Comparative effectiveness research: Policy context, methods development and research infrastructure‡

Sean R. Tunis, a∗† Joshua Benner b and Mark McClellan b

Comparative effectiveness research (CER) has received substantial attention as a potential approach for improving health outcomes while lowering costs of care, and for improving the relevance and quality of clinical and health services research. The Institute of Medicine defines CER as ‘the conduct and synthesis of systematic research comparing different interventions and strategies to prevent, diagnose, treat, and monitor health conditions. The purpose of this research is to inform patients, providers, and decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances.’ Improving the methods and infrastructure for CER will require sustained attention to the following issues: (1) Meaningful involvement of patients, consumers, clinicians, payers, and policymakers in key phases of CER study design and implementation; (2) Development of methodological ‘best practices’ for the design of CER studies that reflect decision-maker needs and balance internal validity with relevance, feasibility and timeliness; and (3) Improvements in research infrastructure to enhance the validity and efficiency with which CER studies are implemented. The approach to addressing each of these issues should be informed by the understanding that the primary purpose of CER is to help health care decision makers make informed clinical and health policy decisions. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: clinical research; research methods; research infrastructure

1. The need for more and better evidence on what works best

In the debate about health-care reform, one area of agreement is the need to address the gaps in quality and efficiency in health care in the United States. While America leads the world in many areas of health-care innovation, it lags behind many developed nations in important health outcomes like mortality rates for conditions amenable to medical care [1] and has much higher health-care costs. Spending on health care will consume approximately 18 per cent of GDP in 2009, or $2.5 trillion—and at current rates of growth, health care will exceed one-fourth of GDP by 2025 [2, 3]. The Congressional Budget Office has determined that health-care costs represent the single greatest challenge to balancing the federal budget [4].

Policymakers are hopeful that health-care spending growth can be reduced to a more sustainable level while improving access to care that improves health. Evidence on the variations in medical care across geographic regions in the United States suggests that as much as 30 per cent of spending reflects medical care of uncertain or questionable value. Overall, the Institute of Medicine has estimated that less than 50 per cent of treatments delivered today are supported by evidence [5]. A major reason for the gap is limited investment and capacity to develop relevant and reliable evidence about which interventions work best for whom—in other words, comparative effectiveness research (CER). Of the nation’s more than $2 trillion annual health expenditure, less than 0.1 per cent is invested in assessing the comparative effectiveness of available interventions [6, 7]. Developing better evidence on which interventions work best is one essential element for meaningful health-care reform.

The American Recovery and Reinvestment Act of 2009 (ARRA) provided $1.1 billion to the Agency for Healthcare Research and Quality (AHRQ), the National Institutes of Health (NIH), and the Secretary of the Department of Health and Human Services (HHS) to support the development and dissemination of evidence on CER. The law also provided

‡ This article was published online on 16 June 2010. An error was subsequently identified. This notice is included in the online and print versions to indicate that both have been corrected [28 July 2010].
for the formation of a Federal Coordinating Council on CER to help direct and coordinate the use of the new funding.

It recently released its definition of CER [8]:

Comparative effectiveness research is the conduct and synthesis of systematic research comparing different interventions and strategies to prevent, diagnose, treat and monitor health conditions. The purpose of this research is to inform patients, providers, and decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances. To provide this information, comparative effectiveness research must assess a comprehensive array of health-related outcomes for diverse patient populations.

In June, the Institute of Medicine released an initial list of ‘national priorities’ for CER, which included both patient-level and health system-level interventions [9]. The Director of the Office of Management and Budget, Peter Orszag, has noted that better evidence is a key step in reducing cost growth and achieving a more sustainable health-care system [10].

The positive expectations of CER may be achievable if the research provides credible, relevant, and timely evidence for individual patient-care decisions, and for policy decisions that affect how care is delivered to populations.

In the health reform bill signed by President Obama in March 2010, a new public-private Agency named the Patient-Centered Outcomes Research Institute (PCORI) was created to support and oversee the conduct of CER. By 2014, funding for PCORI will reach approximately $470 million per year, highlighting the increased urgency of developing a robust framework to guide investments in the methods and infrastructure required to conduct this research.

The future utility of CER depends on how often it will be possible to apply convincing data and methods, to the right question, at the right time. CER can be performed using a broad range of established and emerging methods, which generally fall into five major categories (Table I). The first two categories can be thought of as secondary CER, because they utilize existing studies to reach their conclusions. The methods for primary CER include retrospective and prospective observational studies, and experimental studies including randomized trials.

This paper provides an initial framework for improving research methods and infrastructure for primary CER. Experimental studies are a crucial source of CER information, and for questions that can be addressed effectively with these methods, it is critically important to develop study designs and infrastructure that will generate credible and relevant information, as quickly, efficiently, and inexpensively as possible. Non-experimental approaches using observational data are also a useful tool for CER, and could become increasingly important as such methods are further refined. There have been important advances in the design and use of clinical registries, significant technical advances (and increased funding) that will exponentially increase the availability of encounter-generated data (claims and electronic medical records), and accompanying efforts to improve the validity of methods used to analyze observational data [11–13].

There is widespread agreement, however, that current methods for primary CER often fail to provide sufficient credible, relevant, and timely evidence. Well-designed randomized clinical trials can be extremely costly, even with limited sample sizes and follow-up. Such studies are often criticized as inadequate for decision-making because of limited sample size, insufficient duration of follow-up, or lack of relevance due to their focus on already outdated medical technology. Conversely, observational databases can increasingly provide much larger sample sizes and perhaps more timely analysis, but studies using these databases are not randomized and have been criticized for producing potentially biased results. Further, differences in medical technology, both across areas and over time, have led to concerns about the validity of inferences from particular non-experimental studies.

