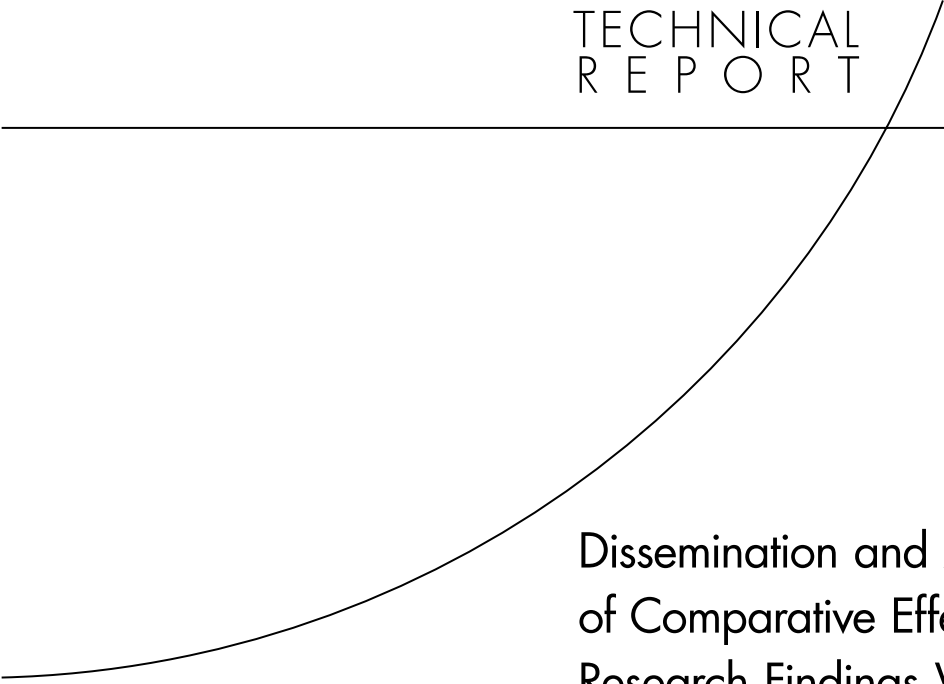


TECHNICAL
R E P O R T



Dissemination and Adoption
of Comparative Effectiveness
Research Findings When
Findings Challenge Current
Practices

Eric C. Schneider, Justin W. Timbie, D. Steven Fox,
Kristin R. Van Busum, John P. Caloyeras

Sponsored by the U.S. Department of Health and Human Services



The research described in this report was sponsored by the U.S. Department of Health and Human Services and was conducted in RAND Health, a division of the RAND Corporation.

The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

© Copyright 2011 RAND Corporation

Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Copies may not be duplicated for commercial purposes. Unauthorized posting of RAND documents to a non-RAND website is prohibited. RAND documents are protected under copyright law. For information on reprint and linking permissions, please visit the RAND permissions page (<http://www.rand.org/publications/permissions.html>).

Published 2011 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665

RAND URL: <http://www.rand.org/>

To order RAND documents or to obtain additional information, contact

Distribution Services: Telephone: (310) 451-7002;

Fax: (310) 451-6915; Email: order@rand.org

Preface

This report summarizes findings from a qualitative analysis of the factors that impede the translation of comparative effectiveness research (CER) into clinical practice and those that facilitate it. A case study methodology is used to explore the extent to which these factors led to changes in clinical practice following five recent key CER studies. The enabling factors and barriers to translation for each study are discussed, the root causes for the failure of translation common to the studies are synthesized, and policy options that may optimize the impact of future CER—particularly ongoing CER that was funded through the American Recovery and Reinvestment Act of 2009—are proposed. This report should be of interest to policymakers, including those in both federal agencies and nongovernmental entities involved with the generation or translation of CER research, as well as to public and private purchasers, clinicians, decision support companies, and patient-advocacy organizations.

This work was funded by the Office of the Assistant Secretary for Planning and Evaluation of the U.S. Department of Health and Human Services. The work was conducted in RAND Health, a division of the RAND Corporation. A profile of RAND Health, abstracts of its publications, and ordering information can be found at www.rand.org/health.

Contents

Preface iii

Figures and Tables xi

Executive Summary xi

Acknowledgments xxxvii

Abbreviations xxxix

CHAPTER ONE

Introduction and Summary of Approach 1

Defining Comparative Effectiveness Research 1

Federal Investment in Comparative-Effectiveness Research 2

Challenges of Translating CER Evidence into Practice 2

Objectives of the Project 3

Framework for Translation of CER into Practice 4

Case-Study Research Approach 8

Literature Review 12

Identifying Potential Discussants 13

Development of the Discussion Guide 13

Recruiting Discussants 14

Discussions with Stakeholders 16

Analysis of Qualitative Data 16

| | |
|--|----|
| Organization of This Report | 16 |
| CHAPTER TWO | |
| CATIE Case Study | 17 |
| Clinical-Practice Context | 17 |
| The Comparative-Effectiveness Question..... | 18 |
| Study Design Characteristics..... | 18 |
| Study Results..... | 18 |
| Impact of the Trial on Clinical Practice | 21 |
| New Evidence Following Initial Release of CATIE Results..... | 21 |
| Key Barriers to Clinical-Practice Change | 22 |
| Key Enablers of Clinical-Practice Change | 24 |
| Conclusions | 26 |
| CHAPTER THREE | |
| COURAGE Case-Study Report | 30 |
| Clinical-Practice Context | 30 |
| The Comparative-Effectiveness Question..... | 30 |
| Study Design Characteristics..... | 31 |
| Study Results..... | 31 |
| Impact of the Trial on Clinical Practice | 34 |
| Key Barriers to Clinical-Practice Change | 34 |
| Key Enablers of Clinical-Practice Change | 37 |
| Conclusions | 41 |
| CHAPTER FOUR | |
| SPORT Case-Study Report | 45 |
| Clinical-Practice Context | 45 |

| | |
|--|-----------|
| Comparative-Effectiveness Question..... | 46 |
| Study Design Characteristics..... | 46 |
| Study Results..... | 47 |
| Impact of the Trial on Clinical Practice | 49 |
| Key Barriers to Clinical-Practice Change | 50 |
| Key Enablers of Clinical-Practice Change | 54 |
| Conclusions | 54 |
| CHAPTER FIVE | |
| COMPANION Case-Study Report | 59 |
| Clinical-Practice Context | 59 |
| Comparative-Effectiveness Question..... | 60 |
| Study Design Characteristics..... | 60 |
| Study Results..... | 61 |
| Impact of the Trial on Clinical Practice | 62 |
| New Evidence Following Initial Release of COMPANION Results..... | 62 |
| Key Barriers to Clinical-Practice Change | 63 |
| Key Enablers of Clinical-Practice Change | 66 |
| Conclusions | 68 |
| CHAPTER SIX | |
| CPOE Case-Study Report | 73 |
| Clinical-Practice Context | 73 |
| Comparative-Effectiveness Question..... | 74 |
| Study Design Characteristics..... | 74 |
| Study Results..... | 74 |
| Additional Evidence Following the Initial Release of CPOE Results..... | 75 |

| | |
|--|----|
| Impact of the Trial on Clinical Practice | 76 |
| Key Barriers to Clinical-Practice Change | 76 |
| Key Enablers of Clinical-Practice Change | 78 |
| Conclusions | 81 |

CHAPTER SEVEN

Factors Influencing the Translation of CER Research into New Clinical Practices:

| | |
|--|-----------|
| A Synthesis of Themes from the Case Studies..... | 84 |
| Root Causes of Failure of CER to Change Clinical Practice..... | 84 |
| Limitations of This Study | 92 |

CHAPTER EIGHT

Policy Implications of the Case-Study Results..... 95

A CER-Enabling Infrastructure..... 96

Policy Options 96

Research and Evaluation..... 103

Concluding Remarks..... 105

References 106

Figures and Tables

Figures

| | |
|---|-----|
| S.1. Conceptual Framework for Translation of CER into Clinical Practice | xii |
| 1.1. Conceptual Framework for Translation of CER into Clinical Practice | 5 |

Tables

| | |
|--|------|
| S.1. Phases of the Translation of CER into Clinical Practice | xv |
| S.2. Case-Study Topics..... | xiii |
| 1.1. Phases of the Translation of CER into Clinical Practice..... | 5 |
| 1.2. Preliminary Case-Study Topics..... | 9 |
| 1.3. Final Case-Study Topics..... | 12 |
| 2.1. Results of the CATIE Trial | 20 |
| 2.2. Key Findings from the CATIE Case Study | 27 |
| 3.1. Results of the COURAGE Trial | 32 |
| 3.2. Key Findings from the COURAGE Case Study..... | 43 |
| 4.1. Results of the SPORT Trial..... | 48 |
| 4.2. Key Findings from the SPORT Case Study | 56 |
| 5.1. Results of the COMPANION Trial | 61 |
| 5.2. Key Findings from the COMPANION Case Study | 70 |
| 6.1. Results of the CPOE Study..... | 75 |
| 6.2. Key Findings from the Bates CPOE Case Study | 83 |

Executive Summary

Background

Insufficient evidence regarding the effectiveness of medical treatments has been identified as a key source of inefficiency in the U.S. healthcare system. Clinicians vary widely in their recommendation and use of diagnostic tests and treatments for patients with similar symptoms or conditions. This variation has been attributed to clinical uncertainty, since the published scientific evidence base does not provide adequate information to determine which treatments are most effective for patients with specific clinical needs.

A dramatic federal investment in comparative effectiveness research (CER) was made possible through the American Recovery and Reinvestment Act of 2009 (ARRA), with the expectation that the results will not only influence clinical practice but will also improve the efficiency of healthcare delivery. To do this, CER must provide information that supports fundamental changes in healthcare delivery and informs the choice of diagnostic and treatment strategies. Many new tests and treatments commonly adopted today are not completely grounded in scientific evidence. Some remain entrenched even when unambiguous scientific evidence about superior alternative approaches emerges. Other new clinical practices are not quickly adopted, either because information about them does not reach decisionmakers in a usable format or because of other barriers to their adoption.

Study Objectives

The project described in this report had three main objectives: (1) to develop a framework to help organize the array of barriers and enablers that influence the translation of CER evidence into new clinical practices; (2) to conduct case studies on the adoption of new clinical practices; and (3) to identify policy options that might facilitate dissemination of CER-based clinical practices.

We designed our organizational framework to isolate key factors affecting each phase of the process of CER translation, beginning with the generation of evidence and ending with the adoption of new clinical practices. The framework was also intended to inform CER development and dissemination activities, as well as future research on translation of CER into practice.

We conducted case studies on the adoption of new clinical practices following the release of five carefully selected CER studies published in the past 15 years, applying our framework to identify key themes relating to the pace of adoption of new practices. We sought information through discussions

with stakeholders representing a broad range of perspectives and by examining the peer-reviewed literature associated with each case study. Synthesizing common themes across case studies provided insight into the root causes for the failure of CER to change clinical practice in a timely manner.

We developed a set of policy options through consultation with an expert panel and with partners at the Office of the Assistant Secretary for Planning and Evaluation (ASPE) in the Department of Health and Human Services (HHS) that might facilitate dissemination of CER-based clinical practices and thus maximize the effectiveness of the federal government's current investment in CER.

The methodology we developed might also be used to inform larger-scale, prospective, in-depth qualitative and quantitative research on the impact of the federal investment in CER.

Methodology

Framework for CER Translation into Practice

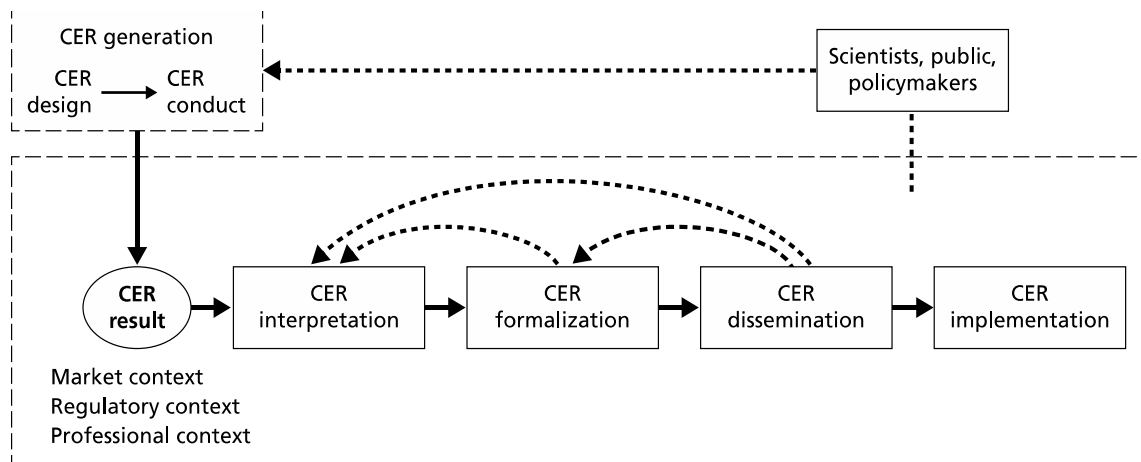
Our conceptual framework posits that the process of CER translation follows five key phases, shown in Figure S.1. While Figure S.1 suggests a generally linear temporal process, the phases are actually somewhat concurrent, and there appear to be multiple interactions between stakeholders at different phases. The phases are described in Table S.1.

Case-Study Research Approach

In selecting case studies, our preliminary intent was to identify and include CER trials that produced results that challenged current clinical practices. We used an environmental scan to develop a preliminary list of case-study topics, which we narrowed to five, based on a number of considerations. We wanted the topics to involve a high burden of illness, high prevalence, high-quality studies, diversity of treatment modalities, and diversity of treatment settings. The case-study topics chosen are shown in Table S.2.

Figure S.1

Conceptual Framework for Translation of CER into Clinical Practice



NOTES: CER dissemination and implementation phases involve dissemination and implementation of *clinical practices* as opposed to *results*. Solid lines represent the pathways through which CER research is translated into clinical practice; dotted lines are feedback loops that influence prior phases, indicating the interactive and iterative nature of the process.

RAND TR924-S.1

Table S.1

Phases of the Translation of CER into Clinical Practice

| Phase | Description |
|----------------|---|
| Generation | Generation includes the design and conduct of the CER study; it involves primarily funders and CER researchers, but research priorities are influenced by the needs of multiple stakeholders, including scientists, the public, and policymakers. |
| Interpretation | Stakeholders ascribe meaning to CER results based on a number of factors, including the strength of evidence, applicability of the evidence to the potential adopter’s practice setting, personal experience, and messages received by other stakeholders (e.g., professional societies, industry, media, and opinion leaders). |
| Formalization | Formalization is the process by which the interpretations of CER results are converted into guidance instruments such as clinical-practice guidelines, performance measures, and quality improvement tools. Multiple stakeholders may play roles in formalization through participation in guidelines committees, regulatory committees, and performance-measure development and endorsement processes. |
| Dissemination | Dissemination is the process by which CER information and/or associated tools designed to influence practice is actively transmitted to stakeholders. It typically promotes (or discourages) implementation of a new practice but may also have the goal of promoting a particular interpretation of the CER results. |
| Implementation | Implementation is the adoption of new clinical practices based on CER results. Implementation decisions may depend on a wide range of factors, including the dissemination of messages and the successful embedding of CER-related clinical guidance into tools that facilitate practice |

change, as well as the local market, regulatory, and professional context that may promote or impede changes. The implementation phase takes place primarily in local practice contexts.

Table S.2

Case-Study Topics

| Topic | Type of Comparison | Reference |
|---|-------------------------------|-----------------------------------|
| Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE): atypical antipsychotic drugs versus conventional antipsychotic drugs for schizophrenia | Medications | Lieberman, Stroup, et al., 2005 |
| Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation (COURAGE): percutaneous coronary intervention versus optimal medical therapy for chronic stable angina | Medication versus procedure | Boden, 2007 |
| Spine Patient Outcomes Research Trial (SPORT): surgical versus nonsurgical treatment for lumbar spinal stenosis | Procedures | Weinstein, Tosteson, et al., 2008 |
| Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure (COMPANION): optimal medical therapy versus cardiac resynchronization therapy versus combined cardiac resynchronization therapy and defibrillator therapy for patients with moderate to severe heart failure | Procedures | Bristow, Saxon, et al., 2004 |
| Computerized physician order entry (CPOE): interventions to prevent serious medication errors | Delivery-system interventions | Bates et al., 1998 |

After selecting the case-study topics, we examined the peer-reviewed literature to obtain information on the extent to which the CER evidence led to practice change, the key stakeholders engaged in translation, and the specific dissemination activities involved. We developed a preliminary list of potential discussants for each topic and extended invitations to a set of discussants who, taken together, could provide perspective on all phases of the CER translation process. We developed a core discussion guide that would enable us to address key topics in a consistent manner across case studies and conducted 53 discussions with individuals or groups, including researchers involved with the CER studies, practicing physicians, leaders of professional societies, representatives of funding agencies, patient advocates, decision support developers, directors of quality improvement organizations, senior executives of health plans, Medicaid directors, journal editors, and leaders of integrated health systems.

CATIE Case-Study Summary

Background

The National Institute of Mental Health (NIMH) funded the \$42.6 million CATIE study in 1999 to compare the effectiveness of a first-generation antipsychotic (perphenazine) to three second-generation medications: olanzapine, quetiapine, and risperidone. CATIE was considered a landmark trial because of its size, duration, and public sponsorship. Furthermore, it was designed to be generalizable to real-world clinical settings by using limited exclusion criteria, enrolling patients from diverse settings, and permitting flexible dosing protocols. Prior to CATIE, the optimal choice of drug treatments for patients with schizophrenia was disputed for at least four reasons: (1) uncertainties surrounding effectiveness in controlling psychotic symptoms; (2) uncertainty about the relative incidence of side effects, including tardive dyskinesia and metabolic side effects; (3) some evidence that second-generation antipsychotics improve cognition; and (4) the costliness of second-generation antipsychotics. These concerns prompted many to call for a rigorous assessment of the overall value of the second-generation antipsychotics.

Results

The initial results from CATIE were released in 2005, and they surprised many. The trial found that perphenazine was as effective as olanzapine in terms of time to discontinuation of medication for any cause, the trial's primary outcome. Patients randomized to olanzapine also had the longest time to discontinuation of any group because of lack of efficacy, the largest weight gain, and significant increases in other variables associated with the metabolic syndrome. The CATIE investigators concluded that perphenazine could not be rejected as an inferior treatment.

Lessons Learned

The CATIE case study highlighted the role of pharmaceutical manufacturers in shaping and reinforcing beliefs about the relative superiority of second-generation antipsychotics, both directly (through marketing and detailing) and indirectly (through key thought leaders) well in advance of the conduct of a CER study. By the time the CATIE results were released, these efforts had cemented beliefs about the various classes of antipsychotics. Practice patterns do not appear to have changed in the five years following the publication of the trial's results. Professional societies did not strongly advocate for practice changes based on the results. Guidelines eventually changed but had limited impact. Performance measures were not updated to reflect the trial's findings and may have continued to reinforce existing prescribing patterns. Professional societies and advocacy organizations challenged the results of the trial in an effort to protect provider autonomy and preserve access to medications, respectively. Public payers were initially unwilling to enact policies that might limit the treatment options of patients with schizophrenia, given the relative lack of access to care for this population and the potential backlash from advocacy organizations.

A number of strategies might be used in CER trials that share some of the characteristics of CATIE to promote the uptake of results into practice. With regard to CER generation, methodological choices (in CATIE, the exclusion of patients with tardive dyskinesia from the perphenazine group) may limit the perceived generalizability of the findings and cause physicians to distrust the results. In the CATIE case,

providers had strong prior negative experiences involving adverse outcomes such as tardive dyskinesia. Failure to design the trial to address one of the main beliefs driving use of a treatment—e.g., that second-generation medications were safer with respect to the incidence of tardive dyskinesia—meant that an issue important to prescribing providers might be perceived as being inadequately addressed by the study.

Interpretation and formalization faced formidable but predictable difficulties given the strength of established beliefs about the superior efficacy and safety of second-generation antipsychotics. Indeed, it proved quite difficult to change the deeply ingrained belief system founded on industry-funded studies. Interestingly, critiques of the study methodology that did not address harms may not have significantly influenced prescribing practices. Likewise, it should be borne in mind that professional societies can be expected to generate guidelines reflecting their professional interests if the study results leave room for such interpretations by failing to produce a “clear winner.” Timely updates to quality measures that reflect the new CER evidence may prove critical to motivating early changes in practice.

Dissemination and implementation strategies should be vigorous and multipronged. In the case of CATIE, academic detailing within closed systems eventually proved effective in some cases, but early efforts were constrained by doubts about the value of the findings, and even a clinical decision support prompt faced initial resistance. A key element to success was presenting physicians with their actual practice data, which often showed how far they diverged from the ideal. Finally, future adverse-event surveillance systems (or registries) may help to resolve lingering questions about the relative side-effect risks of alternative antipsychotics, such as tardive dyskinesia, relative to cardiovascular disease.

COURAGE Case-Study Summary

Background

The COURAGE trial compared the risk of cardiovascular events among patients with stable coronary artery disease (CAD) assigned to a treatment strategy of intensive pharmacologic therapy and lifestyle intervention (optimal medical therapy) alone and patients who were assigned treatment with percutaneous coronary intervention (PCI) followed by optimal medical therapy. No study before COURAGE had included the intensity of medical therapy attempted in the trial, which included the use of aspirin, beta-blockers, angiotensin-converting enzyme (ACE) inhibitors, statins, and clopidogrel, as well as diet, exercise, and smoking-cessation counseling. Medication doses were repeatedly intensified in pursuit of aggressive blood-pressure and LDL-cholesterol targets. The fundamental question addressed by COURAGE was whether the use of PCI to reverse narrowing of the coronary arteries in conjunction with optimal medical therapy would provide patients with stable CAD a greater reduction in the risk of myocardial infarction (MI) and death than medical therapy alone. While PCI has been shown to provide substantial benefit for patients being treated for emergency conditions, the benefit of the procedure in the stable CAD population had not been conclusively demonstrated. In the years preceding COURAGE, most clinical trials that assessed these end points were small and underpowered.

Results

The COURAGE trial found that as an initial management strategy in patients with stable CAD, PCI did not reduce the risk of death, MI, or other major cardiovascular events when added to optimal medical therapy. The findings reinforced existing practice guidelines, which stated that PCI can be safely deferred in patients with stable CAD, even in those with extensive, multivessel involvement and inducible ischemia, provided intensive, multifaceted medical therapy is instituted and maintained.

Lessons Learned

Both our discussions and a high-quality empirical analysis indicate that the COURAGE trial did not have an impact on clinical practice. The trial may have had an important indirect effect on practice by encouraging the integration of appropriateness criteria for coronary revascularization into decision support tools and into data collection for registries. How much these efforts will facilitate practice change remains unclear. Efforts to use appropriateness criteria in quality improvement are nascent, and while they have yet to be used in an accountability or payment context, there is increasing interest in them among policymakers. These initiatives will be most effective once reimbursement systems create demand for them. Changes in the organization of cardiology practices, driven in part by the movement toward accountable-care-organization (ACO) payment models, may be the single most important determinant of the future adoption of findings from COURAGE and other CER evidence.

