a means to reduce differences between groups. The approach requires that a cut-off value divide patients in to two groups: one group receives one treatment and is on one side of the cut-off while patients on the other side receive a different treatment. When this is the case, it might be possible to compare patients just above and just below the cut-off and make the assumption that they are more likely to be similar on unobserved variables. The comparison is then restricted to these two slices of the population closest to the cut-off. ECD makes this approach possible because with large numbers of cases is it more likely that there are enough patients on the two sides of the cut-off to have a sufficient sample size. However in some clinical situations the patients on the two sides of the cut-off may be so different that the comparison would not be valid. Furthermore, using only small subgroup of the population of interest may provide groups that are more comparable, but it produces results that are not applicable to as many patients as results that can be produced from other designs.

The approaches described here as well as others in use and under development address the problem that observational studies might be “comparing apples to oranges” in different ways. Matching Methods, Propensity Scores and Profiling use data on observable characteristics to create comparisons that are more likely to be “orange-to-orange” and “apple-to-apple.” Instrumental variables, fixed effects models, and regression discontinuity models attempt to work around the fact that there may be unobservable differences in the treatment groups in observational studies by creating groups in which these unobservable characteristics are more likely to be similar. What is similar across all the approaches is that they all require data with large numbers of patients and variables. Whereas past applications of these methods have not always been able to address concerns about the validity of observational studies, ECD has potential to make these methods more feasible. While having large quantities of data does not guarantee that the assumptions underlying these methods can be met, it does expand the possibilities.

VI. Adapting RCTs to Study CER and PCOR

While more commonly used in observational study designs, ECD also can be a valuable tool in RCTs. A variety of adaptations to traditional RCTs have been developed that attempt to retain the advantages of randomization while increasing the applicability and value of trial results for clinical decisions. As discussed previously, efforts to make RCTs more pragmatic and focused on effectiveness is a clear goal. Using new methodological adaptations to enhance trials is similarly important. The following box illustrates some of the methodological adaptations used by the Observational Medical Outcomes Partnership (OMOP).⁴³

Testing Analytic Approaches to Match Results from RCT and Observational Study Designs

Recently, the Observational Medical Outcomes Partnership (OMOP), funded by the NIH Foundation, tested the ability to match results from RCTs with observational study designs applied to large administrative and clinical data sets in studies of drug effectiveness. The following analytic approaches were tested:

- Adapted self-controlled case series
- Observational screening
- Disproportionality analysis
- Temporal pattern discovery,
- Bayesian logistic regression,
- Surveillance,
- Sequential probability ratio test,
- Sequential sampling, and
- Incident user design

Results indicated a varying degree of comparability between specific methods, and the OMOP team has concluded that no specific method demonstrated superior performance. However, the aggregate results provide a benchmark and baseline expectation for risk identification method performance. OMOP also found that systematic processes for risk identification can provide useful information to supplement an overall safety assessment, but caution that current methods suggest a substantial chance of identifying false positive associations.

The feasibility of these efforts relies on the ability to link or merge data across delivery settings and institutions and simplify strategies to randomize participants.

Pragmatic or Practical Clinical Trials

Pragmatic Clinical Trials (PCTs) are sometimes also called Effectiveness Trials in order to highlight that they differ from efficacy trials (see Section II for a discussion of the relevant differences). PCTs are basically RCTs refashioned to focus on “who the treatment works for and when” (effectiveness) rather than “does the treatment work under ideal circumstances” (efficacy).⁴⁴ ECD can facilitate pragmatic trials because there is often a need to collect data across sites and to link primary data collected through the trial to existing administrative and clinical data (e.g., patient charts, lab and imaging reports). Often all the data needed would have been either impossible or too expensive to collect without access to HIT and EHRs.

Cluster Randomized Trials

In cluster randomized trials (CRTs) researchers randomly assign groups of individuals, rather than individuals to the different treatments. CRT groups or clusters can be any size or type. Studies have been conducted where the clusters are primary care teams, hospitals, health plans, communities or even states.