Consequently, developing useful evidence from CER on a wide enough range of health-care interventions to have a substantial impact on care is a very challenging task. It will require identification of and investment in improved research methods, as well as a much better capacity for conducting CER studies and acting appropriately on the results. Improving the methods and infrastructure to fulfill the promise of CER will require sustained attention to the following issues:

- Meaningful involvement of patients, consumers, clinicians, payers, and policymakers in key phases of CER study development and implementation.
• Development of methodological ‘best practices’ for the design of CER studies that reflect decision-maker needs and balance internal validity with feasibility and timeliness, particularly with respect to the use of observational data and data from studies that do not rely on traditional randomization methods.
• Improvements in research infrastructure to enhance the validity and efficiency with which CER studies are implemented.

The approach to addressing each of these issues is informed by the understanding that the primary purpose of CER is to help health-care decision-makers make informed decisions—at the level of individual care for patients and clinicians, and at the level of policy determinations for payers and other policymakers. An important corollary of this observation is that, in order to be useful for decision-making, the evidence generated through CER must be valid, relevant, timely, feasible, and actionable. In order to balance all of those considerations, it will be necessary to go beyond the current approaches to conducting clinical and health services research; this research should not be designed and implemented within the traditional research community alone. The primacy of the needs of health-care decision-makers at multiple levels has important implications for collaboration, research methods, and infrastructure needed for CER.

2. Meaningful involvement of decision-makers

Perhaps the most common failure of much of the clinical and health services research done in the past—and a potential explanation for the extensive gaps in practical knowledge about health-care interventions—is the lack of sustained, meaningful engagement of health-care decision-makers in the design and implementation phases of CER studies [14]. To fulfill the objectives of CER, new strategies will be required to support highly diverse, multi-disciplinary collaborative working groups for CER projects. Better informational and technical capabilities for collaborative projects are making these strategies increasingly feasible.

Recommendation 1: The AHRQ should conduct a systematic assessment of best practices for effective engagement of decision-makers during various stages of clinical and health services research, including in priority-setting, protocol development, study implementation, and dissemination.

Recommendation 2: As a condition of receiving federal funding for any CER study, the investigators must form a stakeholder advisory committee composed of individuals who represent groups directly affected by the research, and whose specific functions should be determined based on the findings of the review in Recommendation 1.

Among the elements of the CER process that will require effective dialogue and consensus are:

• Selecting and prioritizing important research questions.
• Refining research hypotheses and arriving at the specific questions to be addressed.
• Feedback on specific elements of draft study protocols, including patient inclusion criteria, outcomes of interest, and methods. This will include advice on how to ensure that important patient subgroups are analyzed.
• Techniques to enhance enrollment of patients and clinical investigators in the trials.
• Use and protection of patient-level information in administrative and clinical databases when used for research.
• Strategies for effective dissemination of the results.

Categories of participants to be included in these CER working groups include:

• Patients and consumers (representative of the general public).
• Practicing clinicians.
• Medical professional organizations.
• Evidence review groups (The Cochrane Collaboration, AHRQ’s Evidence-based Practice Centers).
• Federal Agencies (AHRQ, NIH, FDA, Centers for Disease Control and Prevention, Veterans Affairs, Department of Defense).
• Public and private payers/purchasers (CMS, Medicaid, VA, Wellpoint, Blue Cross and Blue Shield health plans, employers).
• Life sciences industry (drugs, devices, other products and services).
• Representatives of a study’s research team.

Particularly important is fully understanding and incorporating the perspectives of patients and consumers in the clinically and technically complex discussions that take place with respect to the design of health research. Existing models of collaborative clinical and health services research can be used to identify best practices and lessons learned, resulting in a template for successful approaches that could be adopted for future CER projects. Some insights into effectively engaging patients and consumers can be gleaned from the work of the National Breast Cancer Coalition, the HIV/AIDS community, the Juvenile Diabetes Research Foundation, Consumers United for Evidence-based Medicine, and the Citizens Council of the National Institute of Health and Clinical Excellence (NICE) [15].
Statistics in Medicine

S. R. TUNIS, J. BENNER AND M. MCCLELLAN

There does not appear to be any published systematic assessment of the best practices identified by these organizations with respect to the engagement of patients and consumers, nor is there a body of the literature on techniques for effectively engaging other types of decision-makers in clinical or health services research. Even in the absence of fully developed functional models for collaborations, all CER projects should be able to clearly describe the mechanisms that are used to ensure that relevant expert, stakeholder, and decision-maker perspectives are adequately considered in the design, implementation, and dissemination of their work. CER stakeholder advisory committees are one mechanism to accomplish this objective.

Members of the advisory committee should be determined based on the specific health intervention under study and should include representatives of those groups most directly affected by the results. For a decision regarding initial treatment of prostate cancer, for example, the advisory committee might include a representative from a prostate cancer patient advocacy group, a prostate cancer survivor, a representative from a professional society of urologists, the medical director from a private health plan, and the author of a recent review on the clinical effectiveness of existing strategies. Members of the advisory committee are not members of the investigator team, but the committee’s recommendations on each phase of the study outlined above should be documented by the investigators in all study reports and publications.

3. Developing methodological guidance for CER

A crucial requirement of effective CER will be to employ the best possible analytic methods and data in studies of clinical and health policy questions. Ideally, evidence on the question of interest would be available from randomized controlled trials, with patients (and providers or organizations when appropriate) randomized to the alternative treatments, practices, or policies of interest. If randomized controlled trials sit atop the conventional hierarchy of evidence, expert opinion is the least desirable source, generally inferior to non-randomized studies, which may be biased but provide empirical evidence. With the emergence of new questions, new data sources, and improvements on methodologies, more detailed technical evaluation of the suitability of data and methods is possible and necessary.