Several strategies may improve uptake for CER trials that share some of the characteristics of COURAGE. In the generation phase, research should focus on a decision point sufficiently upstream to meaningfully impact decisionmaking. A critical driver of the use of PCI is the initial decision to refer a patient to an interventionist, since this tends to create an expectation that angiography and PCI will follow. The COURAGE trial did not address the initial referral decision directly. Rather, it addressed a decision point later in the pathway to PCI—after patients have undergone angiography—at which the utility of decision support and patient decisionmaking aids may be suboptimal. Current and proposed trials are focusing on decisions that occur prior to angiography, and these may have a greater impact on clinical practice. Other design problems to avoid include the potential for significant patient crossover or excessive time to complete the study. However, discussions with stakeholders suggest that criticisms of the trial design probably had only a minor influence on practice patterns post-COURAGE.

Interpretation and formalization can languish if study findings confirm current guidelines, even if they contradict current practice. Prior to COURAGE, practice guidelines were based on very weak evidence, promoting physicians' inclination to disregard them, but since the COURAGE results reinforced the guidelines, there was less impetus to revise them. A CER result that necessitates a change in guidelines may have more impact. Similarly, unless payers and other stakeholders have the ability to collect relevant appropriateness data, they will have no incentive to develop reimbursement policies based on guidelines or appropriateness criteria.

Dissemination and implementation may be either advanced or retarded by several factors, but in this case, psychological aspects appear vital. For example, while registries may have influenced practice (by incorporating performance measures and appropriateness criteria into their design), their influence to date on appropriate use of elective PCI appears modest. Similarly, payer limits on upstream diagnostic

procedures may have somewhat dampened demand for PCI, as might accountable-care reimbursement schemes in the future. Psychological factors, including concerns about harm and physician response to popular media coverage regarding PCI overuse, may have more significantly modulated the tendency to intervene aggressively. However, strong financial and psychological factors still incline both providers and patients to favor PCI. As one discussant put it, even without financial incentives, “inter-ventionalists love to intervene.” By all accounts, both clinicians and patients may underestimate the effectiveness of optimal medical therapy, and patients may not be informed of, fully understand, or seek out available information on the benefits and risks of PCI. Patient decision aids may play a key role in helping to address this information gap. However, to be maximally effective, such decision aids will have to be implemented in settings where financial incentives do not promote PCI and before patients have progressed to the point where intervention becomes inevitable.

SPORT Case-Study Summary

Background

The principal clinical question motivating SPORT was whether surgical treatment options were superior to nonsurgical treatment for patients with low back pain related to lumbar spinal disorders (disc herniation, spinal stenosis, and degenerative spondylolisthesis). Our case study focused only on the subpopulation with spinal stenosis. Prior to SPORT, the Maine Lumbar Spine Study, a prospective cohort study enrolling 148 patients, was the largest study comparing the effectiveness of alternative treatments for spinal stenosis, and it found that surgical patients had better outcomes than patients receiving nonsurgical treatments. A 2005 Cochrane review summarizing the evidence prior to 2000 suggested that the relative efficacy of surgery was not established, because existing trials were small and enrolled patients both with and without degenerative spondylolisthesis. A large, randomized CER trial was considered necessary to provide stronger evidence on the effectiveness of surgical treatment for spinal stenosis among patients who did not have degenerative spondylolisthesis.

Results

SPORT’s intention-to-treat analysis showed that surgery was more effective than non-operative treatment on the SF-36 bodily pain scale and on patients’ self-reported ratings of symptom improvement but on few other primary or secondary outcomes. However, patients with spinal stenosis had very high rates of crossover after randomization (as was the case for the subpopulations with disc herniation and degenerative spondylolisthesis). Only 67 percent of patients randomized to the surgical arm underwent surgery, while 43 percent of the patients randomized to nonsurgical treatment underwent surgery within two years of the baseline assessment. For this reason, the data from the randomized cohort were combined with data from an independent and concurrent cohort study and analyzed as a single observational study comparing patients who underwent surgery with those who did not (an “as-treated” analysis). The observational analysis found that surgery was superior to nonsurgical treatment across all primary and secondary outcomes, and the advantage was sustained over two years of follow-up.

Lessons Learned

SPORT appears to have had little impact on clinical practice, and the seeds of its low impact appear to have been sown primarily in the generation phase. The study design, which was unblinded, also allowed for very large patient crossover. As a result, what was intended to be a randomized controlled trial (RCT) with an “intention-to-treat” analysis had to also be treated as an observational cohort study using an as-treated analysis. Analogous studies have avoided these difficulties, suggesting that they are not inherent in this type of CER but can be forestalled by careful study design and execution.

In the interpretation phase, the RCT results suggesting limited benefits from surgery were discounted because of high rates of patient crossover. In contrast, the observational cohort study, at least within the spinal surgery community, confirmed the relative advantage of surgery, which was already the prevailing method of treatment. Interpretation was further complicated by the study’s lack of detail on subgroups, which made it hard to determine whom surgery would benefit most, as well as the (possibly erroneous) perception that the surgical techniques used in the study were already outdated. While presenting competing analyses may have opened the results to conflicting interpretation, the observational results alone produced different interpretations regarding the magnitude of the benefit provided by surgery.

The SPORT case study also highlights the challenges in weighing the relative strengths and weaknesses of RCTs and observational cohort studies and the selective use of evidence during the formalization phase. Multiple specialty societies, possibly influenced by various levels of industry sponsorship, issued competing and conflicting guidelines, while relevant data from European studies were generally discounted or ignored. Registries might help to bolster guidelines or generate appropriateness criteria, but since effectiveness outcomes from spinal surgery are often subjective, registries may be best suited to report on harms. Registry penetration appears quite low in orthopedic surgery, and financial incentives are not aligned to promote participation by surgeons.

While dissemination of the SPORT results appeared to be far-reaching, messaging about them emphasized the benefits of surgery rather than the significant clinical improvement among patients in the nonsurgical group and the relatively small difference in clinical benefit between the groups. Referring providers appear to be the optimal point for dissemination of the results, since referral to a surgeon is usually followed by surgery. Intense marketing of spinal hardware by the device industry may override the results of clinical trials, and, as SPORT illustrates, messages may be vague and selective, omitting key evidence provided by trials. Similarly, payers and purchasers, faced with both the “positive” results from the observational cohort analysis and the “equivalence” results from the intention-to-treat analysis, appear to have accepted the primacy of the observational cohort analyses and did not enact policies restricting the use of decompression surgery. However, there are now some early examples of more nuanced and data-driven reimbursement policies focusing on related procedures (e.g., fusion surgery).

Nevertheless, at the implementation phase, strong financial incentives favor surgical over nonsurgical treatment. The alignment of financial incentives among physicians, hospitals, and device manufacturers appears to have increased the use of complex procedures despite uncertainty about their effectiveness

and considerable evidence of greater risks. Countering this trend are radiology benefits managers (RBMs), which may reduce inappropriate upstream diagnostic procedures, and a potential future role for patient decision aids. While the SPORT results can be viewed as both flawed and confirmatory of current practice, the trial was successful in providing quality data on the relative risks and benefits of surgery, and these data have been integrated into patient decision aids. Those tools might ultimately change clinical practice by more fully incorporating patient preferences into decisions about surgery. Currently, few incentives encourage the use of such shared decisionmaking or a more rigorous informed-consent process. The use of these techniques early in the pathway leading to surgery will be critical to their overall effectiveness. Incentives to promote the spread of patient decision aids and efforts to improve the appropriate use of diagnostic imaging represent the most important strategies for changing clinical practice in the future.

COMPANION Case-Study Summary

Background

Cardiac resynchronization therapy (CRT) can improve the health status of patients with heart failure (HF) by electrically stimulating the heart to improve synchronization of pumping. Most HF patients appear to be at high risk for potentially fatal derangements in the heart's electrical activity. Implantable cardioverter defibrillators (ICDs) protect against sudden death from abnormal rhythms but do not reduce HF symptoms. The principal question addressed by the COMPANION trial was whether adding CRT alone or combined with ICD treatment (CRT versus CRT-D) to the medical management of HF patients with conduction abnormalities not only improved functional measures but also reduced hospitalizations and all-cause mortality (Bristow et al., 2004). Previous studies did not have sufficient power to detect a survival advantage from combined therapy.

Results

The COMPANION trial showed that patients assigned to the CRT group and the CRT-D group both had a statistically significant improvement of 17 percent over medical therapy alone in the combined end point of death and hospitalization from any cause. While adding an ICD to CRT did not appear to benefit patients more than CRT alone, it did show a trend toward reducing 12-month all-cause mortality (12 percent versus 15 percent). The results of this study imply a clear survival and quality-of-life benefit from adding CRT (with or without CRT-D) to optimal medical therapy for patients who suffer from HF with delayed ventricular conduction. This contrasted with the current practice at the time, which was to use CRT for HF patients but withhold ICD devices given both safety concerns and a lack of proven benefit.

Lessons Learned

Uptake of the CER results following publication of the COMPANION study has been uneven. Recent estimates indicate that there is both significant underuse of CRT among potentially eligible HF patients and also fairly frequent CRT-D use in patients who lack an indication for it.

In contrast to the other CER case studies, COMPANION generated relatively few controversies in the generation and interpretation phases. The results were fairly readily accepted, the main disputes being over the degree to which they could be generalized to HF patients who did not meet the original inclusion criteria. Formalization of the COMPANION results was relatively rapid: specialty-society guidelines were updated promptly, which promoted their uptake, at least among proceduralists and HF-management specialists. No primary-care specialty-society guidelines were issued. In addition, the specialty-society guidelines left open the appropriateness of CRT-D. This and other factors would in turn contribute to an ineffective dissemination phase.

This case study illustrates the critical role dissemination plays in translating CER research into practice. Essentially all dissemination activities focused on interventional cardiologists and HF specialists rather than referring physicians. Specialty societies, industry, and other continuing-medical-education (CME) producers all directed their educational efforts toward those groups. Most primary-care providers (who manage many HF patients) remain unaware of the COMPANION results. In addition, those primary-care providers and general cardiologists who took an interest in the study findings were confronted with conflicting and ambiguous guidelines. This generated considerable confusion and a reported reluctance to refer patients for CRT. Future similar CER dissemination should focus significant effort on providers further upstream in the decision pathway and on delivering clear, unambiguous referral criteria. However, the COMPANION case is not only a cautionary tale. HF registries have had a significant positive impact through publication of high-profile studies illustrating inappropriate ICD use. Similarly, recent limited experience with clinical decision support tools shows that they can be very effective in prompting appropriate referrals and discouraging inappropriate procedures, but such tools for CRT or CRT-D appear to be rare.

In the implementation phase, imprecise guidelines and evidence-neutral reimbursement policy may contribute to the use of CRT-D for inappropriate indications. Reimbursement policies, particularly those of Medicare, significantly favor CRT-D implantation over CRT alone, despite evidence that adding the ICD has a very high marginal cost relative to the benefits it confers. As with other studies, referral to an interventionist is also tantamount to ordering the procedure. This tendency is compounded by open guidelines that allow the CER results to be cited as justifying use in patients who would not meet study inclusion criteria. While primary-care physicians and some general cardiologists fail to refer many potentially eligible patients, dedicated HF clinics have been much more successful in achieving appropriate referrals, as well as avoiding inappropriate ones. Such clinics may serve as a model for implementing analogous CER results. Currently, patients are not generally equipped to participate as fully informed partners in the clinical decision, and decision aids are not readily available, but it is likely that such decision aids could significantly improve appropriate use of CRT if physicians were given stronger incentives to use them.

CPOE Case-Study Summary

Background

During the 1990s, experts debated the optimal approach to reducing medication errors. Some were not persuaded that traditional paper-based ordering systems were a significant problem or that computer-based ordering systems alone (e.g., for medications and lab tests) would be more effective in reducing the rate of medication errors than nurse-focused, pharmacist-focused, or team-based interventions. The principal CER question leading up to Bates's 1998 study was whether computerized physician order entry (CPOE) could reduce medication errors and medication-related adverse events among hospitalized patients more effectively than other interventions. To address this question, Bates and colleagues compared the effectiveness of CPOE alone with CPOE plus a team intervention for reducing the number of unintercepted serious medication errors. The study, conducted within six units at Brigham and Women's Hospital (Boston, Mass.), used a pre/post design, where the rates of medication errors prior to CPOE adoption were compared with the rates during the ten months following CPOE adoption. The CPOE system allowed physicians to select from a menu of medications defined by the hospital formulary, with default dose and dose ranges provided for each medication, as well as automatic checking for common drug allergies and drug-drug interactions. The team intervention centered on pharmacy-specific process changes, including changing the role of the pharmacist, standardizing labeling of intravenous bags, and implementing a pharmacy communication log so that the nursing staff could better communicate with pharmacy staff.

Results

The analysis indicated that the team intervention had no incremental benefit over the implementation of CPOE alone, so the intervention arms were pooled. Between pre and post periods in the same hospital units, unintercepted serious medication errors (the study's primary end point) decreased by 55 percent. Unintercepted potential adverse drug events (ADEs)—a secondary end point—declined by 84 percent. The authors of the study concluded that, on the basis of these results, other hospitals should consider CPOE adoption as the principal intervention to reduce unintercepted serious medication errors.

Lessons Learned

In contrast to the adoption of a new medication or device that does not require significant changes to the work of clinicians and staff but instead funnels through the existing workflow, the adoption of quality improvement strategies faces serious barriers because they may require significant changes to organization, financing, and staff work. These barriers may neutralize the impact of even outstanding CER evidence. While we selected CPOE as an example of a delivery-system intervention for which there is published CER evidence, it is worth noting that although CPOE has features that are typical of many such interventions, it also has distinct features that may be easier to implement. The CPOE case study suggests a number of lessons about translating delivery-system-related CER into new practices.

First, CPOE is both a new technology and a new set of workflow requirements. These features of the intervention are complex and require substantial up-front investment, as well as coordination, communication, and long-run commitments from numerous stakeholders with potentially conflicting

goals. The staff involved in implementation of CPOE must make nontrivial changes in workflow; like other technology-based delivery-system interventions, CPOE requires dramatic changes in individual process and social interactions with peers.

Second, CPOE is a variable technology with evolving features and functionalities. It has numerous meanings across a wide range of hospitals and different vendors, depending on their needs and existing health information technology (HIT) capabilities. This poses challenges for end users (particularly hospital executives) who wish to use CER evidence for decisionmaking. Our case study suggests that these individuals often struggle to conceptualize the intervention and consequently may find it difficult to assess the applicability of the results to their own settings.

Third, the financial investment in CPOE is substantial, and key leaders must have clear reasons and plans for implementation to overcome resistance from staff. Successful CPOE implementation appears to require financial incentives to improve the business case, and the experience of early adopters suggests that organizational factors and missions can be significant enablers even when financial incentives are not aligned.

Fourth, the target stakeholders for CER results concerning CPOE are more diverse than the typical users of other types of CER studies. They include hospital executives and technology vendors, in addition to physicians, pharmacists, and other clinical staff. This may increase the complexity of messaging to achieve effective and consistent dissemination of the CER results.

Despite the unique features of CPOE, similar delivery-system interventions based on CER evidence (particularly those that improve patient safety) may also benefit from some combination of strong mandates, systematic standards, and financial incentives that improve the business case for implementation.

Root Causes of CER Failure to Rapidly Change Clinical Practice

A myriad of factors influence whether CER is successfully translated into clinical practice. However, our synthesis suggests that some of these factors are “root causes” in the sense that they are fundamental and may represent high-leverage points for action to improve adoption of a new practice. We identified five root causes of failure when CER is slow to change clinical practice. These root causes manifest themselves in somewhat different ways across the case studies, appear to explain the strategies of the many stakeholders with an interest in CER, and typically exert their effects over multiple phases of the CER translation process.

1. Financial incentives are primary drivers of adoption of new clinical practices whether or not the practices are supported by CER evidence. CER results that threaten the financial interests of a stakeholder will be challenged at all phases of the CER translation process.

The most fundamental determinant of successful CER translation is the extent to which the economics of adopting a new clinical practice are favorable to providers and patients. Our case studies on the comparative effectiveness of interventional and noninterventional procedures highlight the perverse consequences of fee-for-service reimbursement as a driver of the use of procedures that CER evidence shows have little or no marginal benefit. Once patients are referred to interventional specialists, even if only for consultation, there is a high likelihood that they will receive an invasive procedure.

Our case studies highlight the role of financial incentives in influencing more than only the implementation phase of the CER translation process. In particular, financial incentives may supersede CER evidence in influencing the adoption of new clinical practices in the following ways:

- Stakeholders with a financial interest in the outcome of a CER study may seek to influence its design in order to increase the odds that its results will favor them, or they may initiate efforts to critique and thus undermine potentially unfavorable CER studies at the time the studies are enrolling participants. Critiques of a CER study design by interested stakeholders may peak when the results are released to maximize the likelihood that the study will be viewed as methodologically weak.
- The interpretation of CER results through a dynamic scientific debate among stakeholders appears to be influenced by financial incentives of the participants.
- The formalization of guidelines and measures based on CER evidence may be influenced in subtle ways by financing, and professionals have few financial incentives to facilitate the development of performance measures unless they will be paid based on the measure results. If guidelines do not evolve with the CER evidence, other formalization activities such as the modification of quality measures will be delayed or simply fail to occur.
- The dissemination of new practices is expensive, and the lack of financing for dissemination activities to support CER-based practices may be an important impediment to change. Aggressive dissemination activities directed toward payers may cause them to focus narrowly on areas where practice variation is extensive, where evidence clearly does not support a practice, and where risks to patients are unambiguous.
- Physicians working within a larger organizational context may be more likely to use performance measurement and feedback, patient decision aids, clinical decision support tools, and registries, all of which have the potential to increase responsiveness to CER.

Despite the seemingly powerful influence of financial incentives favoring both the status quo and an accelerating panoply of new procedures, recent trends, including the emergence of innovative payment models and new types of physician organizations, provide some basis for optimism that CER evidence can be more influential in the future. In addition, activities that curtail the influence of financial interests in each of the phases of CER translation (such as the recent Institute of Medicine [IOM] report calling for greater transparency and integrity in the guideline development process) may reduce the countervailing

forces that work to undermine or neutralize even the best CER evidence. Payers are also actively engaged in horizon-scanning for CER evidence that may form the basis of policies before practices become widespread in the community.

2. Even the best CER studies may fail to produce an unambiguous “winner,” so it may be difficult to achieve a consensus interpretation of the results.

CER studies that produce clear “winners” (i.e., showing unambiguously that one treatment is better than another or that two treatments have essentially equivalent effectiveness) should be more likely to change practice because they are difficult to challenge. However, our case studies suggest that even among the best-designed and -conducted CER studies, unambiguous outcomes are likely to be rare. Many factors increase the risk of an ambiguous result from a CER study, including design factors (e.g., use of active comparison groups rather than placebos), differential weighting of end points by stakeholders, and differences in provider equipoise for recommending treatments. Persuading stakeholders about treatment equivalence may be much more difficult than persuading them of treatment superiority. CER studies that produce ambiguous results open the door to selective interpretation, may undermine consensus interpretation of the results, and may fail to promote guideline updates by professional societies or the formation of coverage policies by payers. In cases where one treatment is found to be unambiguously harmful (e.g., the Women’s Health Initiative), clinical practice has been known to change rapidly, and our findings confirm to some extent that general rule, based on anecdotal reports suggesting decreased use of olanzapine in the post-CATIE period. Adverse-event data from registries may help to identify more unambiguous “losers” over time.

While ambiguity may lead to incomplete use of CER results and may limit the potentially attainable change in clinical practice, the lack of “winners” does not invariably mean that the CER fails to have an impact on clinical practice. Many discussants indicated that the goal of CER is not identifying “winners,” but generating information to help physicians and patients arrive at satisfactory treatment decisions. Several of our case studies might have reassured physicians and patients that moderate-dose conventional antipsychotics or less-aggressive therapies can have benefit comparable to that of more-aggressive therapies.

3. Cognitive biases play an important role in stakeholder interpretation of CER evidence and may be a formidable barrier to clinical-practice change.

At least three cognitive biases may influence the way in which physicians and other stakeholders interpret new CER evidence. First, confirmation bias, the tendency for a stakeholder to embrace evidence that confirms preconceived notions of treatment effectiveness and reject evidence to the contrary (while typically criticizing the studies on methodological grounds), may reinforce established practice patterns. Stronger study designs (emphasizing particularly the generalizability of findings) and careful monitoring of study conduct (particularly to prevent crossover for randomized study designs)

may preempt these critiques and counteract the influence of confirmation bias. A second bias is the belief that intervening aggressively is better than inaction, even when the marginal benefit is small. This bias may be reinforced by perverse financial incentives and by providers' perceived risk of malpractice liability if they fail to act; however, more complete data on treatment harms or heterogeneity of benefits may promote greater equipoise among both physicians and patients. A third cognitive bias, which may be reinforced through messaging by interested stakeholders, is the tendency to perceive new technologies as superior to older technologies—a problem with lengthy CER studies, during which technology often advances. Adaptive study designs could provide the flexibility to allow the evolution of treatments through the course of a trial, but these approaches were not used in any of the trials we studied, with the exception of CATIE.

Strategies for mitigating cognitive biases are available, but their effectiveness is not completely clear. Enhancing the transparency of stakeholder positions by using approaches that foster explicit formal decisionmaking processes is one approach to mitigating cognitive biases in a policymaking context. Disclosure of financial and intellectual conflicts of interest is another strategy used by the IOM and others. Regulation of detailing and direct-to-consumer advertising may also be effective.

4. The questions posed by a CER study and its design may not adequately address the needs of end users or focus adequately on the clinical decisionmaking opportunities that have the greatest potential to influence clinical practice.

Our case studies suggest that CER faces potentially unrealistic expectations on the part of multiple end users. First, there is an unavoidable tension in the design of CER studies between supporting personalized medicine and supporting clinical policymaking, which requires generalizable results on larger populations. Second, as demonstrated in CATIE, CER studies may not be designed with a comprehensive or explicit understanding of the beliefs and concerns of clinical practitioners, such as their preoccupation with the relative safety of classes of antipsychotics rather than their relative effectiveness. Third, head-to-head comparisons of treatments may help providers select appropriate treatments in the later stages of clinical decision algorithms, but CER concerned with upstream diagnostic tests or procedures may have a larger impact on patient outcomes and the overall value of care. If distinct providers (e.g., primary-care providers) are responsible for upstream decisions to refer, and if both providers and patients face weaker incentives to choose an intervention (compared to interventionists), their decisionmaking may be more readily influenced by evidence.