Using groups rather than individuals allows trials to be conducted in typical settings and with a more diverse group of patients than could be obtained in a randomized study of individuals.^{45, 46} Another advantage of CRTs is that they help reduce the chance of patients assigned to the control groups receiving the treatment. For example, it may not be realistic to ask physicians to use a new smoking cessation approach with some patients and not others—but primary care practices could be randomized to use the new approach or not for all their patients. Similarly a media campaign to increase whooping cough vaccinations cannot be randomized to individuals, but cities or states could be randomized to be target markets for the media campaign and then their immunization rates could be compared.

One of the key challenges with CRTs is that analysis has to consider variation within each cluster as well as across the clusters.⁴⁷ To do this and have the desired power to detect statistically significant differences can require much larger numbers of patients than if individuals were randomized. With ECD from data networks, identifying and randomizing clusters and then collecting equivalent data from numerous clusters is possible, making this approach more feasible than it was in the past. CRTs also create challenges for implementation because they often require involvement of front-line staff, which may necessitate training or alteration of existing workflows.

Designed Delay Trials

Designed Delay Trials (DDTs) are also called Wait-list, Phased Implementation, Staggered Start or Randomized Start. These trials take advantage of the fact that often a treatment, program or policy cannot be made immediately available to every person or in every location; implementation has to be done in stages for many practical reasons. Designing the distribution of the treatment by random assignment to delay or no delay rather than allowing implementation to happen in a haphazard fashion, makes it possible to assess the impact of the treatment or policy in the settings where it will be implemented. By controlling the timing of implementation, it is possible to compare the outcomes of the early implementation and delayed groups. The groups are similar in that they are all eligible for the intervention. Random assignment makes them less likely to differ in other ways and limits bias that may result when organizations choose or volunteer to be in the initial or later implementation groups.⁴⁸ Using randomization as a way to determine the rollout of a new treatment, program or service is also a fair way to allocate a scarce resource.

Point-of-Care and Adaptive Trials

Both Point-of-Care Trials and Adaptive Trials are newer and less well known. They share an essential characteristic: key trial elements such as the number of patients enrolled, treatment details, and eligibility criteria can be changed during the trial. The changes need to follow rules developed as part of the trial design. Point-of-Care Trials add an emphasis on embedding the

clinical trial in the apparatus of routine clinical care.⁴⁹ Sequential multiple assignment randomized trials (SMART) use adaptive interventions to study the “development of evidence-based adaptive intervention strategies.”⁵⁰

The advantages are that results may be available in shorter periods of time, treatments can maintain clinical relevance, clinical care can be delivered without intrusion of study personnel or restrictive protocols, and results are more likely to be relevant to regular clinical care. The difficulty is that developing and implementing adaptations to ongoing trials may require complex statistical approaches and expertise. Similarly Point-of-Care trials require EHRs and integrated information systems that can be used not only to identify patients, but also to prompt clinicians to consider enrolling appropriate patients, randomize enrolled patients, track decisions and monitor treatment and allow ongoing analysis of results. This capacity is not currently universally available and these requirements limit these designs to organizations with advanced statistical and informatics capacity.

As ECD becomes more available, the data needed to drive variations on traditional RCTs will become accessible in more situations. Expansions of data networks and technological advances are increasing the complexity of information systems available to larger numbers of clinical practices. Remote access and linked systems allow the inclusion of more sites than would have been feasible in the past. It is this growth and innovation that will make trial adaptations including pragmatic, cluster, designed delay, point-of-care, and adaptive trials feasible and useful for future clinical research, though many of the challenges of data quality are methodological considerations that will continue to be important for the community to address.

VII. Using evidence to make health care decisions

As decision-making bodies are presented with new evidence emerging from studies using ECD with new methods, it is important to consider how current decision-makers, such as guideline setting bodies and purchasers or payers, use CER and PCOR to inform practice and coverage. While an in-depth discussion of ways in which evidence is used to make health care decisions is outside the scope of this brief, it is useful to include an overview of the key methodological considerations used by guideline setting bodies such as the US Preventive Services Task Force (USPSTF); and purchasers or payers.