The decade-long struggle of public and private payers, both in the U.S. and abroad, to make evidence-based policy decisions on the use of molecular imaging (primarily FDG-PET scanning) in oncology exemplifies the problems resulting from the absence of a well-defined and broadly accepted evidentiary framework for conducting CER. PET scans produce images that reflect the metabolic activity of internal structures, and because many cancers are highly active metabolically, there has been great enthusiasm for the use of this technology in patients with cancer. Despite the publication of hundreds of clinical studies on various diagnostic uses of PET, systematic reviews continue to observe that the available evidence of clinical utility is limited or poor quality for many common clinical uses. Over the past several years, Medicare has provided coverage for tens of thousands of PET scans in the context of a national PET registry intended to fill this gap [16], but some analysts argue that such observational data are of limited value in assessing the diagnostic value of imaging technologies [17]. More recently, the National Oncologic PET Registry has been modified to include linkages to Medicare claims data with the intent of monitoring longitudinal outcomes, though questions remain about the validity and credibility of evidence generated through that approach [18]. Because of the limited evidence, the government of Ontario has implemented a conditional reimbursement program for PET scanning, devoting considerable analytic and political resources to conducting randomized trials on the clinical utility of PET in the management of oncology patients [19].

All of this work is taking place in the absence of any common understanding among researchers, decision-makers, and other stakeholders about which evidence can be developed and applied on the clinical utility of FDG-PET for management of oncology patients. Some continue to advocate for RCTs, and others for more sophisticated registries. Strongly divergent views exist about whether it is reasonable to expect compelling, direct evidence that diagnostic technology improves patients’ health outcomes, while others note that the absence of such information makes it impossible to determine whether imaging is justifiable. If measuring health outcomes is necessary, can one reliably derive such information from Medicare or other claims data? Further, even with outcomes information, in the absence of large randomized studies, can the causal impact of this diagnostic technology on outcomes be elucidated?

The lack of a well-defined framework to guide the design of CER studies is equally problematic in other important clinical areas, including cardiac imaging, genetic testing, radiation therapy for cancer, treatments for chronic wounds, complementary and alternative medicine, and disease management programs. In each of these areas, intense debate exists about which methods will yield evidence that is credible and relevant for decision-makers. For the most part, there has been limited systematic effort to reconcile or align the competing views on what constitutes acceptable CER evidence, therefore research activities are generally guided by what is possible, rather than what could be done to generate more valid, relevant, and timely evidence.
3.1. Principles of a methodological framework for CER

3.1.1. Traditional ‘hierarchies of evidence’ are overly simplistic and should not necessarily guide the implementation of CER. The need to think more broadly about developing evidence was described in detail in a recent paper by Sir Michael Rawlins, Chairman of the NICE in the United Kingdom [20]. Hierarchies of evidence should be replaced by accepting—indeed embracing—a diversity of approaches. This is not a plea to abandon RCTs and replace them with observational studies. Nor is it a claim that the Bayesian approaches to the design and analysis of experimental and non-experimental data should supplant all other statistical methods. Rather, it is a plea to investigators to continue to develop and improve their methods; to decision-makers to avoid adopting entrenched positions about the nature of evidence; and for both to accept that the interpretation of evidence requires judgment.

Rawlins correctly points to the need for a more cognitive approach to evidence-based policy making, and his comments mark a broader shift in thinking that recognizes the importance of both experimental and non-experimental methods in CER (Table II). Because knowing that an intervention works under ideal circumstances (efficacy) is necessary but not sufficient for evaluating what is appropriate for patients in real-world practice settings, answering CER questions will require a more nuanced approach to the generation and appraisal of evidence than is reflected in the widely used linear evidence hierarchy [21].

3.1.2. The right approach to a given CER study depends on the circumstances. Rather than a simple hierarchy that governs the selection of CER methods, a more useful framework would enable researchers to apply the best possible data and methods to a given CER question, once that question has been determined to be important to address. This would require consideration of the potential classes of CER questions, and then the strengths and limitations of available data and methods with respect to the different types of questions. The resulting guidance to investigators and funders will accordingly resemble a decision tree more than a pyramid.

Contextual considerations should be used to guide the choice of research methods and data for a given CER question, including: (1) the specific decision to which the evidence will be applied; (2) the nature of the interventions involved (whether it involves a treatment, strategy, or policy); (3) consequences of making the wrong choice; and (4) feasibility of experimental versus observational approaches. Feasibility considerations in observational studies further include the availability of relevant covariate information and/or instrumental variables or ‘natural experiments’ affecting the relevant patient population.

4. Selection and improvement of CER methods

Consistent with the principles above, two forms of guidance to CER investigators are needed, in addition to continued investment in methodological research to enhance CER methods. This methodological guidance and innovation will enable the CER community to design studies that are accurately targeted to produce the information needed by patients, consumers, clinicians, payers, and policymakers. If done through effective multi-disciplinary consultation, the study design recommendations will reflect collective judgments about the inevitable trade-offs between internal validity, feasibility, timeliness, and generalizability.