5. Clinical decision support and patient decision aids can help to align clinical practice with CER evidence, but they are not widely used.

Perverse financial incentives and lack of accountability for implementation have limited the production and dissemination of both clinical decision support tools and shared-decisionmaking aids. Decision support tools to promote evidence-based diagnostic testing and appropriate referral to specialists are

uncommon, although both may lead to the use of treatments that are better aligned with CER evidence. Well-informed patients who make treatment decisions according to their preferences may ultimately serve as a counterweight to providers who lack equipoise. However, the prevalence of direct-to-consumer advertising has grown, and this may create or reinforce misconceptions about treatments. Even if incentives for adopting these tools were better aligned, the challenge of integrating them seamlessly into clinical practice has not been solved, and limited HIT infrastructure and inadequate provider training on shared decisionmaking may continue to pose barriers to the implementation of these tools in the coming years.

Limitations of This Study

Because our primary objective was to identify and synthesize themes across case studies, we struck a specific balance between depth of content and breadth of inclusion of case studies. Our sample of expert discussions and stakeholder perspectives for each of the case studies was limited, and a larger sample of end users of CER from a diversity of practice settings could have identified additional barriers and enablers. The perspectives of the device industry and, to a lesser extent, the pharmaceutical industry are also underrepresented, despite considerable outreach efforts. Most potential discussants declined to participate, citing concerns about the sensitivity of the information they might be asked to share.

We chose not to use a formal qualitative research methodology that included the coding of themes with the use of specialized analytic software. Biased interpretations of the data by the research team were mitigated by requiring a minimum of three investigators to be present for each discussion, and we used email follow-up for areas that needed clarification. Our root-cause analysis drew primarily on themes that were mentioned repeatedly by stakeholders. Because of the limited scope of topics and the limited number of discussions we were able to hold, some of our findings regarding the root causes of failed CER translation or the facilitators of practice change may not be generalizable to other topics or to a broader range of practice settings.

Policy Implications

While the root causes of failure to translate CER evidence into clinical practice are formidable, they are not insurmountable. After reflecting on the policy implications of our case studies, we identified a range of policy options that can address the root causes and promote more effective translation of CER evidence into clinical practice. Each of these policy options can be categorized into one of the following domains: governance, standards, financing, professionalism, marketing and education, and research and evaluation. Each of the policy options would be optimally deployed within a healthcare system having a CER-enabling infrastructure that

1. Enables generation of CER that is more relevant to decisionmakers

2. Enables more-effective translation of research into practice
3. Enables more-effective evaluation of the impact of translation activities.

The policy options presented here are not intended to create a centralized command-and-control infrastructure that determines the CER agenda. Changes in the translation process, such as reengineering financial incentives, must be carried out by a diverse set of public and private stakeholders, and the changes must address a remarkable diversity of payment arrangements. The policy options we suggest could bring greater coherence and transparency to the process of CER translation, achieve greater balance of the influence of stakeholders participating in CER translation, and enhance the voice of the public and patients whose health outcomes depend on effective, safe, and affordable care. Enacting some or all of these options could be expected over time to modify the financial and other incentives that shape clinical decisionmaking so that decisions will be increasingly based on evidence rather than other considerations.

Governance

Create a transparent governance mechanism with oversight of the CER translation process:

1. *Include patient or consumer representatives.* Patients are ultimately the financers of CER and the key beneficiaries of the clinical practices guided by the research. They should be engaged to ensure that the CER translation process is well informed by the end-user perspective and to provide a counterbalance to other stakeholder interests.
2. *Include public and private payer and purchaser representatives.* As stewards of the financing of healthcare and representatives of their member or customer interests, payers and purchasers may be able to identify the specific opportunities for high-value care that would be especially amenable to CER and for which modification of payment or coverage policies could be especially effective in optimizing clinical practice.
3. *Enable and support public comment opportunities.* Vigorous solicitation of public comment with verification and full disclosure of the potential conflicts of interest (both financial and intellectual) of those who offer comments can enhance the credibility of the CER enterprise.
4. *Institute strong policies on disclosure and management of potential conflicts of interest.* CER evidence that is perceived as biased by financial interests will lose credibility and can impede the take-up of new practices; it might also be countered by policies that identify and manage potential financial, institutional, and intellectual conflicts of interest.
5. *Use the governance mechanism to generate a prospective public record of stakeholder expectations.* Documenting the positions of relevant stakeholders at the outset of CER studies with respect to the study objectives and the parameters around which the results should be interpreted creates a public record of expectations of each stakeholder and may discourage post-hoc efforts to undermine the credibility of studies that produce results contrary to the interests of specific stakeholders.

Standards

Support and enhance creation of standards for CER generation and translation:

1. *Incorporate data elements that are critical to translation activities into the proposed national CER registry.* Creating explicit standards for the description of CER study objectives, design, sampling, and causes of heterogeneity in sampled populations within a CER registry may be useful to guide formulation of a consensus interpretation of results and to avoid post-hoc reframing of the questions and implications by stakeholders to serve their interests.
2. *Encourage development of standardized electronic clinical data systems (clinical registries).* Standardized electronic clinical registries can enable rapid and low-cost CER, can provide data for longitudinal tracking systems to evaluate the impact of CER translation activities on clinical practice patterns, and may better support decisionmaking because clinicians and patients perceive data derived from them to be trustworthy.

Financing

Encourage public financing of CER translation and promote the use of CER evidence in payment programs:

1. *Provide direct and indirect support for formalization of CER evidence and the dissemination of CER-based clinical practices.* Public financing for both translation of high-profile CER evidence in guidelines, quality measures, and clinical decision support and subsequent dissemination activities could counteract the influence of industry in the process of formalizing CER evidence that might otherwise undermine the evidence or selectively promote less effective or more costly practices.
2. *Promote the use of CER-based clinical practices through payment policy and incentive programs directed toward providers and patients (e.g., value-based purchasing).* Encouraging “prudent purchasing strategies” for public-sector payers based on CER evidence can assure that the financial incentives of providers and patients in the delivery of healthcare are well aligned to support clinical practices based on CER evidence and discourage practices that are not evidence-based.

Professionalism

Supporting professional consensus across the phases of CER translation:

1. *Foster and support a broad vision of professionalism in the governance of CER translation.* Broadly constituted professional committees may be able to produce balanced, consensus interpretations of CER results and may resolve differences of opinion about interpretation of those results in a transparent manner. Multispecialty clinical registries and health-information exchange may counteract the tendency to focus on narrowly defined subspecialty interests.
2. *Include training for professionals on the role of cognitive biases in diagnostic and treatment decisionmaking.* Training in the nature, role, and impact of cognitive biases can enable professionals to recognize the circumstances under which decisions and clinical recommendations are prone to cognitive biases and to employ specific techniques that modify the decisionmaking context to compensate for these biases.

Marketing and Education

Promote demand for CER-based clinical services through public education and marketing:

1. *Promote patient demand for CER-based clinical services through shared decisionmaking that includes formal patient decision aids.* Shared decisionmaking involving the use of formal decision aids is the most prominent approach for assuring that patients can receive the best available evidence about alternative tests and treatments in a usable form. Decision aids could counter the messages that promote suboptimal clinical practices.
2. *Support “social marketing” campaigns for high-profile CER results to counteract the effects of industry-sponsored detailing and direct-to-consumer advertising.* Marketing campaigns, including detailing to clinicians and direct-to-consumer advertising, are generally aimed at exploiting cognitive biases, and this may impede the uptake of CER-based clinical services. Social marketing—the application of marketing techniques to promote behavioral change—has the potential to increase awareness and demand for evidence-based healthcare services by promoting greater patient engagement in medical decisionmaking.

Research and Evaluation

1. *Support research to identify the gaps in clinical decisionmaking that are the highest-priority topics for end users of CER.* Research on end-user needs can help identify high-priority topics and increase the relevance of CER to payers, professionals, and the public while fostering the selection of approaches for disseminating CER results tailored to the expectations of key stakeholders.
2. *Promote integration of CER registries with clinical registries to support evaluation of the impact of CER studies and the factors associated with successful translation.* Prospective evaluation of the impact of CER could be strengthened if available data sources can provide valid and reliable estimates of current clinical-practice patterns, which may be found in clinical registries. Clinical registries that are capable of providing longitudinal data on patients may enable these more complex studies and increase the relevance of results for end users.

3. *Support projects that develop unbiased and efficient methods for formalization of CER results.* Methods for developing and refining guidelines, performance measures, and clinical decision support tools are still a work in progress. Support for research and demonstration projects that develop and study new methods for formalization could lead to more effective, efficient, and unbiased tools.
4. *Support projects that enhance the utility of CER results by demonstrating and evaluating models for the use of decision aids by clinicians and patients.* Creating more effective decision aids, training professionals to use them, and developing strategies for embedding them in routine practice have all proven challenging. Research and evaluation projects that lead to better decision aids and their more effective use could increase the impact of CER.
5. *Support research into the ways in which CER evidence is used by integrated delivery systems.* Integrated systems may have unique perspectives on which CER topics are likely to have the greatest impact on clinical practice, and their CER translation experience may be invaluable. Future studies might involve engaging these organizations to elicit best practices in CER translation and evaluating which strategies may be transferrable to nonintegrated delivery settings.

The federal government is making a sizable investment in CER in the hope that the results will influence the decisions of clinicians and patients, optimize the quality of care, and lower costs. Given these goals, attention to the root causes of ineffective translation of CER evidence into practice seems critical. The list of policy options we have outlined is not exhaustive, but we believe that these options may provide guidance to a broad set of policymakers concerned with the organization and financing of healthcare. Taken together, the options suggest a number of different paths forward. Exploring multiple approaches may be appropriate in view of the large number of factors that impede CER translation. As the ARRA-financed CER portfolio begins to produce new evidence, a number of opportunities will arise in the near term to experiment with these strategies.

Acknowledgments

The authors thank Caroline Taplin (Task Order Officer) and Kate Goodrich of the Office of the Assistant Secretary for Planning and Evaluation at the U.S. Department of Health and Human Services for their guidance and support throughout this study.

The authors also thank the following individuals, who served as members of an expert panel, for their constructive feedback and insightful discussions: Edgar Black, Medical Director of Policy Resources at Blue Cross Blue Shield Association; Susan Dentzer, Editor-in-Chief of *Health Affairs*; J. Mark Gibson, Deputy Director for Evidence-Based Policy at Oregon Health & Science University; Chris Hafner-Eaton, Health Science Policy Analyst at the National Institutes of Health; Howard Holland, Acting Director of the Office of Communications and Knowledge Transfer at the Agency for Healthcare Research and Quality; Arthur Levin, Director of the Center for Medical Consumers; Brian Mittman, Director of the Department of Veterans Affairs (VA) Center for Implementation Practice and Research Support; and Harold C. Sox, Professor of Medicine at Dartmouth Medical School.

Finally, the authors thank Dmitry Khodyakov of RAND and David Atkins of the VA Health System for their constructive reviews of this study. Their input helped to improve the report.

Abbreviations

| | |
|---------|---|
| ACC | American College of Cardiology |
| ACE | angiotensin-converting enzyme |
| ACO | accountable-care organization |
| ACS | acute coronary syndrome |
| ADE | adverse drug event |
| AHA | American Heart Association |
| ALLHAT | Antihypertensive and Lipid Lowering Treatment to Prevent Heart Attack Trial |
| APA | American Psychiatric Association |
| ARRA | American Recovery and Reinvestment Act of 2009 |
| ASPE | Assistant Secretary for Planning and Evaluation |
| BARI 2D | Bypass Angioplasty Revascularization Investigation 2 Diabetes |
| BCBS | Blue Cross Blue Shield |
| BNP | B-type natriuretic peptide |
| CABG | coronary artery bypass graft |
| CAD | coronary artery disease |
| CATIE | Clinical Antipsychotic Trials of Intervention Effectiveness |

| | |
|-----------|--|
| CEO | chief executive officer |
| CER | comparative effectiveness research |
| CME | continuing medical education |
| CMS | Centers for Medicare and Medicaid Services |
| COMPANION | Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure |
| COPD | chronic obstructive pulmonary disease |
| COURAGE | Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation |
| CPOE | computerized physician order entry |
| CRT | cardiac resynchronization therapy |
| CRT-D | combined CRT and ICD therapy |
| CT | computed tomography |
| CTCA | computed tomography for coronary angiography |
| CVD | cardiovascular disease |
| EHR | electronic health record |
| FDA | Food and Drug Administration |
| GWTC | Get With the Guidelines |
| HF | heart failure |
| HHS | Department of Health and Human Services |
| HIT | health information technology |
| HITECH | Health Information Technology for Economic and Clinical Health Act |
| ICD | implantable cardioverter defibrillator |

| | |
|---------|---|
| IOM | Institute of Medicine |
| IT | information technology |
| LMW | low molecular weight |
| LVEF | left ventricular ejection fraction |
| MASS-II | Second Medicine, Angioplasty, or Surgery Study |
| MI | myocardial infarction |
| MPS | myocardial perfusion single-photon emission computed tomography |
| MRI | magnetic resonance imaging |
| NASS | North American Spine Society |
| NCDR | National Cardiovascular Data Registry |
| NIMH | National Institute of Mental Health |
| NYHA | New York Heart Association |
| OAT | Occluded Artery Trial |
| ODI | Oswestry Disability Index |
| OMT | optimal medical therapy |
| OPT | optimal pharmacological therapy |
| PCI | percutaneous coronary intervention |
| PCORI | Patient-Centered Outcomes Research Institute |
| PCPI | Physician Consortium for Performance Improvement |
| PORT | Patient Outcomes Research Team |
| QALY | quality-adjusted life year |

| | |
|--------|---|
| RBM | radiology benefits manager |
| RCT | randomized controlled trial |
| ROI | return on investment |
| SAMHSA | Substance Abuse and Mental Health Services Administration |
| SCAI | Society for Coronary Angiography and Interventions |
| SF-36 | Medical Outcomes Study 36-Item Short-Form Health Survey |
| SPORT | Spine Patient Outcomes Research Trial |
| VA | Department of Veterans Affairs |

Introduction and Summary of Approach

Defining Comparative Effectiveness Research

Insufficient evidence regarding the effectiveness of medical treatments has been identified as a key source of inefficiency in the U.S. healthcare system. Clinicians vary widely in their recommendation and use of diagnostic tests and treatments for patients with similar symptoms or conditions. This variation has been attributed to clinical uncertainty, since the current published scientific evidence base does not provide adequate information to determine which treatments are most effective for patients with specific clinical needs. A surprisingly high proportion of the healthcare services received by patients today may have limited value, while some care may even be harmful (Owens, Qaseem, et al., 2011). By helping clinicians, patients, payers, and others to better distinguish between effective and ineffective (or potentially harmful) therapies, comparative effectiveness research (CER) has the potential to slow the growth of spending on healthcare without adversely affecting the quality of care.

The Institute of Medicine (IOM) defines CER as the generation and synthesis of evidence that compares the benefits and harms of alternative methods for preventing, diagnosing, treating, and monitoring a clinical condition or improving the delivery of care. It further states that the purpose of CER is to assist consumers, clinicians, purchasers, and policymakers in making informed decisions that will improve healthcare at both the individual and population levels. The Federal Coordinating Council for Comparative Effectiveness Research used similar terms to define CER (Federal Coordinating Council for Comparative Effectiveness Research, 2009).*

CER differs from other clinical research in a number of key ways. First, it is intended to provide information directly relevant to the decisions of a wide range of stakeholders, since it addresses both the clinical and policy aspects in the selection of tests, treatments, and quality improvement strategies. Unlike randomized controlled trials (RCTs), which examine the efficacy of tests, treatments, or quality

* The Federal Coordinating Council for Comparative Effectiveness Research defined CER as follows: Comparative effectiveness research 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 decisionmakers, responding to their expressed needs about which interventions are most effective for which patients under specific circumstances.

improvement strategies in highly specified samples, CER addresses their effectiveness in a broader and more diverse set of populations and settings. Second, CER focuses on comparisons of alternative clinical approaches that have demonstrated efficacy in RCTs but may differ with respect to effectiveness in some populations, side effects, or costs. Third, CER may emphasize collection of data on subgroups in order to facilitate individualized decisionmaking. Fourth, CER uses observational or pragmatic research methods designed to enhance generalizability, including treatment allocation and the measurement of outcomes in real-world settings that are relevant to patients. In particular, CER may draw from a more comprehensive set of population outcomes, examining a wider range of harms, as well as benefits, that may be relevant to patients.

Federal Investment in Comparative-Effectiveness Research

The vast majority of previous CER consists of systematic reviews of published research studies. However, many of the studies included in the reviews are not considered CER; placebo-controlled trials far outnumber head-to-head trials, so comparisons between treatments are mainly indirect. Stakeholders increasingly demand that newly developed CER move beyond reviews, employing head-to-head comparisons of alternative treatments and enhancing internal validity through randomization of study subjects, while also increasing their generalizability to real-world settings by expanding the types of intervention strategies and inclusiveness of patient populations. The goal is to provide data that better guide a larger share of decisions about diagnosis and treatment.

The American Recovery and Reinvestment Act of 2009 (ARRA) provided more than \$1 billion in funding for CER, channeled through the National Institutes of Health (NIH), the Department of Health and Human Services (HHS), and the Agency for Healthcare Research and Quality (AHRQ). The Affordable Care Act established the Patient-Centered Outcomes Research Institute (PCORI), a private nonprofit entity, to oversee such research. Achieving widespread practice change in response to CER will require not only a large number of focused CER studies, but also timely dissemination and adoption of the results. Recognizing this fact, a sizable portion of the ARRA funding was devoted to research on the dissemination and implementation of CER.

Challenges of Translating CER Evidence into Practice

The federal investment in CER made possible through ARRA was a dramatic commitment of resources to clinical research, with the expectation that CER results would not simply influence clinical practice but would also improve the efficiency of healthcare delivery. To do this, CER must provide information that supports fundamental changes to healthcare delivery and informs the choice of diagnostic and treatment strategies. Such information could produce substantial reductions in the growth of healthcare spending. In an economy largely driven by technological innovation, the dissemination of new diagnostic tests, procedures, and treatments is relatively frequent and ongoing. However, many of these new tests and treatments, even those that are widely adopted, are not completely grounded in scientific evidence. Some (including prostate screening tests, use of drug-eluting stents, and certain types of surgery for vertebral fractures) become widely entrenched and are difficult to dislodge despite unambiguous

scientific evidence about superior alternative approaches (Redberg, 2011). At the same time, many new clinical practices are not quickly adopted, either because information does not reach decisionmakers in a usable format or because other barriers prevent their adoption.

If the goal of CER is to quickly achieve maximal impact on practice, historical barriers to the translation of CER evidence into new practices must be overcome. The CER translation process occurs within a complex environment involving multiple stakeholders who influence the way in which CER results are interpreted through the dissemination of messages, policies, or decision support tools that may facilitate or retard the rate at which new practices are adopted. These stakeholders, including pharmaceutical and device manufacturers, payers, federal and state policymakers, medical publishers, and the popular press, each have financial, professional, or advocacy interests that may conflict or align with new research findings. Many current dissemination strategies fail to address the specific social context in which physicians practice and may also fail to address patients' needs for unbiased information. Effective dissemination strategies will be needed to overcome CER translation barriers, but there is limited evidence on the effectiveness of strategies used to disseminate past CER evidence and associated clinical practices.

Objectives of the Project

This project had three main objectives. First, we sought to develop a framework with which to organize the array of barriers and enablers that influence the translation of CER evidence into new clinical practices. Prior frameworks, such as those developed by Rogers and many of those summarized by Greenhalgh, were extraordinarily helpful but were not necessarily tailored to the unique interaction between healthcare innovations, scientific evidence, payment policy, professional norms, and the asymmetry of knowledge between patients and clinicians about new practices and technologies (Rogers, 1995; Greenhalgh, Robert, et al. 2004). We anticipated that a simplified framework would enable us to isolate the key factors affecting each phase of the CER translation process, beginning with the generation of CER evidence and ending with actual adoption of new practices. We believed this framework could also inform future research on translation of CER into practice.

The second objective was to conduct case studies of the adoption of new clinical practices, using five key CER studies published in the past 15 years, and to identify key themes relating to the rapid or delayed adoption of new practices that emerged from the case studies. We held discussions with stakeholders representing a broad range of perspectives and examined the peer-reviewed literature associated with each study. Our framework guided our discussions and helped to ensure that we considered potential factors influencing all phases of the CER translation process. Synthesizing common themes across case studies provided insight into the "root causes" for the failure of CER to rapidly change clinical practice.

Our third objective was to help policymakers develop practical strategies that facilitate dissemination of CER-based clinical practices and thus maximize the effectiveness of the federal government's current investment in CER. To do this, we developed policy options—informed by our expert panel and partners at the HHS Office of the Assistant Secretary for Planning and Evaluation (ASPE)—and identified research needs. Our overall approach comprises a replicable methodology that can inform larger-scale,

prospective, in-depth qualitative and quantitative research on the impact of the federal investment in CER.

Framework for Translation of CER into Practice

On the basis of a review of existing frameworks that have evolved from the science of diffusion, dissemination, and implementation of innovations; discussions with our expert panel; and discussions with other experts, we developed a simplified framework to guide both our collection of case-study data and our subsequent analysis of the barriers to and enablers of CER translation into practice. As we conducted the case studies, we refined the framework to better reflect what we observed and to enable us to synthesize and organize the key themes that emerged (summarized in Chapter Seven).

Our conceptual framework posits that the CER translation process follows five key phases, shown in Figure 1.1. While Figure 1.1 suggests a linear temporal process, the phases are concurrent to some degree, and there appear to be multiple interactions between stakeholders at different phases. The phases are described in Table 1.1.

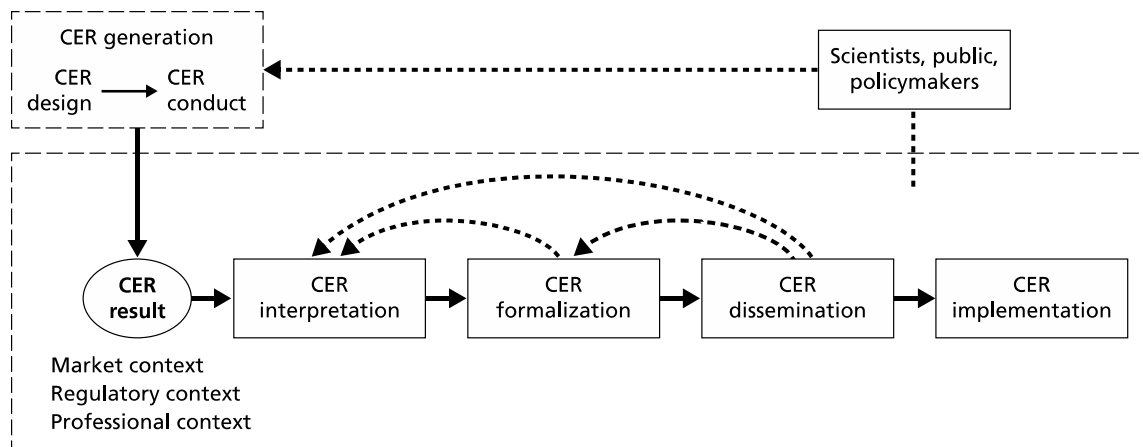
The first phase is *generation*, which includes the design and execution of the CER study. It involves primarily scientists, funders, and the public but is shaped by the needs of many other key stakeholders. In particular, patients or their advocates and clinical professionals may play an important role in helping to ensure that outcomes of CER studies are relevant to them. CER design elements decided during the generation process may strongly influence the relevance of the findings and thus the degree to which they are accepted and lead to changes in clinical practice.