Single studies rarely address all of the issues of interest to decision-makers. As a result, most decision-makers consider the *body* of evidence that is relevant for making a decision.⁵¹ As an example, the USPSTF looks at the chain of evidence that links a screening test to patient outcomes. After examining the relevant body of evidence, the USPSTF makes recommendations based on assessing the *magnitude*

of net benefit on patient outcomes, as well as the *uncertainty around that estimate*, after examining the relevant body of evidence.⁵²

For a screening test, the issues include the:

- accuracy of the test,
- benefits and harms of administering the test,
- benefits and harms of the intervention provided on the basis of test results
- added benefit of providing an intervention early to screen-detected population compared to those detected in a clinic later in the course of disease.

If the studies have examined surrogate (or intermediate) outcomes alone, the USPSTF will look at the strength of evidence that intermediate outcomes are associated (or closely linked) with health outcomes. The chain of evidence is particularly useful when direct evidence for impact of a screening test on patient outcomes from a RCT is not available.

While assessing the internal validity⁵³ of studies is a key component of evaluating the body of evidence,⁵⁴ the USPSTF also uses explicit criteria to assess the generalizability (external validity) of the studies.⁵⁵ To arrive at a recommendation the USPSTF assesses evidence for each key question (or the link in the chain of evidence) as well as assessing the certainty of evidence for the entire preventive service.⁵⁶ When the certainty of evidence is moderate or high, the USPSTF uses all admissible evidence to assess magnitude of the benefits and of the harms (typically by making an Outcomes Table) to assess magnitude of net benefit.⁵⁷ The EGAPP working group has also developed an analytic framework and analogous methods to those of the USPSTF in making recommendations on genetic/genomic tests.⁵⁸

The perspective of both public and private-sector is important to consider because purchasers and payers have different priorities than guideline-producing organizations. While guidelines are typically determined by comparing the balance of harms and benefits of different treatments, payers also focus other criteria such as applicability to their own populations, cost, cost-effectiveness, and feasibility. One example of this is the Washington State Health Care Authority, which uses a robust five step process to determine whether a new technology should be covered.⁵⁹ The Washington State Health Care Authority determination includes a review of the existing evidence, as well as obtaining information from outside stakeholders and advisory groups. The process also requires analysis of “special policy or clinical considerations that could affect the review,” including disproportionate impacts on minority or other disadvantaged groups.

The Blue Cross Blue Shield Technology Evaluation Center and the California Technology Assessment Forum (CTAF) use five criteria to evaluate new technologies.⁶⁰ In addition to requiring approval from relevant government regulatory bodies and evaluating evidence of net benefit and comparative efficacy on health outcomes for a given technology under review, they also evaluate whether the improvements associated with the technology are likely to “be attainable outside the investigational settings,” thus ensuring that the new technology demonstrates not just efficacy, but also effectiveness. United HealthCare is an example of a payer, along with many other health plans, that determines whether the technology is actually “medically necessary” for a specific use based on available evidence, and provides clinically significant improvements.⁶¹

VIII. Conclusion

The availability of large quantities of electronic data; increased standardization of data; and emergence of new data collection strategies are transforming health care, QI efforts, and health research. More data are now accessible on diverse patients, multiple aspects of care, and a range of outcomes than has ever been available before. These developments present opportunities for innovation in CER and PCOR that would not have been feasible in the past. Results from this research have the potential to change clinical care, health services delivery, and ultimately improve the health and wellbeing of patients.

However, these benefits cannot be realized simply because we have more data. Which questions are asked, how the data is collected, and how results are generated matters. Study designs and research methods will determine whether we can leverage these new sources of data to generate useful information. Developing better study designs and more sophisticated analytic methods, and using them appropriately, is essential. This brief outlines some of the promising approaches that are important to understand in order to identify which methods are most appropriate for particular questions using ECD.

Research design using ECD is evolving rapidly, and new methods will be developed as we explore the potential of these data. CER and PCOR are increasingly focusing attention on research questions that are important to patients, providers, caregivers, communities, and other key stakeholders. In response, research design must shift towards practical and innovative strategies that address patient-centered questions by maximizing the usefulness of ECD. Ultimately, if ECD is to lead to answers we can believe in, we need research design and analytic methods that are rigorous, transparent, and focus on producing valid results to improve patient outcomes.