Recommendation 3: At least 10 per cent of funds allocated to CER in the next 10 years should be directed to the Secretary of HHS for use in the development of methodological guidance and innovation. Funded programs should address the needs for:

- Objective reviews of the strengths and limitations of alternative methods for CER, including examples of their implementation and identification of categories of CER topics for which each is potentially appropriate. The Institute of Medicine or another organization capable of convening a broad range of methodologists, clinicians, and other stakeholders should develop a ‘translation table’ relating CER methods to specific types of research questions, as illustrated in Table II. This should build on existing work on methodological and reporting standards for each of the major CER methods. Once this is developed, it can provide a foundation for federal funding of CER to achieve an enhanced impact of CER on decision-making.
- Guidance documents should be developed for the consideration and selection of methods and data sources for specific priority CER topics. As lists of priorities for CER are produced by the Institute of Medicine and others, a process should be developed to identify appropriate research methods for each high-priority research question. This process would include methodologists and content experts, and would carefully consider the pros and cons of different CER methods, then make specific recommendations for study designs necessary to answer these questions. These general design recommendations would be included in the government requests for proposals. The investigators
### Table II. CER methods with examples matched to potential research questions.

<table>
<thead>
<tr>
<th>Description</th>
<th>Example</th>
<th>Advantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pragmatic clinical trials</strong></td>
<td>These RCTs are designed to demonstrate how a medical intervention works in a typical, real-world setting. Features of these trials can include all or a combination of the following: relaxed inclusion/exclusion criteria, relaxed protocol, longer term endpoints, active comparators, and outcome measures of relevance to patients, payers, and physicians [22]</td>
<td>Are newer types of antihypertensive agents, which are currently more costly to purchase on average, as good or better than diuretics in reducing coronary heart disease incidence and progression? [23] The ALLHAT study used patient relevant outcomes, had minimal inclusion criteria, and there was some flexibility in the dosing of the therapies. Because these trials are designed to meet the needs of decision-makers, the results tend to be more generalizable, the outcomes are useful to patients and physicians making tough clinical choices, and the trial maintains all or much of the scientific rigor of traditional RCTs.</td>
</tr>
<tr>
<td><strong>Cluster RCTs</strong></td>
<td>Groups of people are randomized to an intervention instead of randomizing individuals. These groups can be, for example, communities, regional payers, purchasers, delivery systems, clinics, etc. Individuals within a cluster will tend to resemble each other, which needs to be taken into account in the statistical analysis.</td>
<td>What is the comparative effectiveness of the American Cancer Society smoking cessation program versus the American Lung Association smoking cessation program? [24] In this study, different clinics adopted different smoking cessation programs. This approach is ideal for comparing alternative, established therapies with true equipoise, rather than new therapies, and common therapies rather than novel or high-profile therapies. Cluster RCTs can also provide a rigorous evaluation of the effectiveness of therapies in real-world settings, especially when consent by cluster is acceptable.</td>
</tr>
<tr>
<td><strong>Bayesian / Adaptive trials</strong></td>
<td>Unlike traditional RCTs, the Bayesian approach makes use of prior information on a medical intervention to estimate a prior distribution. This prior information is then combined with trial data to create a posterior distribution. Trial data can be analyzed frequently and compared with the prior information to inform the direction of the study.</td>
<td>What is the most effective treatment method for patients with a given biomarker profile? [25] The authors propose to adaptively randomize patients to one of four treatment groups with allocation based on prior accumulated data. The Bayesian approach incorporates prior information into the data analysis, which can lead to mid-course modifications to the trial design, and potentially avoid the need to start new separate trials to reflect knowledge gained during the course of the trial.</td>
</tr>
<tr>
<td><strong>N-of-1 trials</strong></td>
<td>N-of-1 trials are single event case studies to look at the effect of an intervention in an individual. Generally, there are two or more periods, alternating when the participant receives the therapy and one where he does not. This allows physicians to look for clinically meaningful differences in outcomes. [26] Multiple N-of-1 trials can be combined to estimate population effects</td>
<td>What is the optimal drug $x$ in patient $y$ that effectively balances drug’s efficacy with its side effects? A number of other study questions are described in a manuscript by Guyatt et al., 1990 [27]. In general, this design is best for chronic and relatively stable conditions. There is an emphasis on optimizing effectiveness for the individual rather than for a population of patients. This design allows researchers to obtain information on individual treatment response. This approach is useful for chronic conditions with readily assessable primary therapeutic effect.</td>
</tr>
<tr>
<td><strong>Delayed-design or ‘advance coverage’ trials</strong></td>
<td>Many variations exist. The most common version for this design is that participants are randomized either to receive the intervention from the start of the trial, or to have the intervention withheld for a pre-specified amount of time. By the end of the trial, both study groups have received the study intervention [28]</td>
<td>This trial design has been employed for several studies of neuroprotective treatments for Parkinson disease [29]. For example, what is the comparative effectiveness of early versus later initiation of rasagiline on progression of disability in patients with Parkinson disease? [30] All participants are eventually given the potentially beneficial medical intervention, which overcomes some of the ethical concerns raised by traditional RCTs, while maintaining a control group. This approach may be particularly useful for CER when applied to cluster RCTs.</td>
</tr>
</tbody>
</table>
who submit proposals would be free to propose any study design, but would be expected to provide a rationale for
deviating from the recommendations of the expert workgroup.

- Continued research and innovation to develop improved methods for experimental and non-experimental CER.

AHRQ should develop and update annually a list of priority needs for research and development on the methods
of CER, and award competitive grants for these studies.

4.1. Matching CER research questions to appropriate CER methods

The range of potential CER questions and the methods most suitable for answering them suggests that a goal of
this next phase of CER research should include developing more systematic evidence on best practices that could be
applied to the entire portfolio of CER methods, including systematic reviews, decision modeling, retrospective analysis,
non-experimental studies, and experimental studies.