The completion of a study and the publication of its findings initiate the *interpretation* phase, in which individual stakeholders assign a specific meaning to the results. Interpretation is a complex process shaped by the strength of the CER evidence, applicability of the evidence to the potential adopter's practice setting, each adopter's personal experience, and prior expectations of the benefits and harms of each treatment. For example, patients may assume that more costly treatments are more effective. Researchers who conduct systematic reviews also interpret the evidence, albeit with the use of protocols. Interpretation is influenced by messages from other key stakeholders as well, including professional societies, industry, the media, and opinion leaders. Stakeholders whose interpretations of CER results play a large role include

- CER researchers, whose initial presentation of the evidence may play a critical role in shaping interpretations by others
- Professional societies, which may produce consensus statements about the CER

Figure 1.1

Conceptual Framework for Translation of CER into Clinical Practice



NOTES: CER dissemination and implementation phases involve dissemination and implementation of *clinical practices* as opposed to *results*. Solid lines represent the pathways through which CER research is translated into clinical practice; dotted lines are feedback loops that influence prior phases, indicating the interactive and iterative nature of the process.

RAND TR924-1.1

Table 1.1

Phases of the Translation of CER into Clinical Practice

| Phase | Description |
|----------------|---|
| Generation | Generation includes the design and conduct of the CER study; it involves primarily funders and CER researchers, but research priorities are influenced by the needs of multiple stakeholders, including scientists, the public, and policymakers. |
| Interpretation | Stakeholders ascribe meaning to CER results based on a number of factors, including the strength of evidence, applicability of the evidence to the potential adopter’s practice setting, personal experience, and messages received by other stakeholders (e.g., professional societies, industry, media, and opinion leaders). |
| Formalization | Formalization is the process by which the interpretations of CER results are converted into guidance instruments such as clinical-practice guidelines, performance measures, and quality improvement tools. Multiple stakeholders may play roles in formalization through participation in guidelines committees, regulatory committees, and performance-measure development and endorsement processes. |
| Dissemination | Dissemination is the process by which CER information and/or associated tools designed to influence practice is actively transmitted to stakeholders. It typically promotes (or discourages) implementation of a new practice but may also have the goal of promoting a particular interpretation of the CER results. |
| Implementation | Implementation is the adoption of new clinical practices based on CER results. Implementation decisions may depend on a wide range of factors, including the dissemination of messages and the successful embedding of CER-related clinical guidance into tools that facilitate practice change, as well as the local market, regulatory, and professional context that may promote or |

impede changes. The implementation phase takes place primarily in local practice contexts.

- Industry, whose marketing divisions play a key role in developing and disseminating messages
- Funders, who may support and shape the communications strategies for CER results
- Advocacy organizations, which may interpret the evidence on behalf of their constituents
- Systematic reviewers, who interpret CER results in the context of the larger body of evidence and draw conclusions about the meaning of the results.

To produce a change in practice, CER results and their interpretations may pass through a *formalization* phase, in which the clinical practice that will be altered based on the CER evidence is identified and different interpretations are reconciled to produce some form of consensus practice recommendation. In our model, formalization occurs through the activities of national committees that generate or modify clinical-practice guidelines, create performance measures, determine rules for clinical decision support applications, or specify quality improvement strategies. Another type of formalization is the development of “knowledge-summarizing applications” that physicians and patients often use as online reference tools to identify the current recommendations regarding a particular clinical situation. Payers may also formalize CER results through the definition of coverage policy. Because these tools may be derived directly from systematic reviews and clinical-practice guidelines, there may be a lag between the publication of CER evidence and the tools that result from formalization.

Key stakeholders involved in formalization include

- Professional societies, mainly through the development of guidelines, but also through the development of performance measures and quality improvement tools
- Performance-measure developers, who create performance measures based largely on guidelines and other evidence
- Pharmaceutical and device-industry companies that produce detailing materials (e.g., pocket cards) and may support the development of other quality improvement tools
- Clinical decision support (CDS) developers, who may develop alerts, reminders, or clinical-pathway support tools to promote using care in line with the CER evidence
- Payers/purchasers, who develop coverage and reimbursement policies based on the CER evidence
- Policymakers, through the development and/or implementation of performance measures
- Decision-aid developers, who integrate CER evidence into patient decisionmaking tools.

To reach stakeholders, the formalization tools must be part of an active *dissemination* phase. Interpretation, formalization, and dissemination of messages and tools are unlikely to follow a linear process. Rather, as indicated in Figure 1.1, new evidence that evolves and is disseminated through the activities and messages of stakeholders reshapes the context in which interpretation and formalization occur. These processes are likely to be concurrent and may reinforce one another. While a range of stakeholders may play a role in dissemination, key stakeholders include

- Professional societies, which have multiple formal channels for reaching their members, including professional meetings, websites, and journals, through which they disseminate guidelines, consensus statements, and the CER research itself
- Performance-measure developers and CDS developers, who have channels for alerting end users to new tools based on CER evidence
- Payers and purchasers, who may implement new coverage or reimbursement policies or promote new delivery-system interventions consistent with CER evidence
- Advocacy organizations and drug or device manufacturers, who often produce messages in response to new CER evidence to influence other stakeholders
- Medical publishers and mass media, which can reach narrow and large audiences, respectively, to disseminate new CER evidence
- Specialty boards, which may disseminate new findings through certification requirements that require knowledge of relevant CER and demonstration of competency for quality improvement.

Finally, *implementation* occurs when and if stakeholders agree with the formalization of the CER findings sufficiently to adopt the new practice:

- Clinicians' recommendations of CER-informed practices are likely to be highly determined by reimbursement, availability of decision support, and sufficient capacity to deliver the intervention.
- Patients may vary in their treatment preferences and may face out-of-pocket costs associated with the intervention.
- Delivery organizations may have a wide array of decision supports and referral processes in place that facilitate adoption of CER-based clinical practices.
- Payers and purchasers influence implementation through reimbursement levels and financial and nonfinancial incentives.

All of the phases are highly dependent on a range of contextual factors, including the financial incentives defined by opportunities in the marketplace, practice expectations shaped by the professional environment, and the regulatory environment that defines what is possible within a legal or regulatory framework. Different factors may play a role in facilitating or impeding the adoption of new clinical practices during one or multiple phases, and individual factors may interact or offset one another.

Our conceptual framework contains a feedback loop that reflects the tendency of CER studies, through the process of translating research into practice, to generate related questions that were not addressed by the original study and whose answers have value for one or more stakeholders. While many stakeholders may advocate for new CER studies, in general, scientists, policymakers, and the general public have the largest role in the determination of new topics for CER.

Case-Study Research Approach

Identifying Case-Study Topics

In selecting case studies, our preliminary intent was to identify and include CER trials whose results challenged existing clinical practices. To identify such trials, we conducted an environmental scan of the peer-reviewed and grey literatures. We examined summaries of key clinical-research studies published in peer-reviewed journals, such as the Updates in General Internal Medicine series published by *Annals of Internal Medicine*, and we consulted with the project's expert panel, staff at ASPE, and RAND CER experts. We developed a preliminary list of case study topics (shown in Table 1.2) based on the following considerations:

- Conditions must have a high overall burden of illness (life expectancy, quality of life, cost)

- Conditions must be highly prevalent
- Studies must provide high-quality evidence (e.g., randomized designs).

To attain diversity across the case-study topics, we applied two additional considerations:

- Diversity of treatment modalities (e.g., drugs, devices, procedures, delivery system interventions)
- Diversity of treatment settings (e.g., inpatient and outpatient care settings).

While CER results always have the potential to challenge current practices, they do not invariably do so. In some instances, the results simply reinforce predominant practice; in others, the results may be ambiguous. For example, if CER provides estimates of risks and benefits in a population where treatment effects may be heterogeneous, the “right answer” may vary across subgroups or by variants on treatment. The results of some CER studies may suggest modifications to some aspects of practice while reinforcing the status quo for other aspects. For our case studies, we sought to include CER studies that challenged at least some aspect of practice in at least one population.

Table 1.2

Preliminary Case-Study Topics

| Topic | CER Finding | Reference |
|--|--|--------------------------|
| Medications | | |
| High-dose versus usual-dose statin for secondary prevention after myocardial infarction (MI) | Intensive lowering of LDL-C did not significantly reduce the primary outcome of major coronary events. | Pedersen et al., 2005 |
| Beta-blockers versus other drugs for primary hypertension | Beta blockers are less than optimal compared with other antihypertensive drugs. | Lindholm et al., 2005 |
| Diuretics versus calcium channel blockers versus angiotensin-converting enzyme (ACE) inhibitors versus alpha-adrenergic blockers for hypertension (ALLHAT) | Thiazide-type diuretics are superior in preventing one or more major forms of cardiovascular disease (CVD) and are less expensive. | Officers, 2002 |
| Tiotropium versus ipratropium for chronic obstructive pulmonary disease (COPD) | Tiotropium improves health outcomes and is associated with higher costs than ipratropium. | Oostenbrink et al., 2004 |

| Topic | CER Finding | Reference |
|--|---|-----------------------------------|
| Warfarin versus low-molecular-weight (LMW) heparin for outpatient treatment of acute venous thromboembolism | High-dose heparin is more effective than low-dose heparin. | Kovacs et al., 2003 |
| Salmeterol versus fluticasone for COPD | Combination treatment improved symptoms and lung function better than either component alone. | Calverley et al., 2003 |
| Rituximab versus usual care for follicular lymphoma | Rituximab improved outcomes for patients. | Hiddemann et al., 2005 |
| Atypical antipsychotic drugs versus conventional antipsychotic drugs for schizophrenia | Older antipsychotic medications were similar in effectiveness and had lower costs than newer antipsychotic medications. | Lieberman, Stroup, et al., 2005 |
| Procedures | | |
| Carotid artery stenting versus carotid endarterectomy | Rates of death and stroke were lower with endarterectomy than with stenting. | Mas et al., 2006 |
| Surgical versus nonsurgical treatment for lumbar spinal stenosis | Surgery had better outcomes than nonsurgical treatment in the observational data analysis. | Weinstein, Tosteson, et al., 2008 |
| Optimal medical therapy versus cardiac resynchronization therapy versus combined cardiac resynchronization therapy and defibrillator therapy for patients with moderate to severe heart failure (HF) | Cardiac resynchronization reduced hospitalization and improved functional status and survival in patients with moderate to severe HF. | Bristow, Saxon, et al., 2004 |
| Coronary-artery revascularization versus none prior to elective major vascular surgery | Coronary-artery revascularization before elective vascular surgery did not improve long-term outcomes. | McFalls et al., 2004 |
| Preventive screening and diagnostic testing | | |
| Screening for abdominal aortic aneurysm | Screening for abdominal aortic aneurysms reduced mortality. | Lindholt et al., 2005 |
| Virtual colonoscopy versus optical colonoscopy for colorectal cancer screening | Virtual colonoscopy compared favorably with optical colonoscopy. | Pickhardt et al., 2003 |
| Computed tomography for coronary angiography (CTCA) versus usual care for acute coronary syndrome | Use of CTCA for acute coronary syndrome should be avoided because of significant radiation exposure. | Einstein et al., 2007 |

| Topic | CER Finding | Reference |
|---|---|------------------------------------|
| Medication versus procedure | | |
| Percutaneous coronary intervention (PCI) versus optimal medical therapy for chronic stable angina | PCI and optimal medical therapy provided equivalent survival benefit and relief of angina symptoms. | Boden et al., 2007 |
| Primary PCI versus thrombolytic therapy for acute coronary syndrome | PCI was superior to thrombolytic therapy for acute coronary syndrome. | Aversano et al., 2002 |
| Delivery-system interventions | | |
| Computerized physician order entry (CPOE) to prevent serious medication errors | Introduction of CPOE system and a team intervention into the hospital reduced the incidence of serious medication errors. | Bates et al., 1998 |
| Wrong-site, wrong-procedure, wrong-person surgery prevention techniques versus usual care | Implementation of the Joint Commission's "Universal Protocol" reduced the incidence of wrong-site, wrong-procedure, and wrong-person surgeries relative to usual care. | NQF Report on Safe Practices, 2009 |
| Use of critical-care-certified physicians versus usual care | Patients treated by physicians who have specific training and certification in critical-care medicine had better outcomes than those treated by non-critical-care-certified physicians. | NQF Report on Safe Practices, 2009 |

Working from our preliminary list of potential CER studies and using the considerations described above as a guide, we narrowed the list to seven case studies that addressed these considerations. Five topics were selected for investigation (Table 1.3), and the other two were kept as backups.

The Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) trial was selected as one of the most prominent CER studies to involve comparison of two classes of pharmaceutical agents and because of findings that were somewhat controversial. The Spine Patient Outcomes Research Trial (SPORT) and Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation (COURAGE) studies were selected to illustrate the comparison of surgical versus nonsurgical (or procedural versus nonprocedural) management of a chronic clinical condition. The Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure (COMPANION) study was chosen to represent the study of a device with demonstrated efficacy in a subgroup of the potentially affected population. The CPOE study was selected to represent the class of CER involving delivery-system interventions.

Table 1.3**Final Case-Study Topics**

| Topic | CER Finding | Reference | Type of Comparison |
|--|---|-----------------------------------|------------------------------|
| Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE): atypical antipsychotic drugs versus conventional antipsychotic drugs for schizophrenia | Older antipsychotic medications were similar in effectiveness and had lower costs than newer antipsychotic medications. | Lieberman, Stroup, et al., 2005 | Medications |
| Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation (COURAGE): percutaneous coronary intervention versus optimal medical therapy for chronic stable angina | PCI and optimal medical therapy provided equivalent survival benefit and relief of angina symptoms. | Boden et al., 2007 | Medication versus procedure |
| Spine Patient Outcomes Research Trial (SPORT): surgical versus nonsurgical treatment for lumbar spinal stenosis | Surgery had better outcomes than nonsurgical treatment in the observational data analysis. | Weinstein, Tosteson, et al., 2008 | Procedures |
| Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure (COMPANION): optimal medical therapy versus cardiac resynchronization therapy versus combined cardiac resynchronization therapy and defibrillator therapy for patients with moderate to severe HF | Cardiac resynchronization reduced hospitalization and improved functional status and survival in patients with moderate to severe HF. | Bristow, Saxon, et al., 2004 | Procedures |
| Computerized physician order entry (CPOE): interventions to prevent serious medication errors | Introduction of CPOE system and a team intervention into the hospital reduced the incidence of serious medication errors. | Bates et al., 1998 | Delivery-system intervention |

Literature Review

After selecting the case-study topics, we conducted a literature review to obtain additional background information on each topic. Web of Science, an academic citation index, served as the primary tool for our searches. The database allowed us to identify articles that referenced each CER study, such as editorials and letters that might have been used to influence interpretation of the results or promote dissemination, and empirical studies that may have either reinforced or contradicted findings from the

initial CER study. We also searched PubMed to identify related articles that may not have cited the CER study and thus might have been missed in the Web of Science search. In addition, throughout the course of our discussions, informants recommended additional articles to aid our understanding of each topic.

Identifying Potential Discussants

To identify discussants, we developed a list of relevant stakeholder types, which we refined through consultation with our expert panel, ASPE staff, and RAND experts. We then developed a preliminary list of potential discussants based on our literature searches for each case-study topic. We extended invitations to a set of discussants who, taken together, could provide perspective on all of the phases of the CER translation process. Some informants who did not have detailed knowledge of our five case-study topics provided their perspective on CER dissemination more broadly. We refer to these as crosscutting discussions.

Development of the Discussion Guide

We developed a core discussion guide that would enable us to address key topics in a consistent manner across case studies, along with a set of probes that could be used to direct the discussion. The discussions were organized around the following sequence: introductions, the context for the discussion, in-depth exploration of salient knowledge of and experience with the release and subsequent diffusion of the CER findings, a discussion of the individual's dissemination activities at his/her own institution, and key lessons learned about the impact of the CER results. The following discussion topics served as the foundation for our conversations:

- Impressions of the dissemination of results and subsequent changes in clinical practice
 - Whether or not the CER study changed clinical practice
 - The most influential stakeholders in communicating the messages of the CER study
 - The most effective dissemination activities
 - Key barriers to or enablers of change to clinical practice
- Dissemination strategies pursued by the individual discussant or his/her organization
 - Dissemination channels used
 - The target audience for dissemination initiatives
 - The success of dissemination initiatives
 - Key dissemination challenges

- Key lessons learned about what impacts dissemination of CER results
- Recommendations of other discussants with relevant perspectives.

Consistency of topics across discussions allowed us to draw contrasts between different stakeholder perspectives. However, we also tailored specific discussion topics to individual stakeholder types. The nature of the topics evolved organically, based on our emerging understanding of issues within each case study.

Recruiting Discussants

We sent prospective discussants brief email invitations requesting an informal phone conversation regarding the case-study topic. Along with each email, we included details about the study and our own brief summary (typically three pages) of the relevant CER study for each discussant. Individuals who failed to respond to our initial request received two additional email follow-ups, after which no further communication was attempted. Because we guaranteed anonymity for all discussants, we provide only a profile of the discussants, using broad categories:

CATIE (11 discussions)

- Former president of a large psychiatric professional society
- Medical director for a mental health advocacy organization
- Policy expert at a mental health advocacy organization
- Federal agency
- Federal agency (two discussants)
- Federal agency
- Psychiatrist and schizophrenia researcher
- Psychiatrist and schizophrenia researcher
- Psychiatrist and mental health services researcher
- Mental health services researchers (two discussants)
- Mental health services researcher

COURAGE (six discussions)

- Interventional cardiologist and professional-society committee member
- Interventional cardiologist and professional-society committee member
- Interventional cardiologist and professional-society committee member
- Director of cardiovascular research for an integrated health system
- Quality expert at a professional society

- Cardiologist and professional-society committee member

COMPANION (five discussions)

- HF specialist and clinical researcher
- HF specialist and professional-society committee member
- HF specialist and clinical researcher
- Director of cardiovascular medicine at an academic medical center
- Director of electrophysiology at an academic medical center

SPORT (seven discussions)

- Orthopedic surgeon and clinical researcher
- Orthopedic surgeon and clinical researcher
- Orthopedic surgeon and clinical researcher
- Orthopedic surgeon and clinical researcher
- Professional-society committee member
- Professional-society committee member
- Chair of orthopedics at an academic medical center

CPOE (seven discussions)

- Expert in clinical decision support and health information technology (HIT) researcher
- Expert in clinical decision support and HIT researcher
- Executive at an employer coalition
- Chief information officer for a large health system (two discussants)
- Senior medical informaticist at an integrated health system
- Director of clinical informatics at an academic medical center
- Founder of a clinical decision support company

Crosscutting (17 discussions)

- Senior leadership at a specialty board
- Senior leadership at a professional society
- CER expert and clinical decision support developer
- Editor-in-chief of a prominent health journal
- Former editor of a prominent health journal

- Editor-in-chief of a university-affiliated health publishing company
- Patient decision-aid developer and shared-decisionmaking researcher
- Medical director of a large health plan
- Medical director of a large health plan (two discussants)
- Senior leadership at an integrated health system
- Senior leadership at an integrated health system
- Medical director of a state Medicaid agency
- Vice president and former chief medical officer for a large insurer
- Director of medical policy and technology at a large health plan
- Senior leadership at an integrated health system
- Director of cardiovascular research at an integrated health system
- Experts in CER involving pharmaceuticals (two discussants)

Discussions with Stakeholders

We conducted 53 telephone interviews, with an average of eight discussions per case-study topic. A minimum of two RAND researchers and one research assistant participated in each call. The calls lasted from 30 to 60 minutes. The interviews, although open-ended, were structured around the discussion guide for each case study. The RAND callers took extensive notes during the interviews, but they were not recorded or formally transcribed. After each call, the researchers debriefed and discussed key points raised by the discussant and the extent to which themes were consistent with those from prior discussions. Stakeholders were promised anonymity to encourage candid discussions.

Analysis of Qualitative Data

Notes from each discussion were organized by theme and synthesized across individuals within each case study. Information from case-study-specific literature reviews was combined with qualitative data from discussions to generate a narrative summary of the key barriers to and enablers of translating CER into practice. Ambiguities in interpretation and contradictions were resolved through discussion and, when necessary, follow-up correspondence (typically email) with the original discussant. Evidence of emerging patterns was investigated with additional follow-up focused questions or literature review.

Organization of This Report

The remainder of this report is organized as follows. Chapters Two through Six summarize the five case studies. Each chapter includes a brief description of the clinical-practice context leading up to the CER findings and a summary of key results, along with a summary of the key barriers to and enablers of the CER translation process. Chapter Seven presents a summary of themes that emerged consistently, organized around five root causes of failure of CER to change clinical practice, along with an assessment of the study's limitations. Chapter Eight discusses the policy implications of our findings and suggests opportunities for future research.

CATIE Case Study

Clinical-Practice Context

The first antipsychotic medication, chlorpromazine, was introduced in 1954. A major side effect of this and other “first-generation” drugs was the development of extrapyramidal symptoms, including several serious movement disorders such as tardive dyskinesia. In 1990, the Food and Drug Administration (FDA) approved clozapine for a limited set of indications (limited because of the risk of serious, life-threatening side effects), ushering in the second generation of antipsychotic medications. The fundamental clinical difference between the first- and second-generation antipsychotics was the lower risk of extrapyramidal symptoms in the latter. The second-generation medications came to be known as “atypical” antipsychotics. The mechanism of action for these agents and the reasons for individual differences in response are still not completely understood.

The introduction of the atypical antipsychotics was heralded by the medical industry and others as the first breakthrough in the treatment of schizophrenia in nearly 40 years, and their use increased rapidly. Many prescribers came to believe that these medications had cognitive-enhancing properties and were effective in controlling positive, negative, and mood symptoms. In addition, they had enhanced tolerability (Lewis and Lieberman 2008). Prescribing practice prior to the CATIE trial may also have been influenced by treatment guidelines and performance measures that emphasized the superiority of the atypical antipsychotics (Owens 2008). For example, one of the performance indicators within the Substance Abuse and Mental Health Services Administration’s (SAMHSA’s) Uniform Reporting System was a measure of the rate at which second-generation antipsychotics were prescribed (Covell, Finnerty, et al., 2008).