About AcademyHealth

AcademyHealth is the leading national organization serving the fields of health services and policy research and the professionals who produce and use this important work. Together with our members, we offer programs and services that support the development and use of rigorous, relevant and timely evidence to increase the quality, accessibility, and value of health care, to

reduce disparities, and to improve health. A trusted broker of information, AcademyHealth brings stakeholders together to address the current and future needs of an evolving health system, inform health policy, and translate evidence into action.

Acknowledgements

AcademyHealth acknowledges the Agency for Healthcare Research and Quality (AHRQ) for its support of this work. The EDM Forum is supported by AHRQ through the American Recovery & Reinvestment Act of 2009, Grant U13 HS19564-01. AHRQ’s mission is to improve the quality, safety, efficiency, and effectiveness of health care for all Americans. As 1 of 12 agencies within the Department of Health and Human Services, AHRQ supports research that helps people make more informed decisions and improves the quality of health care services. For more information, visit www.ahrq.gov.

We thank Dr. Annette Totten for her leadership developing this brief, and Dr. Michael Stoto and Dr. Gurveet Randhawa for their guidance and editorial assistance.

Suggested Citation

AcademyHealth. “Getting Answers We Can Believe In: Methodological Considerations When Using Electronic Clinical Data for Research,” EDM Forum, December 2012.

Also see www.edm-forum.org to access this publication.

Endnotes

1. The Health Data Initiative Forum. The Health Datapalooza. 2012; <http://www.hdiforum.org/>. Accessed September 5, 2012.
2. Oakes J. Effect Identification in Comparative Effectiveness Research. Washington, DC: AcademyHealth;2012.
3. Patient Protection and Affordable Care Act of 2010, Pub. L. 111-14, 124 Stat. 119 (March 23, 2010).
4. Patient-Centered Outcomes Research. 2012; <http://www.pcori.org/what-we-do/pcor/>. Accessed November 1, 2012.
5. Chan C, Wang B, et al. Randomized Controlled Trials in Bariatric Surgery. Obesity Surgery. October 2012. Epub ahead of print.
6. Shekelle PG, Morton SC, Maglione M, et al. Pharmacological and Surgical Treatment of Obesity. Rockville (MD): Agency for Healthcare Research and Quality (US); 2004 Jul. (Evidence Reports/Technology Assessments, No. 103.) Available from: <http://www.ncbi.nlm.nih.gov/books/NBK37379>
7. Sjöström L, Lindroos A, et al. Lifestyle, Diabetes, and Cardiovascular Risk Factors 10 Years after Bariatric Surgery. NEJM. December 2004;351(26):2683-93
8. National Joint Registry. MHRA Medical Device Alert. 2012; <http://www.njr-centre.org.uk/njrcentre/MetalonMetalHipImplants/tabid/237/Default.aspx>. Accessed October 22, 2012.
9. US Food and Drug Administration. Concerns about Metal-on-Metal Hip Implant Systems. 2012; <http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/ImplantsandProsthetics/MetalonMetalHipImplants/ucm241604.htm>. Accessed October 22, 2012.
10. California Joint Replacement Registry. 2012. Accessed October 22, 2012.
11. American Joint Replacement Registry. 2012; http://orthodoc.aaos.org/ajrr/grp_about.cfm. Accessed October 22, 2012.
12. Canadian Joint Replacement Registry (CJRR) Metadata. 2012. Accessed October 22, 2012.