A number of organizations have developed consensus reporting guidelines for various types of clinical and health
services research (CONSORT, MOOSE, GRACE, STROBE, etc.), and considerable work has been done to describe best
practices in the conduct of many of the major methods for CER [31]. These activities have been extremely valuable
in improving the quality and transparency of such research. To develop guidance for high-quality design, conduct and
reporting of CER, there is a need for a well-defined process for developing consensus about the appropriate use of each
of those methods in addressing specific types of CER questions. This consensus must include both methodologists and
decision-makers, since it is not simply a function of the quality of the research but also the adequacy of that research
to inform clinical and policy decisions.

Eventually, it will be useful to be able to complete a ‘translation table’ similar to the basic draft provided in Table
II, which includes a few proposed uses for some categories of CER trials, based on preliminary work done by several
members of an informal workgroup of methodologists known as the CER Innovation Collaborative (CER-IC) convened by
the Institute of Medicine’s Evidence-Based Medicine Roundtable. While the table represents initial thinking, a thorough
structured process involving experts familiar with each type of method will be required to provide useful methodological
guidance to the CER community. It is presented here in this preliminary form for illustrative purposes only, mainly to
highlight the fact that CER methods must be carefully chosen for each CER question, since these methods have inherent
properties that render them more or less well-suited to specific circumstances.

An increasingly important adjunct to RCTs in the context of CER, particularly for questions involving health care
policies and strategies for delivering care, will be the data collected during the delivery of and payment for health
care. These questions require data beyond those feasible to collect in RCTs, and thus raise concerns about bias. There
have been important advances in methods that improve the validity of analyses of non-experimental data [11, 12],
considerable progress in the design and use of clinical registries [32], and significant technical advances and funding
that will exponentially increase the availability of encounter-generated data (claims and electronic medical records) [13].
As methods and data repositories continue to develop, non-experimental CER is likely to increase in importance.

Consequently, a clearer understanding of appropriate methods for non-experimental effectiveness research is essential.
More specifically, methodological guidance can improve the rigor and internal validity of non-experimental studies. New
methods are being developed to improve the validity of clinical inferences from these sources [33]. Perhaps the greatest
challenge to using observational data sources for CER is decreasing systematic bias due to confounding. Confounders
distort the relationship between the exposure and outcome, and when they are observable within the data, straight-
forward methods exist (e.g. exclusion, stratification, matching, or regression) to achieve unbiased results. Therefore,
potentially important covariates (including those that clinicians think are most important to the decision under study—and
thus may be associated with outcomes) must be identified prospectively, before data sources or analytic methods are
selected [34].

Even when best practices for non-experimental studies are followed, many CER questions will not be adequately ad-
dressed by analyzing routinely collected data from large administrative databases or electronic medical records, due to
the inability to account for unmeasured confounders [35]. A well-known example of this is the widespread use of
hormone replacement therapy based on numerous, large, and consistent epidemiologic studies showing dramatic reduct-
ions in heart disease in women taking these medications. Subsequent clinical trials did not demonstrate cardiac benefits,
but many women may have been harmed by following the unequivocal clinical and policy recommendations resulting
from overconfidence in the non-experimental research on this topic [36]. A leading explanation put forward to explain
this discrepancy is the presence of selection bias and uncontrolled confounding in the results of observational studies.

Methods for conducting sensitivity analyses can reveal the potential impact of unmeasured confounders within a treat-
ment effect estimate [11]. Approaches to both the design and the analysis can be appropriate for addressing unmeasured
confounding. Promising methods include propensity scores [37], inverse probability weighting [38], instrumental vari-
bles [39, 40], restriction [12], and use of external information [41]. Quantitative measures comparing the effectiveness
and statistical efficiency of these methods may help illuminate best practices [42].
Because the various approaches described above will often yield different answers to the same question, principles and methods are also needed for combining findings from multiple sources. For example, evidence of the comparative effectiveness of a drug might include results from RCTs, matched observational studies, and instrumental variable analysis using differences in formularies to estimate incremental effects of exposure to the therapy. These methods can be expected to yield different answers even if they are all done correctly, because they estimate treatment effects differently. Decision-makers need to estimate the overall effect in a population that could potentially be treated, yet the quality and relevance of the evidence from these seemingly disparate analyses may not be obvious. Grading evidence and modeling population effects from multiple empirical studies are potential approaches to guiding appropriate interpretation of non-experimental CER.

4.2. Study design recommendations for specific categories of health-care interventions

The discussion above focused on the conceptual work needed to determine the appropriate uses of existing and emerging CER methods, with the goal of matching any given high-priority CER question with the particular methods that are most likely to generate useful evidence on that question. The categories of available methods primarily address alternative mechanisms by which patients are assigned to study groups, not more specific elements of study design that will sometimes vary in predictable ways depending on the technology or service that is the subject of the study.

The value of developing consensus regarding key features of CER study protocols can be illustrated with the example of treatments intended to improve healing of chronic wounds, such as pressure ulcers or diabetic ulcers. AHRQ has commissioned two systematic reviews of negative pressure wound therapy for chronic wounds, each of which identified significant limitations in the quality of the existing clinical studies. Among a number of recurring deficiencies were the decisions about which patients to include and exclude from the studies, the adequacy of care provided to the patients receiving standard care (i.e. choice of the comparator intervention), and the choice of primary outcomes. Furthermore, these and other reviews of wound care interventions generally exclude non-experimental studies because of concerns about the potential for important unmeasured baseline differences in treatment groups that might explain differences in reported outcomes. This is consistent with the FDA guidance document on treatments for chronic wounds, which also emphasizes the need for RCT designs in order to make reliable comparisons of alternative treatments [43].