In 2003, the results of two key studies called into question beliefs about the superiority of the atypical antipsychotics. The first was a trial funded by the Department of Veterans Affairs (VA) health system that compared a first-generation and a second-generation antipsychotic medication and found that they had comparable efficacy and risk of extrapyramidal symptoms and their use resulted in comparable quality of life (Rosenheck, Perlick, et al., 2003). A meta-analysis published the same year showed that optimal dosing for conventional antipsychotics might avoid the high rates of extrapyramidal symptoms typically seen with first-generation medications (Leucht, Wahlbeck, et al., 2003). Neither of these studies caused the psychiatric community (except for a few dissenters) to doubt the superiority of second-generation medications. However, the VA study did highlight the risk of metabolic side effects from them, and these concerns quickly gained recognition by many physicians. According to one

discussant, the pharmaceutical industry had attempted to suppress information on the metabolic side effects of second-generation medications.

In addition, insurers were becoming concerned about the cost of atypical antipsychotics, which constituted a significant share of the total pharmaceutical expenditures of Medicaid budgets in many states (Parks, Radke, et al., 2008). Advocates for parity in mental health insurance coverage worried that questions about the effectiveness or side effects of atypical antipsychotics might be seized upon by insurers to curtail access to them for a population that was already struggling to obtain adequate coverage and access to therapy.

The Comparative-Effectiveness Question

Despite the lower rate of extrapyramidal symptoms of second-generation antipsychotics, the optimal choice of drug treatments for patients with schizophrenia had been disputed for at least four reasons: (1) uncertainties surrounding effectiveness in controlling psychotic symptoms; (2) uncertainty about the relative incidence of side effects, including tardive dyskinesia and metabolic side effects; (3) some evidence that second-generation antipsychotics improve cognition; and (4) the high cost of the medications. These concerns drove a need to assess the overall value of second-generation antipsychotics.

Study Design Characteristics

The National Institute of Mental Health (NIMH) funded the \$42.6 million CATIE study in 1999 to compare the effectiveness of first- and second-generation antipsychotics. The first-generation antipsychotic perphenazine was chosen to be compared with three second-generation medications: olanzapine, quetiapine, and risperidone. A small number of patients were randomized to receive ziprasidone, which was approved during the course of the trial. CATIE was considered a landmark trial because of its size, duration, and public sponsorship. Furthermore, it was designed to be generalizable to real-world clinical settings through limited exclusion criteria, the enrollment of patients from diverse settings, and the flexible dosing protocols it permitted. The primary outcome measure, time to discontinuation of the study medication for any cause, was also chosen because of its relevance to clinical practice—it represents a summary measure of effectiveness that captures both the effectiveness and tolerability of treatment.

Study Results

The initial results from CATIE were released in 2005 and surprised many (see Table 2.1). The trial found that perphenazine was as effective as olanzapine in terms of time to discontinuation for any cause. Patients randomized to olanzapine had the longest time to discontinuation because of lack of efficacy but had the largest weight gain, as well as increases in other variables associated with the metabolic syndrome. The authors of the primary trial publication concluded that perphenazine could not be rejected as an inferior treatment.

Table 2.1**Results of the CATIE Trial**

| Outcome | Results |
|---|---|
| Primary outcome | |
| Time to discontinuation for any reason | Olanzapine was better than quetiapine and risperidone. Olanzapine was no different from perphenazine or ziprasidone. |
| Secondary outcomes | |
| Time to discontinuation for inefficacy | Olanzapine was better than perphenazine, quetiapine, and risperidone. Olanzapine was no different from ziprasidone. |
| Time to discontinuation for intolerability | There were no differences between groups. |
| Time to discontinuation for "patient reason" | Olanzapine was better than quetiapine and risperidone. Olanzapine was no different from perphenazine or ziprasidone. |
| Duration of successful treatment | Olanzapine was better than perphenazine, quetiapine, and risperidone. |
| Positive and Negative Syndromes Scale (PANSS) | There were no differences between groups. |
| Clinical Global Impressions (CGI) Scale | There were no differences between groups. |
| Discontinuation rate due to intolerable side effects | Olanzapine had the highest discontinuation rate due to intolerable side effects. |
| Quality of Life Scale | There were no differences between groups. |
| Neurocognitive effects | Small improvements occurred for all groups (not clinically significant); there were no differences between groups. |
| Hospitalization for exacerbation of schizophrenia | Olanzapine was better than all other groups. |
| Discontinuation due to weight gain or metabolic effects | Discontinuation due to weight gain or metabolic effects was highest in the olanzapine group. |
| Side effects | |

| | |
|--|---|
| Extrapyramidal side effects or movement disorders | There were no differences between groups. |
| Weight gain | The olanzapine group had greater weight gain than all other groups. |
| Metabolic syndrome (i.e., glycosylated hemoglobin, total cholesterol, and triglycerides) | Olanzapine had greater increases on all variables than all other groups. Only ziprasidone resulted in improvements on all variables. |
| Prolactin levels | Prolactin levels in the risperidone group increased. |

Impact of the Trial on Clinical Practice

While a formal evaluation of CATIE’s impact on clinical practice has yet to be completed, most experts agree that the trial did not prompt physicians to switch from prescribing second-generation antipsychotics to prescribing first-generation medications. However, several discussants reported that after the trial, many psychiatrists began for the first time to question the evidence base supporting the second-generation antipsychotics. Many recognized that they had been misled by a combination of their own optimism about newer treatments and the pharmaceutical industry’s active marketing of second-generation medications. According to one discussant, the trial made psychiatrists more open to returning to first-generation medications because it showed that conventional antipsychotics used at low doses were still effective and not associated with extrapyramidal symptoms. The extent to which this interpretation of the trial extended beyond academic psychiatrists into the community at large is unclear.

Although the relative prescribing of first-generation and atypical antipsychotics might not have changed following CATIE, the use of olanzapine might have decreased over time. Olanzapine is the most efficacious second-generation antipsychotic, but it is also associated with the highest incidence of metabolic complications, including weight gain.

New Evidence Following Initial Release of CATIE Results

Additional results from CATIE and similar trials have consistently supported CATIE’s initial findings. Findings from quality-of-life and neurocognitive substudies showed only marginal improvements across treatment groups, with no significant differences between groups. In 2006, a cost-effectiveness analysis using CATIE data found that the additional cost of atypical antipsychotics was not justified given the similar effectiveness of the first- and second-generation treatments. A second trial that compared the effectiveness of first- and second-generation medications in the United Kingdom, the CUTLASS trial, was released in 2006 and confirmed the results of CATIE across most outcomes. A paper released in 2010 showed that the incidence of tardive dyskinesia has remained relatively constant since the 1980s, despite the widespread shift to second-generation medications (Woods, Morgenstern, et al., 2010).

These findings reinforce the conclusion that the initial concerns about the extrapyramidal side effects of first-generation medications were overblown.

Key Barriers to Clinical-Practice Change

Status Quo Bias Resulting from Consistently Positive Messaging

Psychiatrist discussants indicated that at the time of the CATIE trial, the belief in the superiority of atypicals was nearly universal, for at least four reasons. First, the majority of trials preceding CATIE were funded by the pharmaceutical industry, and they consistently concluded that second-generation antipsychotics were superior. It was only following CATIE that critiques of these trials emerged in the peer-reviewed literature. Experts highlighted the fact that first-generation comparators were used at high doses, often without adjuvant therapies designed to control extrapyramidal side effects; statistical analyses used inferior methods; and superiority claims were often based on statistical significance rather than clinical significance (Carpenter and Buchanan, 2008). Second, industry detailers were highly effective in disseminating early findings that supported atypical antipsychotics and motivating their widespread use in the years preceding CATIE. The labeling as second-generation medications implied that they offered an improvement over first-generation medications. As one discussant noted, “We didn’t call beta-blockers second-generation antihypertensives.” When CATIE’s findings were released, there was a “significant effort [by industry detailers] to counter-detail the trial’s findings” according to one discussant. Third, key thought leaders, many of whom had received industry funding, reinforced these messages within both the academic community and the larger provider community. Fourth, medical education may have helped to promote the use of atypicals over first-generation medications. The consistently positive messaging about the superiority of second-generation medications raised expectations about them (particularly since there had been few advances in antipsychotic therapy over the previous several decades) and appears to have prompted many psychiatrists to view the unexpected results from CATIE with skepticism. Widespread belief about the status quo (i.e., the effectiveness of atypical antipsychotics) appears to have been one of the strongest barriers to changing prescribing practices.

Criticism of the Study Design

Possible weaknesses in CATIE’s design and its potentially limited generalizability were widely debated in the peer-reviewed literature. While many experts defended the trial, the funder (NIMH) did not participate directly in these exchanges. Three points were frequently raised. First, there were concerns that medication doses were not equivalent across treatment groups (Dettling and Angheliescu, 2006), and in some cases, the doses used were not those in standard practice. For example, olanzapine doses were higher than average, while doses of risperidone and ziprasidone were lower. Second, patients with tardive dyskinesia were excluded from receiving perphenazine, and this caused many psychiatrists to criticize the common interpretation of the trial results—i.e., that first- and second-generation medications are equivalent—because it seriously downplayed the risks of tardive dyskinesia for patients taking perphenazine. Third, some experts argued that longer-term follow-up data were needed and that the 18-month study design did not answer key long-term questions, in particular, whether the incidence of tardive dyskinesia and diabetes differed across classes. Many stakeholders felt that these critiques

were intended to draw attention away from the main results of the trial, and subsequent empirical studies debunked many of them. The consensus among our discussants was that the critiques were unlikely to have affected the decisionmaking of the average physician.

Prior Experience with Tardive Dyskinesia

Psychiatrists with prior experience managing the adverse effects of first-generation medications, particularly tardive dyskinesia, may have been less likely to embrace CATIE's findings. Physician discussants impressed upon us how debilitating and potentially irreversible this side effect can be. One physician commented, "It's a reality for those of us who grew up in that era and had severe experiences with first-generation drugs. . . . You would try to get the patient down on the lowest dose [possible] . . . to control their psychosis." In the past, before psychiatrists ever initiated therapy using first-generation medications, they would have to warn patients about these side effects in writing in order to protect themselves from malpractice lawsuits. Tardive dyskinesia may therefore remain a key concern for many in the psychiatric community, and CATIE did not directly address this issue, since patients with tardive dyskinesia were excluded from receiving perphenazine.

Lack of Positive Messaging from Professional Societies About Changing Practice

Discussants indicated that one of the chief concerns of the American Psychiatric Association (APA), the largest professional society of psychiatrists, was the preservation of physician autonomy—potentially motivated by the historic lack of insurance coverage for mental health services (an issue that has only begun to be addressed in recent years). Lack of parity between mental and nonmental healthcare coverage was mainly attributable to the limited evidence base in psychiatry. One common interpretation of CATIE's results was that neither first- nor second-generation antipsychotics were effective, and psychiatrists might have perceived this interpretation as a threat to their autonomy if the results were to be used by payers and purchasers to limit access to certain medications. The APA argued that antipsychotics were not interchangeable and that an individualized treatment approach was appropriate, which implied unlimited access to these medications despite their higher costs (Duckworth and Fitzpatrick, 2008).

Limited Impact of Changes to Guidelines

In 2009, the APA published a Guideline Watch, which updated its pre-CATIE guidelines. It stated that the "distinction between first- and second-generation antipsychotics appears to have limited clinical utility" (Dixon, Perkins, et al., 2009). The new guidelines also recommended the use of low-dose medications when first-generation agents were selected. This change was a watershed in that it was the first instance of APA guidelines explicitly stating that, as a group, second-generation medications offered no advantages over first-generation medications. This contrasts with the Schizophrenia Patient Outcomes Research Team (PORT) guidelines, which consistently acknowledged the relative equivalence of first- and second-generation medications despite the fact that it was an unpopular view within the community (Buchanan, Kreyenbuhl, et al., 2010). Neither of these guidelines appears to have had a substantial impact on clinical practice.

Limited Leverage of Payers

Given the cost differential between first- and second-generation antipsychotics, one might have expected payers to modify prescribing formularies or introduce other restrictions on the use of the latter, but public payers did not adopt stricter formularies following the trial, causing some researchers to lament that state Medicaid agencies have been “largely silent” following the release of CATIE’s results (Rosenheck, Leslie, et al., 2008). One discussant referred to the failure by payers to more actively encourage use of first-generation medications as a “lost opportunity.” A total of 21 states had introduced prior authorization programs for one or more atypicals at the time CATIE began, but only four more did so by June 2006 (Polinski, Wang, et al., 2007).

In general, payers run the risk of being shamed by advocacy organizations when they attempt to implement policies that restrict access to medications. In their messaging, advocacy groups have stressed the vulnerability of patients and their extreme price-sensitivity. Following CATIE, some experts advocated step therapy—starting initial treatment with equivalent but less-expensive drugs—with prior authorization for any deviation (Rosenheck and Sernyak, 2009), while others argued against it citing the risks of homicides or suicides as potential unintended consequences of treatment failure for this population. There is little empirical evidence about the impact of step therapy on health outcomes for patients with schizophrenia. Many payers may view formulary policy as the “third rail” of state Medicaid policy and avoid using it as a potential policy lever (Rosenheck, Leslie, et al., 2008).

While step-therapy programs might have saved states considerable sums of money, Medicaid programs have typically relied on bulk purchasing with the use of preferred drug lists as a way to control their pharmacy budgets. A discussant from an advocacy organization noted that the National Association of State Mental Health Policy Directors successfully disseminated a preferred drug list template for antipsychotics that promoted access to different types of second-generation antipsychotics because of their distinct side-effect profiles as demonstrated in CATIE. While we do not know the impact of this template, it may have preserved relatively open access to second-generation medications in many states. Private payers, particularly small payers, may have had some success in implementing step therapy, according to one government expert.

Quality Measures Favored Second-Generation Antipsychotics

Prior to the release of CATIE, SAMHSA used a performance indicator as part of its reporting requirements. The measure appears to have assessed the rate at which second-generation medications were prescribed (Covell, Finnerty, et al., 2008) and continued to be used in the years following CATIE.

Key Enablers of Clinical-Practice Change

Practice-Change Interventions

Interventions such as computerized clinical decision support and academic detailing based on CER results are often effective in changing prescribing practices, but their impact on implementation of CATIE’s findings is unclear. Our case-study results suggest that these interventions may have limits. For example, one initiative undertaken in the VA health system required physicians to fill out a

computerized form before prescribing a second-generation antipsychotic—a minor administrative hurdle designed to cause them to consider the rationale behind their choice (Rosenheck and Sernyak, 2009). Not only did it fail to change practice, it drew the attention of the pharmaceutical industry and actually resulted in new policies in the VA that forbade such administrative barriers. One prominent integrated delivery system did not even attempt a strategy of academic detailing following CATIE, because existing prescribing patterns were considered too well established for it to succeed. Despite these early setbacks, discussants representing several public and private payers reported achieving physician buy-in and a successful practice change to step therapy. In all of these efforts, prescribers were presented with data on their actual prescribing patterns, as compared with best practice. The data often demonstrated that their prescribing was much further from ideal practice than they believed. One discussant described this as applying “gentle but unyielding pressure,” because the process of gaining physician participation often took years.

Potential for Harm

In CATIE, olanzapine was associated with greater weight gain and increases in glycosylated hemoglobin, total cholesterol, and triglycerides. Despite having superior efficacy, olanzapine’s adverse-event rate combined with the elevated cardiovascular mortality risk of patients with schizophrenia (patients have a 26-year-shorter life expectancy on average) may have contributed to discussants’ perceived decrease in olanzapine use over time. While metabolic parameters can be monitored to ensure that they remain controlled, many physicians may have concluded that the metabolic side effects should be avoided at all costs (Carpenter and Buchanan, 2008). Psychiatrists might also view the decision to continue use of second-generation antipsychotics as the result of a calculation of risk tradeoffs—tardive dyskinesia from first-generation medications versus metabolic side effects from the newer medications—and not about efficacy differences at all. Many psychiatrists might favor the preservation of patients’ quality of life (controlling tardive dyskinesia) over preserving life expectancy (avoiding diabetes and cardiovascular disease). In support of this argument, some experts claim that the determinants of lower life expectancy in patients with schizophrenia are still unclear, implying that it might be related to reasons other than cardiovascular disease.

Coordination Between Government Agencies

SAMHSA is dedicated to translating evidence and disseminating best practices in the treatment of mental health and substance abuse. NIMH and SAMHSA engaged in a series of discussions following the CATIE trial, and SAMHSA’s messaging focused on enhancing patient engagement with regard to antipsychotic prescribing, because of the observed heterogeneity in benefits and harms. SAMHSA also helps to disseminate CER evidence on mental health programs through the National Registry of Evidence-Based Programs and Practices. This customizable database includes information on study design, populations studied, and outcomes (including costs) and provides references to studies that have replicated the results. Users can identify evidence that pertains to specific populations of interest. The database does not contain any information on pharmacologic interventions but theoretically provides a model for doing so in the future.

Medicaid Spending Cuts

Experts predict that anticipated cuts in state Medicaid spending may accelerate the switch to generic antipsychotics, and because all first-generation medications are available in generic form, this may effectively accelerate the adoption of step therapy. It is possible that formulary policies that were unpalatable several years ago might be tolerable during times of shrinking public budgets.

Adverse-Event Surveillance

The CATIE trial was criticized because its time frame was not long enough to assess the rate of adverse events—particularly the incidence of tardive dyskinesia—with first-generation medications or the incidence of diabetes or metabolic syndrome with second-generation anti-psychotics. Adverse-event surveillance programs may help to resolve questions about the long-term safety of antipsychotics. Current initiatives, such as the FDA’s Sentinel Initiative to develop the ability to use distributed-data networks to track adverse events from drugs and devices, may therefore be an enabler of practice change in the near future.

Timely CER

CATIE took more than five years to complete. Timely CER is critical, because the treatments being compared are likely to have been in use for years, and practice patterns can become fixed, as they did in antipsychotic prescribing. Experts have proposed trial designs, including adaptive designs, that allow new treatments to be integrated into studies as they progress. This could greatly improve the efficiency of these trials and protect against the criticism that their findings are outdated because of the use of treatments that do not reflect current practices. CATIE used an adaptive design by allowing patients to be randomized to ziprasidone after it became FDA approved. More timely CER using these new trial designs may represent a future enabler of clinical-practice change.

State and Local Regulatory Policy Restricting Pharmaceutical-Industry Influence on Prescribers

A number of states and medical centers have adopted policies to limit the influence of the pharmaceutical industry. Several states have passed laws barring physicians from accepting gifts, and as of December 2009, nine states (California, Florida, Maine, Massachusetts, Minnesota, New Hampshire, South Carolina, Vermont, West Virginia) and the District of Columbia had laws or resolutions governing pharmaceutical marketing practices (National Conference of State Legislatures, 2010). Many academic medical centers have implemented similar policies, and others that are even more far-reaching bar detailers altogether. The effectiveness of these policies and the prospects for their continued diffusion are not known.

Conclusions

Our case study of the CATIE trial highlighted the role of pharmaceutical manufacturers in shaping and reinforcing beliefs about the relative superiority of second-generation antipsychotics, both directly (through marketing and detailing) and indirectly (through key thought leaders), well in advance of the conduct of a CER study. By the time the CATIE results were released, these efforts had succeeded in

cementing beliefs about the different classes of antipsychotics, and practice patterns do not appear to have changed in the five years since. Professional societies did not strongly advocate for practice changes based on CATIE’s results (except through an eventual change to guidelines, which appears to have had limited impact). Performance measures were not updated to reflect the trial’s findings and may have continued to reinforce existing prescribing patterns. Professional societies and advocacy organizations challenged the results of the trial in an effort to protect provider autonomy and preserve access to medications, respectively. Public payers were initially unwilling to enact policies that might limit the treatment options of patients with schizophrenia given the relative lack of access to care for this population and the potential backlash from advocacy organizations.

A number of strategies might be used to promote the uptake of results of CER studies similar to CATIE into clinical practice. Methodological choices (such as the exclusion of patients with tardive dyskinesia from the perphenazine group) may limit the perceived generalizability of findings and cause physicians to distrust results. In the case of tardive dyskinesia, providers had strong prior negative experiences involving adverse outcomes. Failure to design a trial to test one of the main beliefs driving use of a treatment meant that an issue important to prescribing providers might be perceived as being inadequately addressed.

Interpretation and formalization of CATIE’s results faced formidable but predictable difficulties given the strength of established beliefs about the efficacy and safety of second-generation antipsychotics. Indeed, it proved very difficult to change the deeply ingrained belief system founded on industry-funded studies. Interestingly, critiques of the study methodology that did not address harms may not have significantly influenced prescribing practices. Likewise, it should be borne in mind that professional societies can be expected to generate guidelines reflecting their professional interests if study results enable such an interpretation by failing to have a clear winner. Timely updates to quality measures that reflect the new CER evidence may prove critical to motivating early changes in practice.

Dissemination and implementation strategies should be vigorous and multiple. Academic detailing within closed systems eventually proved effective in the case of CATIE, but early efforts were constrained by doubts about its value and even a clinical decision support prompt faced initial resistance. Presenting physicians with their actual practice data, which often showed how far they diverged from the ideal, was key to success. Finally, future adverse-event surveillance systems (or registries) may help to resolve lingering questions about relative side-effect risks among patients taking alternative antipsychotics.

Key findings from the CATIE case study are summarized in Table 2.2.

Table 2.2

Key Findings from the CATIE Case Study

| Phase | Key Findings |
|-------|--------------|
|-------|--------------|

| | |
|----------------|---|
| Generation | <ul style="list-style-type: none"> • The trial design downplayed the risk of tardive dyskinesia (a severe harm) in patients taking the first-generation drug. • Follow-up may have been too short to capture true side-effect incidence. • There was a perception that medication dosages did not reflect typical practice. |
| Interpretation | <ul style="list-style-type: none"> • Prior experience with tardive dyskinesia from first-generation drugs biased practitioners against the CER results. • Critiques of the study design (later mostly debunked) may have obscured the primary CER result but may not have affected prescriber decisions. • Widespread belief in the superiority of second-generation drugs (driven by uniformly positive messaging from the pharmaceutical industry) enhanced skepticism of CER results. • Labeling of drugs as “second-generation” promoted belief in their superiority. |
| Formalization | <ul style="list-style-type: none"> • Professional society guidelines were driven by a desire to maintain practitioner decisionmaking autonomy. • Guidelines reflected a concern that payers would limit access to more-expensive medications if the drugs were judged either interchangeable or ineffective. • Quality measures continued to favor the use of second-generation medications. |
| Dissemination | <ul style="list-style-type: none"> • Changes in professional-society guidelines (stating that treatments were equivalent) appear to have not significantly affected practice. • Academic detailing was forestalled by a belief that prescribing patterns were already too entrenched. • Regulatory and administrative restrictions on industry marketing may have favorably altered prescribing patterns. |
| Implementation | <ul style="list-style-type: none"> • Advocacy groups and the pharmaceutical industry vigorously opposed any prescribing guidelines interpreted as step-therapy policies. • Payers nevertheless have increasingly instituted step-therapy guidelines promoting initial treatment with equivalent but less-expensive drugs. • Medicaid budget cuts may be accelerating a move toward initially prescribing equivalent generic (mostly first-generation) drugs. • Buy-in usually follows when physicians are presented with actual practice data, which often demonstrate wide divergence from ideal practice. |

- The CER results demonstrated the need for practitioners to weigh alternative side effects (e.g., quality of life versus longevity).
-

COURAGE Case-Study Report

Clinical-Practice Context

Treatment guidelines for patients with stable coronary artery disease (CAD) recommend an initial strategy of intensive medical therapy along with risk-factor reduction and initiation of lifestyle interventions, collectively known as optimal medical therapy (OMT). However, over the past 30 years, percutaneous coronary intervention (PCI) has increasingly been used as an initial strategy for patients with stable CAD. Approximately 85 percent of all PCI procedures are performed on patients with stable CAD (Boden, O'Rourke, et al., 2007); the remaining 15 percent are performed on patients with acute coronary syndrome (ACS), an emergency condition. PCI is also indicated for the relief of angina (chest pain); however, a large percentage of patients who undergo elective PCI are asymptomatic (Diamond and Kaul, 2007).