13. Gartlehner G, Hansen RA, Nissman D, et al. Criteria for Distinguishing Effectiveness From Efficacy Trials in Systematic Reviews. Rockville,MD: AHRQ;2006: Apr. (Technical Reviews, No. 12.) 2, Methods. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK44028/>
14. Oakes, op. cit.
15. Shrank WH, Patrick AR, Brookhart MA. Healthy user and related biases in observational studies of preventive interventions: a primer for physicians. *J Gen Intern Med.* 2011;26:546–50.
16. Brookhart MA, Patrick AR, Dormuth C, et al. Adherence to lipid-lowering therapy and the use of preventive health services: an investigation of the healthy user effect. *Am J Epidemiol.* 2007;166(3):348–54.
17. Abernethy AP, Ahmad A, Zafar SY, Wheeler JL, Reese JB, Lysterly HK. Electronic patient-reported data capture as a foundation of rapid learning cancer care. *Med Care.* Jun 2010;48(6 Suppl):S32-38. (WAS 12)
18. Valuck RJ, Anderson HO, et al. Enhancing Electronic Health Record Measurement of Depression Severity and Suicide Ideation: A Distributed Ambulatory Research in Therapeutics Network (DARTNet) Study. *J Am Board Fam Med.* 2012 Sep;25(5):582-93.
19. ImproveCareNow. <https://improvecarenow.org>. Accessed January 2, 2013.
20. Enhanced Registry. <https://enhancedregistry.org>. Accessed January 2, 2013.
21. “Building Electronic Clinical Data Infrastructure to Improve Patient Outcomes: CER Project Profiles,” EDM Forum, AcademyHealth, November 2012. Accessible at <http://tinyurl.com/blmntk5>
22. Sittig DF and Hazlehurst BL. Informatics grand challenges in multi-institutional comparative effectiveness research. *JCER.* 2012;1(5):373-376
23. Sabharwal R, Holve E, Rein A, and Segal C. Approaches to Using Protected Health Information (PHI) for Patient-Centered Outcomes Research (PCOR): Regulatory Requirements, De-identification Strategies, and Policy. EDM Forum, AcademyHealth, March 2012.
24. Holve E, Segal C, Hamilton Lopez M. Opportunities and challenges for comparative effectiveness research (CER) with Electronic Clinical Data: a perspective from the EDM forum. *Med Care.* Jul 2012;50 Suppl:S11-18.
25. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Med Care.* Jul 2012;50 Suppl:S21-29.
26. Maydanchik A. Data Quality Assessment. Technics Publications: Bradley Beach, NJ; 2007.
27. Viswanathan M, Ansari M, Berkman N, et al. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews. Rockville, MD: AHRQ;2012.
28. Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions. Version 5.1.0. Available from <http://www.cochrane-handbook.org>.: The Cochrane Collaboration; 2011. Cochrane book series.
29. Sturmer T, Glynn R, Rothman KJ, Avorn J, and Schneeweiss S. Adjustments for Unmeasured Confounders in Pharmacoepidemiologic Database Studies Using External Information. *Med Care* 2007;45: S158–S165
30. Howards P, Schisterman E, Poole C, Kaufman J, and Weinberg C. Toward a Clearer Definition of Confounding Revisited With Directed Acyclic Graphs *Am. J. Epidemiol.* 2012;176(6): 506-511
31. Fitzmaurice G. Adjusting for Confounding. *Nutrition.* 2004;20:594–596
32. Oakes, op. cit.
33. Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol.* May 2009;62(5):464-475.
34. Luce BR, Kramer JM, Goodman SN, Connor JT, Tunis S, Whicher D, Schwartz JS. Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. *Ann Intern Med* 2009;151:206-9.
35. D’Avolio L et al. Implementation of the Department of Veterans Affairs’ first point-of-care clinical trial. *JAMIA* 2012;19(e1):e170-6
36. Schneider B, Carnoy M, Kilpatrick J, Schmidt WH, Shavelson RJ. Estimating Causal Effects Using Experimental and Observational Designs. Washington, DC: American Educational Research Association;2007.
37. Sox HC, Goodman SN. The methods of comparative effectiveness research. *Annual review of public health.* Apr 2012;33:425-445.
38. Rassen JA, Solomon DH, Curtis JR, Herrinton L, Schneeweiss S. Privacy-maintaining propensity score-based pooling of multiple databases applied to a study of biologics. *Med Care.* Jun 2010;48(6 Suppl):S83-89.
39. Kaplan SH, Billimek J, Sorkin DH, Ngo-Metzger Q, Greenfield S. Who can respond to treatment? Identifying patient characteristics related to heterogeneity of treatment effects. *Med Care.* Jun 2010;48(6 Suppl):S9-16.
40. Brookhart MA, Rassen JA, Wang PS, Dormuth C, Mogun H, Schneeweiss S. Evaluating the validity of an instrumental variable study of neuroleptics: can between-physician differences in prescribing patterns be used to estimate treatment effects? *Med Care.* Oct 2007;45(10 Supl 2):S116-122.
41. Maclure M, Carleton B, Schneeweiss S. Designed delays versus rigorous pragmatic trials: lower carat gold standards can produce relevant drug evaluations. *Med Care.* Oct 2007;45(10 Supl 2):S44-49.
42. Kahn, op. cit.
43. Observational Medical Outcomes Partnership (OMOP). 2012; <http://omop.fnih.org/>. Accessed October 20, 2012.
44. Gartlehner, op. cit.
45. Glynn RJ, Brookhart MA, Stedman M, Avorn J, Solomon DH. Design of cluster-randomized trials of quality improvement interventions aimed at medical care providers. *Med Care.* Oct 2007;45(10 Supl 2):S38-43.
46. Mazor KM, Sabin JE, Boudreau D, et al. Cluster randomized trials: opportunities and barriers identified by leaders of eight health plans. *Med Care.* Oct 2007;45(10 Supl 2):S29-37.
47. Zurovac J and Brown R. Orthogonal Design: A Powerful Method for Comparative Effectiveness Research with Multiple Interventions. Center on Healthcare Effectiveness, Mathematica Policy Research, April 2012.
48. Fiore LD, Brophy M, Ferguson RE, et al. A point-of-care clinical trial comparing insulin administered using a sliding scale versus a weight-based regimen. *Clin Trials.* Apr 2011;8(2):183-195.
49. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *Am J Prev Med.* 2007;32 (5 Suppl):S112–S118.
50. U.S. Preventive Services Task Force . U.S. Preventive Services Task Force (USPSTF) procedure manual, Section 5, methods for arriving at a recommendation. Rockville, MD: AHRQ;1999. Available at <http://www.uspreventiveservicestaskforce.org/uspstf08/methods/procmmanual5.htm>.
51. Sawaya, G.F., Guirguis-Blake, J., LeFevre, M., et al. Update on the Methods: estimating certainty and magnitude of net benefit. *Ann Intern Med.* 2007;147:871-875.
52. The Task Force prefers large, well-conducted RCTs to determine the benefits and harms of preventive services. In many situations, however, such studies have not been or are not likely to be done. When these studies can be done, and other evidence is insufficient to determine benefits and/or harms, the Task Force advocates that large, well-conducted RCTs be done. It notes that small, poorly-conducted RCTs are often of little use. Multiple, large, well-conducted observational studies with consistent results showing a large effect size that does not change markedly with adjustment for multiple potential confounders may also be judged sufficient to determine the magnitude of benefits and harms of a preventive service. Large, well-conducted observational studies can provide essential additional evidence even in situations where there are adequate RCTs. For more information see: <http://www.uspreventiveservicestaskforce.org/uspstf08/methods/procmmanual4.htm>
53. The USPSTF has provided useful criteria for evaluating the internal validity of specific study designs. These may vary depending on the study design, but generally include adequate sample sizes, maintenance of comparison groups, consideration of relevant outcomes, and adjustment for confounders. The USPSTF considers diagnostic accuracy separately since RCTs are not frequently used to determine the accuracy of a test.