The manufacturers of treatments for chronic wounds have different perspectives about what constitutes adequate evidence to demonstrate clinical benefit for regulatory or reimbursement purposes. Many public and private payers provide coverage for and spend considerable sums on treatments for chronic wounds that do not meet the evidentiary standards applied in the systematic reviews of these treatments. All of this creates an environment of tremendous uncertainty for researchers planning to design future CER studies for treatment of chronic wounds. Without some effort to systematically gather and integrate different perspectives on the design of studies for treatment of chronic wounds, future studies will reflect a range of different designs, perpetuating the current situation in which the overall quality of evidence is poor and not aligned with what decision-makers would most like to know.

Effective investment in CER studies of chronic wounds would be supported by agreement among experts, stakeholders, and decision-makers on best practices for designing these studies. Are there acceptable alternatives to traditional RCT methods for comparing interventions? Would adaptive designs be a potential option for greater efficiency and lower costs? Is it possible to better formalize the elements of standard care for purposes of the control-group intervention? Are there specific primary outcome measures that are preferable to others that might be used? For head-to-head comparative studies, are there alternatives to very large, blinded RCTs that would provide reasonable evidence on the comparative risks and benefits of competing forms of the same underlying intervention? Without better defined, shared principles on critical elements for CER research protocols, there is a high risk that the investment in CER studies for chronic wounds and other topics will not meet the requirements of internal validity, feasibility, timeliness, and relevance.

The same lack of agreement on best practices has hindered the collection of evidence on non-invasive imaging in the diagnosis of coronary artery disease (coronary CT angiography or CCTA). Several recent systematic reviews of CCTA noted limitations in the existing evidence, with most studies being single center trials enrolling high-risk patients and reporting only on diagnostic accuracy compared with invasive angiography. Medicare came to similar conclusions after reviewing the evidence in March 2008, and concluded that large, rigorous RCTs with measurement of hard clinical endpoints (such as cardiac death and acute myocardial infarction) would be necessary to provide adequate evidence of the clinical utility of coronary CT angiography [44]. Many individuals from the cardiology and imaging community firmly believe that the current evidence is sufficient to conclude that CCTA is an important advance in management of suspected coronary disease, and do not believe that additional studies are necessary [45]. The National Heart, Lung, and Blood Institute has recently funded a 10 000 patient, pragmatic RCT to study the clinical utility of CCTA in low and intermediate risk patients, a study that will provide critical information for clinicians and patients in about 5 years.
However, there remains significant controversy regarding the appropriate design of studies to evaluate new and existing non-invasive cardiac imaging technologies. Once again, investments in CER studies by public or private entities will be challenging without greater clarity on the set of acceptable designs for these studies.

Progress on CER could be expedited by a process for developing technology-specific methodological guidance on major categories of health services and strategies. The purpose of such a guidance would be to assist product developers and clinical researchers in designing CER studies that would provide patients, clinicians, and policymakers with a reasonable level of confidence about the relative effectiveness of the available health-care options. This guidance would need to be developed through a collaborative effort involving a broad range of experts and stakeholders and would develop specific recommendations for the design of CER studies that would aim to balance the validity, relevance, feasibility, and timeliness. Such a process would be necessary to allow for informed, collective judgments to be applied to decisions about which methods would provide sufficiently credible answers to important CER questions within specific classes of health interventions, while also considering the importance of relevance and timeliness in the process of making real-world, real-time decisions. A useful framework to guide the choice of methods in CER will require a level of specificity that can best be addressed by focusing on closely related groups of technologies for which decisions about study eligibility criteria, comparators, outcomes, and other study design elements may be similar.

It will inevitably be challenging to integrate the conflicting interests of the different stakeholders to successfully reach consensus on CER methods, given the significant implications of the alternative methods for patients, product developers, payers, and other stakeholders. However, there are potential advantages of increased consistency, transparency, and certainty for all stakeholders, which suggest that agreement on some common principles will eventually be achievable. Furthermore, it will be preferable to have these debates about methods before studies are conducted rather than once the investment of time and resources has already been made.

Intervention-specific evidentiary guidance would draw upon the methodological framework described above that matches various CER methods to the type of question for which those methods are and are not well-suited. These principles would be considered in the process of developing guidance for specific categories of technologies, since the type of research questions that will be important for those technologies can be matched to the various categories of available methods that may be appropriate. The guidance documents would provide greater specificity by seeking to create consensus not only on the mechanisms used for patient assignment to treatment and comparison groups but also on all other elements of the study design.

Based on part on the contents of a draft version of this paper provided to congressional staff in June 2009, the new CER institute created in the health reform legislation (PCORI) will include a methodology committee with responsibility for developing methodological guidance and standards for CER. The legislation states that these CER methods should balance validity with relevance, feasibility and timeliness, and should be that multiple experts and stakeholders should be involved in the guidance development process.

### 4.3. Work in progress to develop methodological guidance for CER

The Center for Medical Technology Policy has been working on early prototypes of collaboratively developed Effectiveness Guidance Documents (EGDs) to provide specific recommendations to product developers and clinical researchers about the design of clinical studies that will produce the evidence desired by patients, clinicians, and payers. Each EGD will focus on a specific category of health-care technology. For example, draft documents are under development for the treatment of chronic wounds, non-invasive cardiac imaging, and gene expression profiling for management of breast cancer. Methods for CER studies are also being developed through multi-disciplinary projects focused on the design of studies of new indications for approved cancer drugs, imaging in oncology, complementary and integrative medicine and joint replacement surgery (see www.cmtpnet.org).

The goal is to describe clinical studies that would provide decision-makers with a reasonable level of confidence that the technology improves health outcomes. In this respect, the guidance documents are intended to provide technology-specific methodological roadmaps for the design of prospective CER. For therapeutic interventions, the primary focus will be on the evidence of comparative clinical effectiveness, and for diagnostic interventions the primary focus will be on comparative clinical utility—how the diagnostic information affects clinical management and whether this leads to improved health outcomes.