The increase in the use of PCI for patients with stable CAD appears to have been prompted by a variety of factors. Cardiologists may have extrapolated from the successful use of PCI for ACS (Boden, 2007). Although the procedure carries some complication risks, including mortality, these risks are generally perceived to be quite low (Lin, Dudley, et al., 2007). The introduction of intracoronary drug-eluting stents may have further encouraged the use of PCI in stable CAD patients by lowering the risk of restenosis. Other technological innovations, including new screening tests for CAD, may have increased the number of patients diagnosed with the condition. Studies show that these tests are often performed for patients who are asymptomatic (Diamond and Kaul, 2007; Lin, Dudley, et al., 2007).

Studies have documented wide variations in PCI use associated with patients' sex and race and between geographic regions, suggesting that referral and treatment decisions may be influenced by factors other than clinical parameters (Lin, Dudley, et al., 2007). Financial incentives are likely to be a potent driver of PCI utilization. The Medicare fee-for-service payment schedule provides generous incentives for PCI, while reimbursement for the management of OMT is relatively less well compensated.

The Comparative-Effectiveness Question

The fundamental question motivating the COURAGE trial was whether using PCI to reverse flow-limiting stenoses (narrowing of coronary arteries) would reduce the risk of MI and death among CAD patients more than OMT would. While PCI had been shown to provide substantial benefit for patients with ACS, its benefit for the stable CAD population had not been conclusively demonstrated. Previous studies had

shown that PCI could decrease angina frequency and improve short-term exercise performance in that population, but it did not clearly reduce cardiovascular-event frequencies over either short or long time frames. However, most of the clinical trials that assessed these end points were small and underpowered (Boden, O'Rourke, et al., 2007), enrolling a total of only 1,872 patients (Diamond and Kaul, 2007). A meta-analysis of trials among patients with stable CAD published in 2005 showed that PCI achieved no reduction in acute coronary events or death compared with OMT (Katrakis and Ioannidis, 2005).

The superiority of PCI in improving quality-of-life outcomes for patients with stable CAD had also not been demonstrated conclusively. Few large-scale studies had addressed this question, and moreover, the treatments used in prior studies were already out of date, including neither the current treatment approaches for medical management nor the use of intracoronary stents for patients who received PCI.

Study Design Characteristics

The COURAGE trial compared the risk of cardiovascular events among patients with stable CAD assigned to a treatment strategy of intensive pharmacologic therapy and lifestyle intervention (OMT) alone with those assigned to treatment with PCI followed by OMT. No previous studies involved the intensity of OMT attempted in COURAGE, which included the use of aspirin, beta-blockers, ACE inhibitors, statins, and clopidogrel, as well as diet, exercise, and smoking-cessation counseling. Medication doses were repeatedly intensified in pursuit of aggressive blood-pressure and LDL-cholesterol targets.

Previous CAD trials suffered from limited generalizability because they enrolled highly selected patient populations. Patients were commonly under the age of 65, less likely to have depressed left ventricular function, less likely to have clinical instability, and less likely to have undergone previous coronary artery bypass graft (CABG) surgery or PCI than patients typically seen in nonexperimental settings (O'Rourke, 2008). COURAGE was designed to enroll a moderate- to high-risk population with less-restrictive inclusion criteria.

The primary outcome for the efficacy study was a composite outcome of death from any cause and nonfatal MI. A quality-of-life substudy assessed changes in angina frequency and quality of life using the Seattle Angina Questionnaire and the RAND-36 health survey. The trial's "nuclear substudy" compared the effectiveness of PCI with or without OMT in reducing the frequency of results suggestive of ischemia among a subset of patients who underwent myocardial perfusion single-photon emission computed tomography (MPS) (Shaw, Berman, et al., 2008).

Study Results

The primary COURAGE trial results were published in 2007 (see Table 3.1). As an initial management strategy in patients with stable CAD, PCI did not reduce the risk of death, MI, or other major cardiovascular events when added to OMT. These findings reinforced existing practice guidelines, which stated that PCI can be safely deferred in patients with stable CAD, even in those with extensive,

multivessel involvement and inducible ischemia, provided intensive, multifaceted OMT is instituted and maintained.

Both treatment groups—those who had PCI and those who did not—had marked improvements in health status during follow-up. The PCI group had small, but significant, lowering of angina prevalence and improved quality of life for about 24 months, but the relative advantage of PCI disappeared by 36 months (Weintraub, Spertus, et al., 2008). A cost-effectiveness analysis using patient-level data from the trial found that the addition of PCI to OMT cost from \$168,000 to \$300,000 per life-year or quality-adjusted life-year gained and thus offered little value based on standard benchmarks for cost-effectiveness (Weintraub, Boden, et al., 2008). In the trial’s nuclear substudy, patients assigned to the PCI group had a greater reduction in ischemia, and patients with moderate or severe pretreatment ischemia appeared to have better outcomes for the primary end point (death or MI), although the results were not statistically significant.

The COURAGE trial was not alone in suggesting that more-aggressive procedure-based treatment approaches might provide no greater benefit than OMT. The Occluded Artery Trial (OAT) published in 2006 (Hochman, Lamas, et al., 2006) and the Bypass Angioplasty Revascularization Investigation 2 Diabetes (BARI 2D) trial published in 2009 also suggested that interventional procedures might not be better than OMT (Frye, August, et al., 2009). The OAT trial showed that patients in whom revascularization was attempted 3 to 28 days after MI (traditionally considered beyond the time frame for myocardial salvage) had no better outcomes (and potentially fared worse) than patients who were treated with OMT. In the BARI 2D study, stable CAD patients with diabetes were randomized to cardiac bypass surgery, PCI, and intensive medical therapy, and the authors found no difference in survival or cardiac events between treatments.

An alternative interpretation of the trial results suggested that the nuclear substudy finding suggesting superiority of PCI over OMT among patients with higher levels of baseline ischemia was actually the most important finding, as this result partially confirmed similar findings from observational studies. While the impact of the substudy findings on practice is still unclear, they have the potential to promote better tailoring of PCI use to patients’ underlying risk of future cardiac events.

Table 3.1

Results of the COURAGE Trial

| Outcome | Results |
|--|--|
| Primary Outcome | |
| Composite: death from any cause <i>and</i> nonfatal MI | No difference: 19% (PCI) versus 18.5% (OMT), p = NS. |
| Secondary Outcomes | |

| Outcome | Results |
|---|---|
| Composite: death, nonfatal MI, stroke, and hospitalization for unstable angina with negative biomarkers | No difference: 20% (PCI) versus 19.5% (OMT), p = NS. |
| Hospitalization for ACS | No difference: 12.4% (PCI) versus 11.8% (OMT), p = NS. |
| MI | No difference: 13.2% (PCI) versus 12.3% (OMT), p = NS. |
| Stroke | No difference: 2.1% (PCI) versus 1.8% (OMT), p = NS. |
| Other | |
| Additional revascularization ^a | 21.1% (PCI) versus 32.6% (OMT), p < 0.001. |
| Percentage angina-free | Higher in the PCI group through 24 months but not statistically different at 36 months (final measurement). |
| Seattle Angina Questionnaire | Scores improved in both groups; scores were higher in the PCI group from 6 to 24 months for most domains, but by 36 months, the PCI group no longer provided a significant advantage for any domain. |
| Physical limitations | |
| Change in angina severity | |
| Frequency of angina | The PCI group had more “clinically significant improvements” in physical function, angina frequency, and quality of life for the first 6 months, but these differences were no longer significant by 12 months. |
| Satisfaction with treatment | |
| Quality of life | |
| RAND-36 | Scores improved in both groups; scores were higher in the PCI group at 3 months for most domains, but after 12 months, PCI provided no significant advantage for any domain. |
| Physical functioning | |
| Role limitations due to emotional problems | The PCI group had more “clinically significant improvements” in physical functioning and role limitation due to physical problems at 6 months, but these differences were no longer significant by 12 months. |
| Vitality | |
| Emotional well-being | |
| Social functioning | |
| Pain | |
| General Health | |

| Outcome | Results |
|--|--|
| Sensitivity analysis: effect of crossovers | Patients assigned to the OMT group who crossed over had changes in quality-of-life outcomes similar to those of patients who did not cross over. |
| Sensitivity analysis: complete cases | 328 (PCI) and 303 (OMT) patients had complete data at 36 months; there were no differences in results. |
| Subgroup analyses (efficacy outcomes) | There were no significant interactions by subgroup. |
| Subgroup analyses (quality-of-life outcomes) | There was interaction between the treatment and baseline tertile of Seattle Angina Questionnaire scores for the following domains: physical limitation, angina frequency, and quality of life (in favor of PCI). |

^aRevascularization was performed for angina that was unresponsive to OMT or when there was objective evidence of worsening ischemia on noninvasive testing, at the discretion of the patient's physician.

Impact of the Trial on Clinical Practice

A recent study examined the impact of COURAGE using data from the National Cardiovascular Data Registry (NCDR) and found that rates of OMT prior to PCI did not change following the trial (Borden, Redberg, et al., 2011). A recently released analysis from the Northern New England Cardiovascular Disease Study Group indicates that PCI rates among patients with stable CAD declined 26 percent shortly after the trial results were released (Neale, 2011). However, northern New England is thought to have more conservative practice patterns than other parts of the country, and these results may not reflect national trends.

Prior to the release of Borden's study, our discussants generally thought that the trial may have been successful in changing practice by promoting a more "conservative" style. One expert said that the trial "took the pressure off interventionists from doing procedures." Others felt that the intense publicity surrounding the trial put the issue of overuse of elective PCI in the spotlight, and this negative publicity drove interventionists to adopt a more conservative practice style. Still others appear to have interpreted the success of COURAGE less in terms of the absolute change in PCI rates than in how the trial has contributed momentum to the development and use of appropriateness criteria (discussed below).

Key Barriers to Clinical-Practice Change

Study Design and Conduct

A number of aspects of the trial's design were criticized by experts:

1. Some interventionists perceived that the trial was designed with the objective of disproving that PCI was beneficial. This may have polarized opinions even before the results were released.
2. There was controversy about whether or not COURAGE enrolled a population with lower risk than the typical population of patients with stable CAD. In the view of some experts (mainly interventionists), the timing of randomization—following angiography—guaranteed that high-risk patients would be triaged immediately to PCI rather than being enrolled in the trial, leaving a population of lower-risk patients to be enrolled. These experts cited the cohort’s average ejection fraction (60.8 percent) and low cardiac death rate (0.45 percent/year) (Bangalore and Messerli, 2007; Tommaso, 2008) as evidence of the cohort’s low risk. Others argued that COURAGE enrollees represented a relatively high-risk population, noting that their five-year MI rates were similar to those of patients with ACS—a condition associated with a high mortality rate (Diamond and Kaul, 2007). Some compared the risk profile of COURAGE enrollees to the average risk of patients undergoing elective PCI in the NCDR and found them to be comparable (Peterson and Rumsfeld, 2008).
3. Most critically, because randomization occurred after angiographic findings were known, the trial appeared to be designed to inform decisionmaking only after angiography. In practice, however, diagnostic angiography and PCI are often performed in tandem without an opportunity for shared decisionmaking between the two procedures. The trial did not explicitly address decisions that may occur before angiography. Experts suggested that trials designed to explicitly address the upstream decisionmaking process (prior to angiography) would be more likely to have an impact. Trials using this design are currently under way and are considered to be “more promising.”
4. The trial may have been underpowered. The intent was to detect a survival difference of 22 percent—the expected benefit of PCI in patients with ACS—between groups (Diamond and Kaul, 2007; Tommaso, 2008). However, some interventionists felt that detecting this difference in a population with stable CAD was unrealistic, because of the low severity of illness of COURAGE enrollees.
5. Crossover was significant; 33 percent of patients randomized to the OMT group crossed over and received PCI. The relatively high rate of crossover meant that the apparent equivalence in primary outcomes could be due to the effectiveness of OMT or to the added benefit of PCI among those patients who crossed over. Thus, this may have affected stakeholders’ interpretations of the trial’s results. Crossover rates are frequently higher in CER trials involving surgery. In the second Medical, Angioplasty, or Surgery Study (MASS-II), which compared CABG surgery, PCI, and OMT, the crossover rate was nearly 24 percent (Diamond and Kaul, 2007).
6. Some criticized the poor quality of PCI procedures performed during the study. Experts pointed to the high revascularization rate in the PCI group (21 percent) and the low rate of multiple-stent use. Critics noted that only 36 percent of patients received more than one stent, even though 70 percent had two-vessel disease (Kereiakes, Teirstein, et al., 2007). Others said that this frequency is comparable to the experience of patients elsewhere, including those within the New York state angioplasty registry (Diamond and Kaul, 2007). Critics also noted that PCI was

successful in only 89 percent of cases whereas a 95-percent success rate better reflects current practice (Prasad, Rihal, et al., 2008).

7. Other aspects of the trial raised questions about the validity and/or generalizability of its findings. The strict enrollment criteria of COURAGE might also have reduced the generalizability of the results. The small number of patients enrolled (2,287, only 6.4 percent of the 35,539 patients screened) suggested that the results had limited generalizability (Kereiakes, Teirstein, et al., 2007), and some criticized the absence of any data characterizing the non-enrolled patients. However, similar trials, including MASS-II, also had low enrollment rates (2.9 percent of 20,769 screened patients). The fact that drug-eluting stents were not used in the trial was also cited as a design flaw. While many experts agree that their inclusion would not have changed the frequency of primary trial end points (Bhatt, 2007; Boden, 2007), drug-eluting stents might have improved the performance of PCI with respect to angina symptoms and quality-of-life outcomes (Diamond and Kaul, 2007). A sensitivity analysis conducted as part of the cost-effectiveness analysis showed no impact of the use of drug-eluting stents (although that study was performed on a limited sample) (Kirtane and Cohen, 2008).

Financial Incentives

Financial incentives are widely believed to be a potent driver of PCI use. While hospital and physician fees for elective PCI may average \$20,000 per case, managing drug therapy and delivering lifestyle counseling are poorly reimbursed. One CER researcher commented that “[doctors] are paid a lot to do an angioplasty and nothing to talk about treatment options.” There appear to be few countervailing incentives that would reduce the use of PCI for patients with stable CAD. Interventional cardiologists may recommend PCI because they believe that other cardiologists and primary-care physicians expect them to do so and that advising against the procedure might jeopardize future referrals. Most experts indicated that few patients fail to receive PCI once they are referred to an interventionist. Competition for patients between interventionists and cardiac surgeons may reinforce these patterns. Elective PCI is also a key revenue source for hospitals. The emergence of for-profit heart hospitals has been driven, in part, by the high margins from elective PCI procedures. To date, few payers have imposed significant barriers to the use of PCI.

Psychological Issues Driving PCI Use

The strong urge to open all significant coronary lesions amenable to PCI once they have been detected through angiography has been referred to as the “oculostentotic reflex” (Lin, Dudley, et al., 2007), which suggests that psychological factors may drive the use of PCI. The published literature and expert accounts describe at least three specific drivers. First, the procedure-based solution to the problem of CAD addresses a physician’s desire to “do something,” while OMT is perceived as inaction. Eliot Freidson made the observation many years ago that physicians often prefer action to inaction even when there is little chance of success (Freidson, 1970). In many cases, the desire to act is a joint sentiment. Patients often desire an active treatment approach; to some, chest pain may provoke significant anxiety even if it has a stable pattern (Lin, Dudley, et al., 2007). Finally, bad outcomes leave a strong impression on

physicians, especially if an action was not taken. One or two negative personal experiences may contribute to a physician pursuing more-aggressive management of CAD (Lin, Dudley, et al., 2007).

Bias Against OMT

Historically, the value of PCI relative to OMT was perceived to be high, because OMT had limited efficacy. However, our discussants described OMT today as being “far superior” to past therapy. Some speculated that the low event rates in trials might be due to the efficacy of modern OMT (O’Gara 2010). While physicians may be aware of the benefits of modern OMT, patients may be less convinced because they are difficult to observe directly. Physicians are also aware that adherence to OMT is relatively poor (Kereiakes, Teirstein, et al., 2007), because of cost barriers and other factors (O’Gara, 2010). Thus, many may view the level of OMT attained in the COURAGE trial as unrealistic and not achievable in practice.

Lack of Change to Practice Guidelines

In the years preceding COURAGE, guidelines published by the American College of Cardiology (ACC), the American Heart Association (AHA), and the Society for Coronary Angiography and Interventions (SCAI) consistently recommended OMT as the favored first approach for patients with stable CAD (Boden, 2008). While COURAGE might have potentially strengthened that recommendation, no major changes were made to practice guidelines in the four years since the findings were released, although an updated set of guidelines has been in development for more than three years. Despite the rigorous guideline development process at the ACC, experts feel that the societies do a poor job of disseminating and promoting adherence to cardiology guidelines. While provocative, the nuclear substudy findings from COURAGE were not seen as definitive and therefore did not lead to updated guidelines.

Duration of the Trial

Enrollment of patients for COURAGE began in June 1999, and the results were published in April 2007, a span of nearly eight years. Over the course of the trial, standards of practice changed dramatically. In particular, drug-eluting stents became the standard of care, despite the fact that few enrollees received them. As mentioned earlier, some experts therefore questioned the generalizability of the trial, while others showed empirically that this limitation did not affect the results. One discussant, noting that PCI technology had evolved during the past decade, bemoaned the fact that the use of outdated procedures is “*always* the first critique of studies that fail to show a benefit for a procedure.” A recently proposed trial to validate the nuclear substudy findings in COURAGE is projected to take between seven and eight years to complete and may pose similar challenges. Some argue that the increasing use of international clinical trials may help to shorten the time required to release results and will thus improve their relevance. The cardiologists we spoke with emphasized that the appropriate role of observational CER studies in this area of cardiology has not been clearly articulated.

Key Enablers of Clinical-Practice Change

Changing Models of Physician Organization and Payment

The past several years have seen a dramatic shift in the way physicians are organized—away from privately owned solo and small practices to employment in larger, organized practice groups. More than

50 percent of the physicians in the United States are employed by hospitals or practice in integrated delivery systems, and the trend is increasing (Kocher and Sahni, 2011). Younger physicians view employment by health systems as providing greater lifestyle flexibility, albeit at a potentially lower salary (Harris, 2010). At the same time, the administrative costs of maintaining a small private practice, including those of bill collection and transitioning to electronic health records, are increasing. The move to risk-based payment such as that proposed for integrated delivery systems and other accountable-care organizations (ACOs) may also drive this trend as practice consolidation enables more effective management of financial risk (Kocher and Sahni, 2011). The chief executive officer (CEO) of the ACC estimated in 2010 that Medicare's 2009 cuts to payments for cardiologists may have reduced the share of cardiologists working in private practice by half in only one year (Harris, 2010).

If the shift from private practice to large group practice entails a shift to a salaried payment model, organizations that bear financial risk may have strong incentives to maximize the use of OMT and target the use of PCI to patients who are most likely to benefit. These programs may involve appropriateness reviews, performance measurement, or financial incentives and thus may reduce the likelihood of using PCI for patients with stable CAD.

Conflicting Interpretations of Results by Professional Societies

Paradoxically, divisiveness among subspecialists may have inadvertently helped disseminate the COURAGE results. Many noninterventional cardiologists were concerned about overuse of PCI in the years preceding publication of the COURAGE findings. They therefore embraced the findings, while interventionists tended to criticize the study as flawed. At scientific meetings where the results were discussed, divisions between the specialties were observed to be particularly acute. Journalists took note, and a number of media stories appeared, raising questions among the public about the appropriateness of PCI.

Refined Appropriateness Criteria and New Efforts to Increase Their Use

Appropriateness criteria for the use of coronary revascularization existed before COURAGE, and most of our discussants believed the trial did not lead to the updates of the criteria that were published in 2009 (Patel, Dehmer, et al., 2009). Nevertheless, appropriateness criteria may be an important step toward reducing variation in the use of PCI, and COURAGE may have stimulated interest in integrating their use into clinical practice. Our discussants identified few contexts in which appropriateness criteria were being used consistently by providers as part of a quality improvement program or payment initiative. Some insurers, however, may be starting to embrace appropriateness criteria for diagnostic imaging procedures, which constitute a key step in the diagnostic pathway leading to PCI. Payers have come under fire for using radiology benefits managers (RBMs) with strict prior authorization requirements for cardiac imaging. Blue Cross Blue Shield (BCBS) of Delaware, for example, has been the subject of investigation for contracting with an RBM that has denied nuclear stress tests even when the decision did not agree with the ACC's appropriateness criteria (Miller, 2011). The health plan is now moving to implement a clinical decision support tool that provides direct feedback to ordering clinicians based on appropriateness criteria (Mississippi Chapter of the ACC, 2011).

Some of our discussants were enthusiastic about the role of appropriateness criteria, while others argued that they have limited clinical utility because of the high proportion of indications rated as “uncertain.” One expert suggested that appropriateness criteria are not more widely used because only 40 percent of the indications in current appropriateness criteria are associated with objective evidence and define with clarity what is appropriate or inappropriate.

Growing Prominence of Clinical Registries

Numerous cardiovascular disease registries have emerged and evolved over the past several decades into large-scale quality improvement initiatives that many experts consider to be important potential drivers of practice change. For example, the NCDR Cath/PCI registry has updated its data collection form to be able to capture elements needed to rate the appropriateness of PCI procedures. Providers participating in this registry will soon receive feedback reports on the appropriateness of their cardiac interventions, along with benchmarking data. This will be the first time this type of reporting has been performed on such a large scale and at regular intervals. Efforts are under way within the VA health system to implement a similar technology. The ability to provide this type of information to providers is seen as a key element of outpatient quality improvement initiatives, and it can also help to improve the validity of the appropriateness criteria. Experts indicate that registries will become much more useful for CER as clinical end points are rigorously collected and audited.