54. U.S. Preventive Services Task Force. U.S. Preventive Services Task Force (USPSTF) procedure manual, Appendix VII, criteria for assessing internal validity of individual studies. Rockville, MD: AHRQ;1999. Available at: <http://www.ahrq.gov/clinic/uspstf08/methods/procmanualap7.htm>.
55. U.S. Preventive Services Task Force . U.S. Preventive Services Task Force (USPSTF) procedure manual, Appendix VIII, criteria for assessing external validity (generalizability) of individual studies. Rockville, MD: AHRQ;1999. Available at: <http://www.ahrq.gov/clinic/uspstf08/methods/procmanualap7.htm>.<http://www.uspreventiveservicestaskforce.org/uspstf08/methods/procmanualap8.htm>
56. Ibid.
57. Teutsch SM, Bradley LA, Palomaki GE, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: methods of the EGAPP Working Group. *Genetics in Medicine*. January 2009;11(1):3-14
58. Washington State Health Care Authority: Health Technology Assessment. 2012; http://www.hta.hca.wa.gov/tech_process.html. Accessed December 13, 2012.
59. Blue Cross Blue Shield Association: Technology Evaluation Center Criteria. 2012; <http://www.bcbs.com/blueresources/tec/tec-criteria.html>. Accessed December 13
60. United HealthCare Online: Medical & Drug Policies and Coverage Determination Guidelines. 2012; https://www.unitedhealthcareonline.com/b2c/CmaAction.do?channelId=c94e74bc62c010VgnVCM100000c520720a_____. Accessed December 13, 2012.
61. EDM Forum, op. cit.
62. OMOP, op. cit.
63. FAIR Health: About the Program. 2012; <http://www.fairhealth.org> . Accessed October 22, 2012.
64. The New York City A1C Registry. 2012; <http://www.nyc.gov/html/doh/html/diabetes/diabetes-nycar.shtml>. Accessed October 20, 2012.
65. US Department of Defense. Technology NCFa. DOD, VA Release Mobile App Targeting Post-traumatic Stress. 2012; <http://www.defense.gov/news/newsarticle.aspx?id=117339>. Accessed November 4, 2012.