EGDs are envisioned to be analogous to FDA guidance documents, which are also targeted primarily to product developers and clinical researchers, and which provide recommendations on the design of clinical studies that are intended to support regulatory decision-making. EGDs will serve a comparable function for product developers and clinical researchers, but are focused on the design of clinical studies to support ‘post-regulatory’ decision-making. These post-regulatory decisions include individual clinical decisions made by patients and consumers, clinical recommendations made by clinicians, clinical policies generated by medical professional societies, and reimbursement decisions made by payers. Since there is no single organization that represents the universe of post-regulatory decision-makers, neutral
forums that bring all of the relevant perspectives into a sustained dialogue are helpful for generating study design recommendations that align the information needs of decision-makers, the diverse interests of stakeholders, and the research activities of the CER community.

By including the relevant FDA regulatory experts in the EGD development process, it is hoped that EGDs will reflect optimal alignment between study design elements intended for regulatory approval and those targeted to clinical and health policy decision-making. This may help to avoid the need for multiple studies to address these different evidentiary purposes.

These methodological guidelines are being developed through a collaborative, multi-disciplinary process and provide specific recommendations for the design of CER studies that aim to balance internal validity with relevance, feasibility and timeliness. EGDs could be used by CER researchers conducting either publicly or privately funded research as one source of input during the process of protocol development. To the extent that they accurately reflect the information needs of patients, consumer, clinicians, payers, and policymakers, and to the extent that they successfully balance scientific and practical consideration, they should be a useful guide for designing CER studies.

4.4. Methodological research and innovation

Finally, efficient production of valid comparative effectiveness evidence over the long-term will depend on further refinements in the scientific methods. For those CER questions in which primary research is necessary, methodological research will be needed to develop and refine methods that are efficient, valid, generalizable, and relevant. Attention to methods research in these areas has been limited to date because the demand for these studies, and the funding available to support them, has only recently been identified as a public policy priority, mainly as part of the increased focus on CER.

A particularly important aspect of CER methods research will be the development of improved approaches to account for differences in treatment response in subgroups of patients enrolled in these studies. With the rapid scientific discovery of genetic and molecular markers for the development of disease and responsiveness to treatment, the evidentiary framework for CER will need to be further developed to produce information that is informative for homogeneous subgroups of patients as well as for individuals [46].

A number of investigators are now working on methods research that can better address heterogeneity in clinical trials, during both trial design and analysis of trial results [47]. The FDA is also aware of the increasing interest in methods to reliably analyze clinical trial data to assess subgroup effects, with a focus on determining the potential importance of biomarkers in predicting different responses to treatments, particularly in cancer.

The enthusiasm for determining subgroups in which the benefits or risks of treatment are markedly different from the average must be tempered by the recognition that such findings may or may not prove to be reliable. Perhaps the most compelling illustration of the potential for error was provided by Richard Peto, who performed a retrospective subgroup analysis of data from a very large trial of interventions for the treatment of acute myocardial infarction, and found that the benefits of immediate aspirin therapy—which is consensus standard of care—were not measurable for patients belonging to 2 of 12 astrological signs [48]. These results clearly reflect a statistical artifact, not a genuine clinical finding about these two arbitrary subgroups. This highlights that large sample size and statistical significance are insufficient to guarantee that results are accurate or clinically meaningful and that there is a need to recognize the methodological challenges of generating reliable information about subgroups simply by increasing the diversity of patients in CER studies. Useful information about subgroups will likely require much larger or more informative studies to ensure that the subgroups can be analyzed with reasonable statistical power and further refinement of methods to more accurately adjust for baseline differences in non-experimental studies. The ability of CER to enable more personalized medical decision-making ultimately depends on adequate investments in larger studies with appropriate data and better methods.
address the needs for:

- Ensuring that data standards created through the expansion of health information technology and deployed through electronic medical records are capable of supporting practice-based clinical research.
- Development of informatics grids and other architecture to link practice-based research networks (PBRNs), creating a national network with sufficient scale for conducting priority CER trials.
- Incentives for participation of investigators and patients in CER trials.
- Standard contract language for CER trials that use network infrastructure.
- Ethical guidance to institutional review boards that addresses human subjects protection issues commonly encountered in CER trials.

By virtue of its focus on the effectiveness of interventions in actual practice, rapid expansion of CER nationally will require the capacity to collect data efficiently from sites where care is delivered. Many observers believe that expanding the ability to conduct research in primary care and other practice environments will both improve the efficiency of research by adding to the numbers of investigators and potential human subjects, as well as pave the way for rapid application of results in the practice settings where research was conducted. In 1998, PBRNs were noted by the IOM to be ‘the most promising infrastructure development that [the committee] could find to support better science in primary care’ [49]. In 2006, the NIH released an Inventory and Evaluation of Clinical Research Networks (IECRN), which counted nearly 250 research organizations nationwide and identified 29 among their ‘best practice’ research networks. The IECRN recognized the stability of funding for PBRNs as a ‘pressing concern’ [50].

One goal of the NIH Clinical and Translational Science Awards (CTSA) program is to improve the effectiveness and efficiency of clinical trials in medical research centers. A national consortium, the CTSA now include 39 institutions in 23 states. By 2012 the program is expected to span approximately 60 CTSA with an annual budget of $500 million. One area of focus for the program is to speed the initiation of clinical studies by improving processes while controlling costs and reducing the time taken to complete protocol approvals by ethics committees and contract negotiations. Many stakeholders have observed that ethics approvals and contract negotiations are key bottlenecks in the existing clinical research system.