Unlike their inpatient counterparts, which have a much longer history, outpatient cardiac registries are only now gaining in prominence. The ACC Pinnacle registry, which contains data from 700 providers, is a key source of data for improving the quality of outpatient care in cardiology—not least because it includes data on utilization both before and after patients undergo coronary intervention. The registry has evolved and now provides feedback reports at the individual-provider level at regular intervals; it is moving to a dashboard model in the next few months that will allow providers to have even more timely data for quality improvement.

Online Dissemination Tools

Experts have pointed to a wide range of online resources cardiologists can use to learn about new research, including Medscape, theheart.org, and Cardiosource (which is sponsored by the ACC). These websites not only present data from new trials but may offer features that are particularly valuable for the dissemination of research findings. The editors of Cardiosource are currently developing a point/counterpoint discussion between two experts about findings from research studies and their implications for practice. The intent is to address controversies head on, to discuss the limitations of each study, and to determine the types of patients to which the results apply. The website is also developing an interactive forum in which physicians can engage in discussions about research findings. Neither of these features is linked with continuing medical education (CME) credits; rather, they are designed for clinicians interested in a deeper understanding of the research and its implications for practice.

Growing Availability of Patient Decision Aids

Previous studies have shown that a substantial percentage of patients believe incorrectly that PCI improves their prognosis (Diamond and Kaul, 2007). Many cardiologists we spoke with believe that patients may not be adequately informed about the benefits of modern OMT, and if they were, many might decide to forgo PCI. Decision aids developed using CER results and other evidence may create a context for improved decisionmaking by patients. Although currently there are few financial incentives to support the use of patient decision aids, experts acknowledge that they would be more likely to be used if an ACO model were adopted.

Shared-decisionmaking demonstration projects are under way, but the approach appears to have made few inroads in cardiology. Nevertheless, stakeholders we spoke with indicated that two separate coalitions are seeking to launch shared-decisionmaking initiatives for patients with stable CAD. These efforts are taking place within systems where payment models are conducive to such a program. Patient decision aids are being used in some interventional cardiology practices to generate individualized predictions of risks and benefits based on data extracted electronically from a provider's electronic health record (EHR). While this level of sophistication may not be necessary to successfully implement these types of decision aids, few practice settings would be able to implement such tools today. Apart from shared-decisionmaking aids, patient portals such as the ACC's Cardiosmart.org are available to inform patients generally of the risks and benefits of different cardiac tests and procedures. However, the frequency with which patients use this information and its impact are unclear.

Increased Attention to Risks of Procedures

While there is broad consensus that PCI is relatively safe, cumulative exposure to radiation from various tests and imaging procedures is a growing concern. Performance measure developers such as the AMA Physician Consortium for Performance Improvement (PCPI) have begun discussing quality measures relating to assessment of radiation exposure, and implementing such measures might promote more awareness of these harms in the future. In addition, an editorial accompanying the release of the COURAGE trial's quality-of-life results highlighted the small, but real, mortality risks associated with PCI. CER trial results that include data on the risks of each strategy, in addition to comparing the marginal benefit of each, may be useful for physicians and patients, who may assign different values to various benefits and risks.

Experimentation with Value-Based Insurance Design

Value-based insurance design is a strategy for increasing the use of high-value medical care through the use of cost-sharing levels that are inversely proportional to the amount of clinical benefit. In such schemes, copayments for OMT might be lower than those for PCI or might be waived altogether. For example, Aetna and other insurers have offered free preventive medications to diabetes and heart-disease patients (Diamond and Kaul, 2007). In the past, this approach has not been widely adopted, but one large insurer is currently laying the groundwork to implement value-based benefit plans, in which reimbursement levels would be linked to the most cost-effective treatment, and patients would pay the

difference. The insurer is planning to implement this policy across a wide range of clinical conditions, all of which have strong enough evidence to justify its use.

Conclusions

Our analysis of the COURAGE trial and subsequent events suggests that the findings did not have an impact on clinical practice despite speculation that physicians might have pursued a more conservative approach to the management of stable CAD following the trial. The trial may have an important indirect effect on practice by encouraging the integration of appropriateness criteria for coronary revascularization into decision support tools that can be updated as new CER emerges. Efforts to use appropriateness criteria in quality improvement are nascent, and while they have yet to be used in an accountability or payment context, there is increasing interest among policymakers in pursuing them. How much these efforts will facilitate practice change remains unclear, but it seems likely that integration with cardiac registries and incorporation into decision support tools at the point of care could make a difference. Such initiatives will have greater effectiveness once reimbursement systems create demand for them. Changes in the organization of cardiology practices, driven in part by the movement toward ACO-based payment models, may be the single most important determinant of the future adoption of findings from COURAGE.

Several strategies may improve uptake for CER trials that share some of the characteristics of COURAGE. In the generation phase, the research focus should be on a decision point sufficiently upstream to impact decisionmaking meaningfully. A critical driver of the use of PCI is the initial decision to refer a patient to an interventionist, since this tends to create an expectation that angiography and PCI will follow. The COURAGE trial did not address the initial referral decision directly. Rather, it addressed a later decision point—after patients have already undergone angiography—at which the utility of decision support and patient-decisionmaking aids may be suboptimal. Current and proposed trials are focusing on decisions that occur prior to angiography, and these may have a greater impact on clinical practice. Other design problems to avoid include the potential for significant patient crossover and excessive time to complete studies. However, discussions with stakeholders suggest that these criticisms of the COURAGE trial design are likely to have played only a minor role in influencing practice patterns.

Interpretation and formalization can languish if study findings confirm current guidelines, even if they contradict current practice. Prior to COURAGE, practice guidelines were based on very weak evidence, promoting physicians' inclination to disregard them, but since the COURAGE results reinforced the guidelines, there was less impetus to revise them. A CER result that necessitates a change in guidelines may have more impact. Similarly, payers and other stakeholders must have the ability to collect relevant appropriateness data, or they will have no incentive to develop reimbursement policies or quality measures.

Dissemination and implementation may be either advanced or hindered by several factors, but in this case, psychological aspects appear to be key. While registries may have influenced practice (by incorporating performance measures and appropriateness criteria into their design), their influence on

appropriate use of elective PCI appears modest. Similarly, payer limits on upstream diagnostic procedures may have somewhat dampened demand for PCI, as might accountable-care reimbursement schemes in the future. Psychological factors, including concerns about harm and physician response to popular media coverage regarding PCI overuse, may modulate the tendency to intervene aggressively, but strong financial and psychological factors still incline both providers and patients to favor PCI. As one discussant put it, even without financial incentives, “interventionists love to intervene.” Patients may underestimate the effectiveness of optimal OMT, and patients may not be informed of, fully understand, or have access to available information on the benefits and risks of PCI. Patient and clinician decision aids may play a key role in helping to remedy this. However, to be maximally effective, such decision aids will have to be implemented in settings where financial incentives do not promote PCI and before patients have progressed along the referral pathway to the point where intervention becomes almost inevitable.

Key findings from the COURAGE case study are presented in Table 3.2.

Table 3.2**Key Findings from the COURAGE Case Study**

| Phase | Key Findings |
|----------------|---|
| Generation | <ul style="list-style-type: none">• There is a perception that the study design led to enrollment of low-risk patients, thereby limiting generalizability.• Focus on decisionmaking post-angiography is low-leverage; CER on upstream decisionmaking is seen as potentially more influential.• Patient crossover between treatments muddles the interpretation of findings.• The study was arguably underpowered, but it also took eight years to complete. |
| Interpretation | <ul style="list-style-type: none">• The multiple professional societies involved allowed different interpretations to persist.• A perception that study treatments were already outdated affected interpretation. |
| Formalization | <ul style="list-style-type: none">• The findings confirmed existing guidelines (but not current practice), so no additional formalization occurred initially.• Health plans are able to collect only limited data on PCI indications, limiting their ability to monitor appropriateness or inform payment policy. |
| Dissemination | <ul style="list-style-type: none">• Criticisms of methodology in the peer-reviewed literature promoted popular media coverage.• Specialty-society guidelines, while developed rigorously, were not promoted energetically.• Registries may help disseminate findings through performance measurement and use of appropriateness criteria. |
| Implementation | <ul style="list-style-type: none">• Reimbursement significantly favors PCI.• Referral to an interventionist can be interpreted as endorsement of the procedure.• Psychological factors in both patients and proceduralists drive a desire to fix all stenoses.• Payer limits on upstream diagnostic procedures, while controversial, may limit inappropriate use of PCI.• The move away from private practice in favor of affiliation with medical centers may alter financial incentives.• Potential harm from radiation exposure is an emerging issue that may dampen overuse. |

- Using patient decision aids covering treatments for angina may improve outcomes, but physicians need incentives to use them.
-

SPORT Case-Study Report

Clinical-Practice Context

Spinal stenosis is a narrowing of the vertebral canal that compresses spinal nerves and may cause back and leg pain and difficulty walking. Lumbar spinal stenosis is one of the most common degenerative conditions of the spine and is the most common reason for lumbar spine surgery on the elderly (Weinstein, Tosteson, et al., 2008). Treatment of spinal stenosis may involve nonsurgical care, decompression surgery (involving the removal of bone and ligaments around the stenosis), or spinal fusion surgery (with or without the use of implants).

Rates of spinal surgery increased dramatically in the Medicare population between 1992 and 2003 (Tosteson, Skinner, et al., 2008), and nearly \$20 billion is spent annually on these procedures. Between 1980 and 2000, surgery for spinal stenosis was the fastest-growing type of lumbar surgery in the United States (Deyo, Mirza, et al., 2010). Rates of spinal stenosis surgery vary by more than a factor of five across geographic regions, raising concerns that many of these procedures may be inappropriate (Weinstein, Tosteson, et al., 2008).

Between 2002 and 2007, the overall rate of lumbar spinal stenosis surgery decreased slightly in the Medicare population, but the proportion of surgeries involving complex fusion procedures increased fifteenfold (Deyo, Mirza, et al., 2010). This increase follows an overall trend dating back to 1996, when the FDA first approved the use of intervertebral fusion cages (Deyo, Gray, et al., 2005). The rate of spinal fusion surgery increased 77 percent between 1996 and 2001, while the rates of other orthopedic surgical procedures increased modestly (e.g., the rate of knee and hip arthroplasty increased 13 to 14 percent) (Deyo, Nachemson, et al., 2004). However, as of 2001, the greatest growth in lumbar fusion surgery was for the treatment of herniated discs rather than spinal stenosis (Deyo, Gray, et al., 2005).

The risks of spinal stenosis surgery may be relatively low, but the procedure is disproportionately performed on the elderly, who have somewhat higher risks due to comorbidity. Analysis of Medicare data suggests that 3.1 percent of spinal stenosis surgery patients experience major medical complications, and 30-day mortality rates are approximately 0.4 percent. Both rates increase with age (Deyo, Mirza, et al., 2010). Complex fusion operations are associated with a 5.2-percent rate of major medical complications, compared with 2.1 percent for decompression-only procedures. Mortality rates are twice as high with complex surgery (0.6 percent versus 0.3 percent), and patients undergoing complex fusion operations remain hospitalized for two additional days, on average.

Over time, improvements in surgical and anesthetic techniques and supportive care have probably lowered the risks of surgery (Deyo, Mirza, et al., 2010) and may have fueled growth in the volume of these procedures. Improvements in diagnostic imaging technology such as axial-spine imaging may also have contributed to this growth (Deyo, Nachemson, et al., 2004), because surgeons often rely on imaging for diagnosing spinal stenosis and for determining the appropriateness of surgery (Haig and Tomkins, 2010). New devices, including spinal-fixation devices, computer-guided and minimally invasive surgery, bone-graft substitutes, and supplements such as bone morphogenetic proteins, may also contribute to the increasing use of surgery (Deyo, Nachemson, et al., 2004). No RCTs or prospective-cohort studies existed at the time intervertebral fusion cages were first approved for use (Deyo, Nachemson, et al., 2004).

Comparative-Effectiveness Question

At the time of SPORT, the principal clinical question was whether surgical treatment options were superior to nonsurgical treatment for patients with low back pain related to lumbar spinal disorders, including disc herniation, spinal stenosis, and degenerative spondylolisthesis. Prior to SPORT, the Maine Lumbar Spine Study, which enrolled 148 patients, was the largest study comparing the effectiveness of alternative treatments for spinal stenosis. Using a prospective-cohort design to compare surgical and nonsurgical approaches, this study found that patients who underwent surgery had better outcomes at three months, and these results held over the first year of follow-up (Atlas, Deyo, et al., 1996) before declining slightly over the next four years (Atlas, Keller, et al., 2000). While the surgical-intervention group appeared to have better outcomes in this study, nearly 25 percent of those undergoing surgery had no benefit, depending on the particular outcome (Atlas, Deyo, et al., 1996).

A 2005 Cochrane review summarizing the evidence prior to 2000 suggested that the relative efficacy of surgery was not established, because existing trials were small and enrolled patients both with and without degenerative spondylolisthesis (Gibson and Waddell, 2005). A large, randomized CER trial was considered necessary to provide stronger evidence on the benefits of surgical treatment for spinal stenosis among patients who did not have degenerative spondylolisthesis.

Study Design Characteristics

SPORT was launched in March 2000 to compare the outcomes of surgical and nonsurgical treatment for patients with lumbar intervertebral disk herniation, spinal stenosis, or degenerative spondylolisthesis (Weinstein, Lurie, et al., 2006; Weinstein, Lurie, et al., 2007; Weinstein, Tosteson, et al., 2008)

The spinal stenosis trial compared posterior decompressive laminectomy with usual care, which included physical therapy, education or counseling with home exercise instruction, and treatment with nonsteroidal anti-inflammatory drugs, if tolerated (Weinstein, Tosteson, et al., 2008). The specific type of treatment in the nonsurgical group was left to the discretion of each physician, because of the likelihood of patient heterogeneity in preferences and response to these treatments and limited data on the efficacy of individual nonsurgical treatments. Inclusion criteria required that patients have a history

of neurogenic claudication or radicular leg symptoms for at least 12 weeks. As such, the trial was considered to have enrolled a relatively severely affected population.

The trial included both a randomized cohort and an observational cohort (for patients who met eligibility criteria but who refused to be randomized). Outcomes were assessed in both cohorts over a two-year follow-up period.

Study Results

The results of the trial for patients with spinal stenosis are presented in Table 4.1. The intention-to-treat analysis showed that surgery was more effective than nonsurgical treatment on the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) bodily pain scale and on patients' self-reported ratings of symptom improvement but on few other primary or secondary outcomes. Patients with spinal stenosis had very high rates of crossover after randomized treatment assignment (as was the case for the two other subpopulations with disc herniation and degenerative spondylolisthesis). As shown in Table 4.1, only 67 percent of patients randomized to the surgical arm underwent surgery, while 43 percent of those randomized to nonsurgical treatment underwent surgery within two years of the baseline assessment. Patients who crossed over to receive surgery had high levels of self-rated disability, more psychological distress, worse symptoms, and, at baseline, a stronger preference for surgery.

Because such a large proportion of participants in the nonsurgical arm of the trial crossed over, the authors concluded that the relative superiority (or equivalence) of the treatments could not be determined from the intention-to-treat analysis. They therefore combined the data from the prospective-cohort study with data from the randomized cohort to create an observational cohort and analyzed the outcomes "as treated" (e.g., for patients who underwent surgery versus those who did not). The resulting observational-cohort study was not a randomized design, so the analysis adjusted for known baseline differences between patients in the two groups. This observational analysis found that surgery was superior to nonsurgical treatment across all primary and secondary outcomes. These effects held through two years of follow-up.

The clinical-trial protocol did not specify the type of procedure to be used for patients randomized to surgery, but in the vast majority of cases, decompression surgery rather than fusion surgery was the technique selected (89 percent versus 11 percent). Thus, this trial did not inform conclusions regarding the relative benefits of decompression surgery and fusion surgery (with or without instrumentation).

A cost-effectiveness analysis using patient-level data from SPORT found that surgery for spinal stenosis was moderately cost-effective at \$77,600 for each quality-adjusted life year (QALY) gained, while surgery for spondylolisthesis was not cost effective (\$115,600 per QALY gained).

Table 4.1**Results of the SPORT Trial**

| Outcome | Randomized Cohort^a (intention-to-treat analysis, N = 289) | Combined Randomized and Observational Cohorts^a (as-treated analysis, N = 654) |
|--|---|---|
| Primary outcomes | | |
| SF-36 bodily pain scale | Surgery preferred | Surgery preferred |
| SF-36 physical function scale | No difference ^b | Surgery preferred |
| Oswestry Disability Index (ODI) ^c | No difference | Surgery preferred |
| Secondary outcomes | | |
| Self-reported improvement | Surgery preferred | Surgery preferred |
| Satisfaction with current symptoms | No difference | Surgery preferred |
| Outcome | Randomized Cohort^a (intention-to-treat analysis, N = 289) | Combined Randomized and Observational Cohorts^a (as-treated analysis, N = 654) |
| Satisfaction with care | No difference | Surgery preferred |
| Stenosis Bothersomeness Index | No difference | Surgery preferred |
| Leg Pain Bothersomeness Index | No difference | Surgery preferred |
| Low Back Pain Bothersomeness Index | No difference | Surgery preferred |
| Other | | |
| Underwent surgery | 67% (surgical arm) ^d 43% (nonsurgical arm) | |
| Type of surgery | | |
| Decompression | 89% | |
| Noninstrumented fusion | 4% | |
| Instrumented fusion | 7% | |

| | |
|----------------------------------|-----|
| Intraoperative complication rate | 8% |
| Postoperative complication rate | 12% |

^aTwo-year outcomes are reported.

^b“No difference” implies no statistically significant difference between surgical and nonsurgical groups.

^cAmerican Academy of Orthopedic Surgery’s MODEMS version of the ODI was used.

^dIn the observational cohort, 96 percent of patients who initially chose surgery underwent surgery at two years; 22 percent who initially chose nonsurgical management had undergone surgery at two years.

Impact of the Trial on Clinical Practice

Most experts believe that SPORT had little or no impact on the use of surgery for the three spinal disorders studied. Notably, the publication of the SPORT results included both the intention-to-treat results and the post-hoc observational-cohort study results, potentially adding an element of confusion to their interpretation. The majority of our discussants felt that the intention-to-treat analysis was “uninformative” because of the very high level of crossover, while the observational-cohort analysis was considered to be of high quality and to have produced useful information. Interpretations of the observational-cohort study results generally divided along specialty lines: spinal surgeons embraced them, while nonsurgical spine physicians were more skeptical. Because the observational-cohort results were broadly consistent with those of two earlier studies—the Maine Lumbar Spinal Study and the Finnish trial by Malmivaara (Malmivaara, Slatis, et al., 2007)—the SPORT findings appear to have reinforced the conventional wisdom that surgical intervention is the preferred strategy for treating spinal stenosis.

Discussions with several orthopedic surgeons indicated that the interpretation of the results was more nuanced than we expected. Several discussants said that the patients participating in the observational study who selected nonsurgical treatment did “fairly well” (despite the fact that those results were not emphasized in the dissemination phase) and that surgical patients did not have sizable improvements in health outcomes. They pointed out that the goal of the trial was not to determine whether surgery “worked,” since prior studies had already provided some evidence, but rather to provide estimates of the magnitude of benefits and harms, which clinicians and patients could then use in making decisions about the value of surgical treatment. This perspective reflects the unique weights that individual physicians and patients may assign to various treatment effects when interpreting CER evidence.

Key Barriers to Clinical-Practice Change

Issues with the Interpretation and Dissemination of the SPORT Results

SPORT is an unusual example of CER in that two distinct interpretations based on the intention-to-treat analysis and the observational-cohort study could have different effects on clinical practice, depending on which is most prominent in the practice community and among patients. The SPORT results could be expected to either confirm current practice (yielding no change in the use of surgery), suggest that surgery is used too frequently given its limited benefit (yielding a reduction in the use of surgery), or suggest that surgery is not used frequently enough given its potential for benefit (yielding an increase in the use of surgery). The wide acceptance among key stakeholders of the observational-cohort study results—which tended to favor surgical intervention—suggests that observational CER study designs can provide useful guidance to practice and policy if a randomized clinical trial is not a feasible alternative (or is flawed because of crossover, as in this case).

Financial Incentives

Among orthopedic surgical procedures, spinal surgery is relatively highly reimbursed. The financial incentives of physicians, hospitals, and device manufacturers are aligned in a way that may promote the use of more-complex surgery. SPORT did not specifically address the relative benefits of decompression surgery and complex spinal surgery (i.e., fusion). In particular, differences in professional fees for the two types of surgery are quite large. Most of the experts we spoke with agreed that the trial results probably have not contributed to the growth in complex surgery for spinal stenosis—a trend that began nearly a decade ago. Others believe that some surgeons have used the SPORT results as justification for recommending more complex procedures, because it included patients who underwent fusion surgery.

Two other types of financial incentives may be contributing to the growth of surgery, as well as complex surgery. First, some medical device distributors pay surgeons “dividends” based on the number of devices they use. The Office of the Inspector General of HHS and the Centers for Medicare and Medicaid Services (CMS) have both warned that these arrangements may violate federal anti-kickback statutes and laws governing patient referrals (Carreyrou and McGinty, 2011). The prevalence of these arrangements is unclear but likely to be small. Second, patients who are covered by workers’ compensation policies may also be more likely to undergo complex spinal surgery, as insurers in this area are less likely to impose restrictions on coverage than those in the commercial health insurance market. Experts noted that repeat surgery rates are higher in patients covered by workers’ compensation policies.

The Importance of Referring Physicians in the Treatment Decision

Nearly all experts expressed the view that consultation with a spine surgeon greatly increases the likelihood that a surgical procedure will be recommended. This suggests that dissemination of the CER results to referring physicians may be as important as dissemination to spine surgeons. The appropriateness and timing of the referral decision could be an important area of focus for decision support tools that use evidence from SPORT.

Challenges of Conducting CER Using Randomized Surgical Trials

Evidence from past trials suggests that RCTs in surgery are fraught with methodological challenges. First, the “blinding” of treatment assignment in surgical trials represents an ethical challenge, even though studies have shown that certain spinal procedures have little benefit. According to our experts, surgical interventions are associated with strong placebo effects, so randomization is critical to balance treatment groups with regard to surgical preferences and expectations of outcomes. Randomization is even more critical when outcomes are subjective, such as the assessment of pain, which is a common outcome for spinal surgery trials. Controlled trials involving sham surgery may have a role in some contexts where ethical concerns are mitigated by the low risks of minor incisions, such as in knee arthroplasty; however, sham spinal surgery would be difficult, if not impossible, to implement without ethical concerns.