Appendix I: Sources of Electronic Clinical Data

ECD is available from many sources including electronic health records (EHRs); administrative claims; operational or community-level data, such as vital statistics; and patient reported outcome (PRO) surveys and assessments collected for research studies and clinical care. These types of data may be aggregated by a variety of organizations such as: payers (both government programs and private insurers); providers, ranging from primary or ambulatory clinics, specialty clinics, hospitals or group practices to integrated health systems, and academic medical centers; public health entities such as health departments and hospital associations; community-based organizations, including community health centers; and research organizations.

Several initiatives designed to improve research on comparative effectiveness of treatments and care delivery focused on ECD that is linked across various sites and sources. Among these are a series of eleven projects funded through the Agency for Healthcare Research and Quality (AHRQ) to further develop the infrastructure and methods needed to collect and use ECD for CER, PCOR, and QI.⁶²

The AHRQ projects cover a broad range of issues and data types. For example, the Population-Based Effectiveness in Asthma and Lung Disease (PEAL) Network is linking data from Medicaid and health plan populations in several states to create to compare adherence to - and effectiveness of- different asthma control regimes. The Indiana PROSPECT study is incorporating multiple types of data including patient-centered outcomes and genomic information into a health information exchange that will facilitate research on medications to treatment the symptoms of Alzheimer's.

Large-scale efforts to work across organizations to leverage research expertise and broaden engagement with various stakeholders are also expanding in light of the potential to use ECD for CER and PCOR. For example, the Observational Medical Outcomes Partnership (OMOP), is testing the ability to conduct drug surveillance by using data (both claims and electronic clinical data) from heterogeneous sources including EHRs.⁶³ The National Private Insurance Claims Database⁵¹, which merges data from a variety of care settings and payers allows the possibility of following patients over time even if their source of health coverage and providers change, is another example. Public Health agencies such as New York City's Department of Mental Health and Hygiene are merging payment and provider data to more effectively monitor the health of populations (one example is the New York City A1C Registry⁶⁴) and construct comprehensive dataset based on diseases, exposures or patient subgroups.

Technologies are also increasingly used to collect data directly from patients, allow patients and clinicians to follow individual trends in treatment, and aggregate data across patients for CER. Developers in the mHealth – or “mobile health” – space are actively building the tools needed so that data from mobile devices is meaningful for clinical care. One example of these types of products includes a mobile application created with the Veterans Administration to support collection of patient-generated data for use in clinical care for Post –Traumatic Stress Disorder (PTSD).⁶⁵

These are promising examples that seek to elevate the re-use of electronic clinical data for discovery and quality improvement. In addition to the brief examples provided in this section, additional links and references are provided in the recommended resources at the end of the brief. Further examples on networks, study designs, and informatics strategies are also provided.