While these initiatives represent promising directions, the new demand for large-scale and coordinated CER adds urgency to the need for these infrastructure improvements. Moving toward a national network of practice-based and medical center-based investigators with the tools to conduct CER using best practices—and doing so efficiently—will require significant investments targeted to specific barriers.

5.2. Infrastructure for learning from the delivery of health care

Recommendation 5: At least 10 per cent of funds allocated to CER in the next 10 years should be directed to the Secretary of HHS for use in the development of infrastructure for learning from the delivery of health care. Funded programs should address the needs for:

- Distributed data networks for administrative and clinical databases—including Medicare and Medicaid data—and procedures for private-sector databases to be added to the network, procedures for investigator access to the network, and appropriate safeguards for ensuring the privacy and security of protected information.
- Technical data standards and a common vocabulary to be used by all linked systems. The necessity of these standards to support CER should be a high-priority spending consideration for the funds allocated to expand health information technology.
- Incentives for organizations with relevant data to adopt these standards and participate in research networks.
- Ethical guidance to HHS, other data owners, study sponsors, and investigators that balances the need for evidence to inform decisions with the need to safeguard personal health information.

For non-experimental studies, increasingly rich clinical and administrative data generated electronically through routine health care encounters can and will serve as a timely source of CER information. These data could be a valuable resource for evaluating treatments and outcomes across a broad range of patients, clinicians, and practice settings and for identifying heterogeneity in treatment effects between defined subgroups of patients. The methodological challenges should not be underestimated, but the combination of data sources with the right level of clinical richness can allow valid inference of many important comparative effectiveness questions. These data sources may also be used to provide longitudinal data to supplement data collected in clinical registries and clinical trials. Data networks under development include the Developing Evidence to Inform Decisions about Effectiveness network supported by AHRQ, and the Sentinel network for monitoring post-market safety, supported by the FDA. Patient-level administrative data from Medicare and Medicaid, the two largest health insurance programs in America, are currently not readily available to researchers.
Distributed data networks have the advantage of allowing those organizations with data to keep it behind firewalls (thus avoiding pooling) while still using it for collaborative research. This is done using a set of informatics ‘pipes’ that deliver standardized computer code that queries and analyzes each database in a similar manner. The results from each database can then be combined for statistically robust estimates of comparative effectiveness. Linking together a variety of databases in such a network adds not only to the quantity of patients who can be studied but also to the quality of studies that can be performed, if clinical and administrative data can be linked at the patient level. However, it is in the area of linkage that many persistent challenges and barriers remain. The informatics demands of data-sharing are perceived by most organizations as an additional burden, and thus far there has been only modest federal support for these efforts and no incentives provided to organizations to encourage participation. In addition to the technical challenges, proprietary concerns, as well as concerns about data privacy and security, are additional barriers to widespread use of these data sources [51].

Potentially valuable sources of linkable data include the large pools of claims from Medicare, Medicaid, and private insurers. Although claims data typically lack detailed clinical information, they have been used extensively for health services research and have generated important information about drug safety and effectiveness. Some examples of useful database combinations that would improve the validity of CER findings while preserving generalizability include (1) Medicare Part A, B, and D data with or without Medicaid and Minimum Data Set (MDS), which would allow the analysis of medical interventions and drugs in nursing home residents with detailed clinical information; and (2) Medicare A, B, and D data with the Medicare Current Beneficiary Survey, which would allow studies in elderly outpatients. It will also be important to develop a pathway for incorporating private databases, which can provide important additional detail, such as inpatient hospital records.

Finally, related to the need for methodological standards discussed earlier is the need for quality assurance in such a distributed data network. Because non-experimental research can be easily biased by the use of inferior methods and data to control for confounding, there is a risk that the credibility of the data network—and of CER in general—may be undermined by improper use. Thus, appropriate credentialing systems should be developed to assure effective utilization of such a data network.

6. Summary and conclusion

One of the most important consequences of the recent funding for CER has been to finally focus serious attention on what CER is, how it is different from other clinical and health services research, and what will be required to ensure that it can be conducted successfully. This paper identifies a number of issues related to CER methods and infrastructure that need to be addressed as the CER enterprise is expanded:

- Strategies and mechanisms that allow for the meaningful engagement of patients, consumers, clinicians, payers, and policymakers in key phases of CER must be developed and replicated.
- A clearly articulated framework will be necessary to ensure that high-priority CER questions are addressed with the methods that are most likely to provide meaningful, relevant evidence.
- Significant and sustained investment will be necessary to improve existing research methods for CER.
- Investment in the data collection infrastructure for experimental and non-experimental research is critical to building the capacity to conduct rigorous studies on a national scale.

While initial funding for CER should be devoted to addressing a broad range of high-priority clinical topics, a substantial portion of these initial funds should be directed to building the methodological framework and data collection infrastructure discussed in this paper.

Acknowledgements

This paper was adapted from Tunis SR, ‘Strategies to Improve Comparative Effectiveness Research Methods and Data Infrastructure,’ and from McClellan M and Benner JS, ‘Implementing Comparative Effectiveness Research: Priorities, Methods, and Impact.’ (c) 2009, The Brookings Institution. This publication was produced by The Engelberg Center for Health Care Reform at the Brookings Institution with support provided by The Hamilton Project. Used with permission.

The paper was substantially improved by the careful review and thoughtful feedback from a number of colleagues including Joe Selby, Rich Platt, Josh Ofman, Ellen Sigal, Rob Califf, Tim Taylor, Don Rubin, and Sebastian Schneeweiss. Finally, Danielle Whicher assisted with writing, reviewing, and editing sections of this manuscript.

References