The other main challenge in CER trials involving surgical interventions is crossover, which threatens a trial’s internal validity. The high rate of crossover in both directions (from the nonsurgical arm to the surgical arm and vice versa) significantly weakened SPORT’s intention-to-treat analysis. Some experts indicate that crossover is not inevitable but, rather, can be carefully controlled by trial investigators. They cite trials conducted in European countries that tend to have much lower rates of crossover (Deyo, Natchemson, et al., 2004) as evidence that these problems can be overcome. SPORT was conducted at a small number of academic institutions specializing in spine surgery, where, according to one expert, surgeons and referring physicians may have lacked equipoise in their willingness to recommend surgery. Surgeons in other settings might be better able to control crossover in these types of trials. Our discussants suggested that non-adherence in SPORT may also not have been detected sufficiently early.

Trial Results Provided Insufficient Detail to Enable Clinicians to Tailor Treatment

The primary comparison of two or more treatments among cohorts analyzed by treatment may not provide detailed data on subgroups of patients that physicians could identify as particularly likely to benefit from a particular treatment. Thus, physicians may ignore the group-level data. This might be especially important for SPORT, where patients may have experienced a wide range of anatomical abnormalities and differences in pain symptoms.

Reluctance of Clinicians to Use All Available Evidence

Most of the strongest evidence on the limited efficacy of spinal surgery comes from randomized trials conducted in Europe. European trials have tended to use specific nonsurgical treatments (unlike SPORT) and in this context have shown that patients can improve dramatically without surgery. According to one expert, American spinal surgeons are either unaware of the results of European studies or believe that the results are not generalizable to the U.S. context. This suggests that available evidence about the benefits and harms of surgery for lumbar spinal stenosis is not being used and therefore does not influence decisionmaking. However, there may be key differences that do limit the generalizability of non-U.S. trials. For example, some pain-management options, such as cognitive behavioral therapy, may not be widely available in the United States, may not be reimbursed, or both. Also, in the United States, the financial incentives associated with fee-for-service reimbursement may lead to patients proceeding

more quickly to surgery, so the profile of patients enrolled in European trials may be clinically different from that of those evaluated for surgery in the United States.

Conflicting Guidelines from Multiple Specialty Societies

Physicians who treat patients with spinal conditions may draw on any of several practice guidelines, and these guidelines are often in conflict with one another, possibly because of limitations of the evidence regarding the procedures. One discussant mentioned that the American College of Physicians and the American Pain Society have produced more-conservative practice guidelines, while the North American Spine Society (NASS) has been far more likely to recommend surgical procedures. With strong backing from the device industry, the International Society for the Advancement of Spine Surgery was recently formed, because many thought that the leadership of NASS and its journal were not favorable enough to the industry.

The Absence of Registries

RCTs may exclude high-risk patients and are often of inadequate size or duration to identify harms, particularly rare events (Chou and Helfand, 2005). At the same time, administrative data, such as claims, do not contain information on patients' symptom severity, extent of anatomic abnormalities, functional impairments, or specific implants used during surgery (Deyo, Mirza, et al., 2010). The dearth of available data with which to estimate the benefits and risks of alternative treatment strategies for patients with lumbar spinal stenosis suggests a potential role for registries. Given the findings from earlier research that more-invasive surgical procedures are associated with greater complication rates (Deyo, Mirza, et al., 2010), experts have suggested that registries might be most helpful in providing additional data about the risks of surgery. They may fill gaps in evidence and also serve as the basis for development of performance measures and appropriateness criteria for procedures. Currently, there are few registries for spinal surgery. The American Academy of Orthopedic Surgeons attempted to develop a registry nearly a decade ago, but the effort was severely underfunded and the registry never materialized. NASS is seeking to develop a registry, and this effort is being viewed as a key vehicle for future CER relating to spinal surgery. This research could help support the development of appropriateness criteria, which, according to one surgeon, is "inevitable."

Stakeholders identified at least three factors that may have contributed to the lack of registries for spinal surgery. First, surgeons have few incentives to develop a registry and may face the risk of bringing to light performance problems and other safety risks. As one expert suggested, "Not everyone can practice at the level of the SPORT surgeons." He noted that the majority of spine surgeries are performed in low-volume practices, where outcomes may be worse than in high-volume practices, and that professional societies have historically challenged studies that assess volume-outcome relationships in orthopedic surgery for this very reason. Second, patient-reported outcomes, the primary measure of benefit in spinal surgery, display remarkable variability and may limit the power-of-outcomes analyses. For this reason, some experts believe that registries may be most useful for quantifying the harms associated with spinal surgery. Third, the lack of standardized definitions for both diagnosing conditions and measuring outcomes poses a significant barrier. Diagnosing orthopedic conditions (including spinal stenosis) is challenging, because patient-reported symptoms and imaging results are not always

correlated, so diagnostic criteria have not been developed. According to one expert, the FDA is the entity best positioned to enforce standardized definitions for measuring outcomes of surgery and adverse events, but it has not done so to date. Claims data, while difficult to use for measuring benefits, might be well suited for measuring harms, according to this expert, but these data have not been used effectively. Registries in other countries have identified device problems much earlier than those in the United States. One expert suggested that neither hospitals nor physicians have any interest in drawing attention to the risks of surgery; in contrast, patients have an interest, but they have few advocates.

Lack of Performance Measures and Appropriateness Criteria

Development of quality measures for orthopedic surgery, including measures of readmissions and mortality for select procedures, is advancing. However, there are few measures for spinal surgery specifically, and no appropriateness criteria currently exist. According to one expert, “It is in everyone’s interest to *not* have clear definitions of appropriateness.”

Medical Innovation and Marketing

New surgical devices and their marketing increased significantly over the past decade and may be key drivers of the use of complex procedures in spinal surgery. According to experts, the rapidity of technological development and its marketing pose a barrier to CER, which often takes several years to complete. Even in the absence of evidence about the benefit of new technologies, arguments in favor of new devices are often compelling and may speed their adoption. This may be an important factor in regional variation in the use of complex surgery, as some surgeons have a desire to be local innovators (Deyo, Mirza, et al., 2010). However, experts noted that hardware was not involved in a large share of spinal stenosis surgery procedures.

The device industry played an active role in the dissemination of findings from SPORT, although it initially attempted to discredit the study even before the results were released. After the results were found to be favorable to surgery, the industry began a campaign to promote awareness of the benefits of surgery. The messaging was vague and neglected to focus on the procedures used in the trial, only 7 percent of which involved spinal hardware.

Few Payers Have Restricted Insurance Coverage for Lumbar Surgical Procedures

By nearly all accounts, payers and purchasers have few policies in place that limit access to surgery for patients with lumbar spinal stenosis. This appears to be the case for both Medicare and private health plans. Recently, BCBS of North Carolina discontinued coverage of fusion procedures for spinal stenosis following a comprehensive review of the evidence. A broad coalition of professional societies representing spine surgeons submitted comments and were successful in modifying parts of the coverage policy. According to one expert, the action taken by BCBS represents a “shot across the bow,” and similar scrutiny may be placed on spinal fusion surgery in the future.

Key Enablers of Clinical-Practice Change

Growth in the Use of Shared-Decisionmaking Tools

According to experts, many patients may not understand the benefits and risks of surgery for lumbar spinal stenosis before undergoing the procedure. In particular, it is unclear whether patients are fully aware of the possibility that surgery may not improve their symptoms (which occurs in approximately 25 percent of patients) or that they may need a repeat procedure in the future. In many cases, physicians may believe that they have effectively communicated this information, but discussants mentioned to us that some physicians are just not good at enabling informed consent. The majority of experts we spoke with felt that use of these procedures was driven more by physician preferences than by patient preferences. Many believe that spine doctors tend to advocate for surgical treatment of spinal stenosis and that more-informed patient decisionmaking could help to better address patients' preferences (Deyo, Mirza, et al., 2010).

While the use of decision aids appears to be a growing trend, many experts believe that they are not used frequently enough in spinal surgery. Some suggested that evidence from SPORT has been directly incorporated into patient decision aids for spinal stenosis surgery, including those developed by the Foundation for Informed Medical Decision Making, but their existence does not guarantee that decision aids will be used. One expert mentioned that decision-quality measures might help promote the use of decision aids and that such measures may soon be ready for use. Others argue that the effectiveness of decision aids is unclear. According to one discussant, surgeons thought the decision aid developed by Dartmouth for prostate surgery was too "anti-surgery," while internists generally thought it was too "pro-surgery."

Emergence of Radiology-Benefits Managers

Orthopedic surgeons indicated that diagnostic imaging is excessive, and experts stated that RBMs are being used more commonly by health plans to increase the appropriateness of diagnostic imaging procedures. RBMs may use administrative controls such as prior authorization or utilization review to limit the use of such procedures. Because spine magnetic resonance imaging (MRI) is commonly used to diagnose spinal stenosis and is potentially a key factor in the decision to refer patients for further evaluation, RBMs may ultimately play a significant role in influencing which patients undergo surgery. However, most of the experts with whom we spoke had limited knowledge about the impact of RBMs to date.

Conclusions

SPORT appears to have had little impact on clinical practice, and the seeds of its low impact were sown mostly in the generation phase. The study design, which was unblinded, allowed for a very large patient crossover. As a result, what was intended to be an RCT with an intention-to-treat analysis had to also be analyzed as an observational-cohort study using an as-treated analysis. Analogous studies have avoided these difficulties, suggesting that they are not inherent to this type of CER but can be forestalled by careful study design and execution.

In the interpretation phase, the RCT results suggesting limited benefits from surgery were discounted because of high rates of patient crossover. In contrast, the observational-cohort study, at least within the spinal surgery community, served to confirm the relative advantage of surgery, which was already the prevailing method of treatment. Interpretation was further complicated by the study's lack of detail on subgroups, which made it hard to judge who would most benefit from surgery, as well as the (possibly erroneous) perception that the surgical techniques used in the study were already outdated. Presenting competing analyses may have opened the results to conflicting interpretation, but the observational results alone produced different interpretations regarding the magnitude of the benefit provided by surgery.

In its formalization phase, SPORT again highlighted the challenges in weighing the relative strengths and weaknesses of RCTs and observational-cohort studies and the selective use of evidence. Multiple specialty societies, possibly influenced by various levels of industry sponsorship, issued competing and conflicting guidelines, while relevant data from European studies was generally discounted or ignored. Registries might help to bolster guidelines or to generate appropriateness criteria, but since many outcomes from spine surgery are subjective, registries may be best suited to report on harms. Orthopedic surgery has few registries, and financial incentives are not aligned to promote participation by surgeons.

While dissemination of the SPORT results appeared to be far-reaching, messaging about them emphasized the benefits of surgery rather than the significant clinical improvement among patients in the nonsurgical group and the relatively small difference in benefit between groups. Referring providers appear to be the optimal point for dissemination of the results, since referral to a surgeon is usually followed by surgery. Intense marketing of spinal hardware by the device industry may override the results of clinical trials, and, as SPORT illustrates, messages may be vague and selective, omitting key evidence provided by the trials. Similarly, payers and purchasers, faced with both the "positive" results from the observational-cohort analysis and the "equivalence" results from the intention-to-treat analysis, appear to have initially embraced the observational-cohort analyses and did not enact policies regarding decompression surgery. However, there are now some early examples of more-nuanced and data-driven reimbursement policies focusing on related procedures (e.g., fusion surgery).

Nevertheless, in the implementation phase, strong financial incentives continue to favor surgical over nonsurgical treatment. The alignment of financial incentives among physicians, hospitals, and device manufacturers appears to have increased the use of complex procedures despite uncertainty about their effectiveness and considerable evidence of greater risks. Countering this is the increasing use of RBMs, which may reduce inappropriate upstream diagnostic procedures and may play a potential role in the use of patient decision aids. While the SPORT results can be viewed as both flawed and confirmatory of current practice, the trial was successful in providing quality data on the relative risks and benefits of surgery, and these data have been integrated into patient decision aids. Those tools might ultimately change clinical practice by incorporating fully informed patient preferences into decisions about surgery. Currently, few incentives encourage the use of such shared decisionmaking or more rigorous informed-consent processes. The use of these techniques early in the pathway leading to surgery will be critical to their overall effectiveness. Incentives to promote the spread of patient decision aids and efforts to

improve the appropriate use of diagnostic imaging represent the most important current strategies for changing clinical practice in the future.

Key findings from the SPORT case study are presented in Table 4.2.

Table 4.2

Key Findings from the SPORT Case Study

| Phase | Key Findings |
|----------------|---|
| Generation | <ul style="list-style-type: none"> • Blinding participants regarding their treatment assignment was impractical, but in other surgical studies, it has revealed large placebo effects and a more limited benefit from surgery. • The study design enabled very large (over 40 percent) crossover from both treatment arms, whereas analogous studies have limited noncompliance. |
| Interpretation | <ul style="list-style-type: none"> • The intention-to-treat results was ignored and the as-treated results from the observational-cohort study were considered the primary results. • Publication of both the intention-to-treat and as-treated results allowed different stakeholders to glean confirmation of their competing viewpoints. • The study provided insufficient detail on subgroups to judge which patients were most likely to benefit from surgery. • The perception that study treatments were already outdated affected interpretation. |
| Formalization | <ul style="list-style-type: none"> • Multiple specialty societies generated competing and conflicting guidelines, possibly influenced by their level of industry sponsorship. • Strong evidence from European trials (which concluded that structured nonsurgical treatment works well) appears to have been generally ignored. • Registries might help refine appropriateness criteria and quantify risks and benefits, but they are best suited to measure harms, so physicians have limited incentive to support their development. |
| Dissemination | <ul style="list-style-type: none"> • Decisionmaking by referring physicians is a key factor, so dissemination of CER results should focus on them. • Performance measures and appropriateness guidelines for spinal surgery could work, but they are not yet widely employed. • The device industry attempted to preemptively discredit the study, then reversed course to promote it when the results proved favorable—this can be anticipated. |

| Phase | Key Findings |
|----------------|--|
| Implementation | <ul style="list-style-type: none"><li data-bbox="477 275 1357 300">• Reimbursement significantly favors surgery, especially more-complex procedures.<li data-bbox="477 331 1365 394">• Workers' compensation reimbursement for both surgeons and patients also favors more-complex procedures.<li data-bbox="477 426 1289 451">• Referral to a surgeon can be interpreted as endorsement of the procedure.<li data-bbox="477 483 1438 577">• Payers have begun to impose modest limits specifying the indications for which they will compensate spinal procedures; while controversial, they may limit inappropriate surgeries.<li data-bbox="477 609 1398 672">• Payers are also employing RBMs to rein in inappropriate upstream diagnostic imaging procedures.<li data-bbox="477 703 1419 762">• Using decision support tools covering treatments for back pain may improve outcomes, but physicians are skeptical and have little incentive to use them. |

COMPANION Case-Study Report

Clinical-Practice Context

The five-year mortality rate following a diagnosis of heart failure (HF) is 50 percent, even with optimal medical therapy. HF is also associated with a poor functional status (Friedewald, Boehmer, et al., 2007). For patients with both HF and delayed electrical conduction in the ventricles (wide QRS), treatment with cardiac resynchronization therapy (CRT), using an implantable pacemaker to stimulate both ventricles, has been shown to be effective in improving functional status (Abraham, Fisher, et al., 2002). However, the efficacy of CRT had not been established for those HF patients who also require an implantable cardiac defibrillator (ICD) to treat intermittent life-threatening abnormal heart rhythms.

CRT-device implantation has been described as “a complex and rapidly developing clinical science” that requires a broad range of new skills (Burkhardt and Wilkoff, 2007). The implantation procedure and the optimization of device functioning are both technically challenging, and there are no standards for device optimization. CRT devices may be implanted by a range of physicians, including cardiologists, thoracic surgeons, and electrophysiologists. Race and sex disparities in ICD use have been well documented and may influence the uptake of CRT-D (combined CRT and ICD therapy).

Many risks are associated with CRT procedures. Acute dislodgements during the initial procedure are associated with increased odds of other adverse events, including cardiac arrest, cardiac tamponade, device infection, pneumothorax, and in-hospital death (Cheng, Wang, et al., 2010). Anecdotal evidence suggests that infection rates relating to the devices are rising (Poole, Gleva, et al., 2010). The initial implantation of a CRT-D device can have a 10-percent or greater in-hospital complication rate than implanting an ICD alone (Reynolds, Cohen, et al., 2006). Among patients who undergo generator replacement and lead addition, the six-month major complication rate among those who upgraded from ICD to CRT-D or underwent revision of a CRT device is nearly 19 percent (Poole, Gleva, et al., 2010). Complication rates and procedure times appear to decrease somewhat with the experience of the physician (Leon, Abraham, et al., 2005).

Failure rates for CRT therapy can be up to 40 percent, even for patients who have a wide QRS interval (Burkhardt and Wilkoff, 2007; Friedewald, Boehmer, et al., 2007), and the device may last only ten years before needing replacement. The implant failure rate is approximately 5 percent per year, and failure is more common at low-volume facilities, particularly community hospitals (Friedewald, Boehmer, et al.,

2007). According to anecdotal reports, a physician's decision to refer a patient for CRT may depend on the outcomes of the first several patients he or she refers (Friedewald, Boehmer, et al., 2007).

The costs of implantation and follow-up of CRT-D therapy over a roughly seven-year period are more than \$13,000 higher than those of CRT therapy alone (Budde, 2006). The nonsurgical medical care of HF is time-intensive and underreimbursed (Fonarow, Yancy, et al., 2008), and the decision to proceed with device therapy often requires multiple, in-depth discussions with patients. Virtual (electronic) follow-up of outcomes with these devices is also currently unreimbursed (Friedewald, Boehmer, et al., 2007).

Comparative-Effectiveness Question

Heart failure occurs when the heart can no longer pump blood adequately. CRT appears to improve this condition by electrically stimulating the heart to achieve better coordinated blood-pumping contractions. Most HF patients also appear to be at high risk for potentially fatal derangements in the heart's electrical activity. Should this occur, ICD devices may be used to shock the heart back into a normal electrical rhythm, potentially averting death, but ICDs do nothing the rest of the time to reduce HF symptoms. The principal question addressed by COMPANION was whether adding CRT with or without ICD treatment to the medical management of HF patients with a wide QRS not only improved functional measures but also reduced hospitalization rates and all-cause mortality (Bristow et al., 2004). Previous studies, including the MIRACLE trial (2002), demonstrated that CRT alone improved the New York Heart Association (NYHA) functional class, exercise duration, and quality of life of HF patients who did not require an ICD or traditional pacemaker, but they did not have sufficient power to discriminate a mortality benefit (Abraham, Fisher, et al., 2002). In patients who require ICD implantation mainly for secondary prevention of sudden cardiac death and who also have HF, an ICD has been shown to significantly reduce mortality but does not appear to improve HF symptoms. Two trials—CONTAK-CD (Higgins, Hummel, et al., 2003) and MIRACLE-ICD (Young, Abraham, et al., 2003)—examined whether adding CRT to the ICD in these patients improved outcomes. Both had fairly short follow-up periods and insufficient power to discriminate differences in mortality or hospitalization, but they did demonstrate improved patient NYHA functional status.

Study Design Characteristics

COMPANION was an RCT with three treatment arms, designed to evaluate the efficacy of CRT and CRT-D in patients with HF (NYHA class III or IV), cardiomyopathy (with a left ventricular ejection fraction < 35 percent), delayed ventricular conduction (QRS >120 msec), and no specific indication for an ICD or pacemaker. A total of 1,520 patients were randomized. Patients were assigned in a 1:2:2 ratio to receive optimal pharmacological therapy (OPT), OPT + CRT, or OPT + CRT-D. Follow-up duration was until the primary end point (which averaged about 16 months for those who reached an end point during the two years of the study). The primary-efficacy end point was a composite of either death from any cause or hospitalization from any cause.

Study Results

The results of COMPANION were published in 2004 (Table 5.1) (Bristow et al., 2004). The CRT treatment group showed a statistically significant improvement of 17 percent in the combined end point over those who received OPT alone. The CRT-D treatment group showed a similar 17-percent improvement. While adding an ICD to CRT did not significantly improve the combined end point, it did reduce 12-month all-cause mortality (a secondary outcome) from 15 percent to 12 percent. It is worth noting that the OPT-only group had a 13-percent dropout rate (before an end point was reached), compared with only a 2-percent dropout in the two treatment groups. However, the treatment groups experienced initial procedure failure rates during implantation of the CRT and CRT-D devices of 13 percent and 9 percent, respectively. The results of this study imply a clear survival and quality-of-life benefit from adding CRT (with or without CRT-D) to OPT for patients who have HF with delayed ventricular conduction. The practice at the time was to use CRT for HF but withhold ICD devices in these patients given both safety concerns and a lack of proven benefit.

Table 5.1

Results of the COMPANION Trial

| Outcome | Results | | |
|---|-----------|-----------|-------------|
| | OPT Alone | CRT + OPT | CRT-D + OPT |
| Primary outcome | | | |
| Death or hospitalization (any cause), 12-month rate | 68% | 56%* | 56%* |
| Secondary outcomes | | | |
| Death or hospitalization (cardiac cause), 12-month rate | 60% | 45%* | 44%* |
| Death or hospitalization (HF), 12-month rate | 45% | 31%* | 29%* |
| Mortality (any cause), 12-month rate | 19% | 15% | 12%* |
| Change in six-minute walk distance at three and six months (meters) | 9±84 | 33±99* | 44±109* |
| | 1±93 | 40±96* | 46±98* |
| Improvement in quality of life at three and six months (Minnesota HF Scale) | 9±12 | 24±27* | 24±28* |
| | 12±23 | 25±26* | 26±28* |
| Improvement in NYHA functional class | 24% | 54%* | 55%* |
| | 38% | 61%* | 57%* |

| Other | | | |
|---|-----|-----|----|
| Moderate or severe event associated with device implantation | | 10% | 8% |
| Failure of initial device-implantation attempt | | 13% | 9% |
| Sensitivity analysis: study withdrawal (prior to primary end point) | 13% | 2% | 2% |
| Sensitivity analysis: mortality status unknown at end of study | 4% | 1% | 1% |

NOTE: Asterisks denote statistically significant differences compared with OPT alone.

Impact of the Trial on Clinical Practice

Medicare began covering CRT-CD in 2005. HF registries (IMPROVE HF, Medicare NCDR-ICD) subsequently began to track long-term outcomes. Data from the IMPROVE-HF registry indicate that wide variation remains across different practices in the proportion of CRT devices implanted that also incorporate ICD functions (CRT-D) (Mehra, Yancy, et al., 2009). While HF is the fastest-growing cardiac diagnosis, patients are, on average, ten years older and otherwise sicker than those in COMPANION or similar studies, and the benefit of CRT-D therapy for these older patients is unclear. The registries and experts also suggest that there is both significant overuse and underuse of CRT-D therapy between regions—not all HF patients who are candidates for CRT-D therapy are offered it, and other HF patients may be receiving CRT-D therapy inappropriately. Some of the reasons are clear: About 80 percent of HF patients are treated by noncardiologists, who tend not to refer eligible patients to an interventional cardiologist for CRT therapy (Friedewald, Boehmer, et al., 2007). Interventional cardiologists, by contrast, may overutilize CRT and ICD: Many patients who received CRT were either not on OPT first or did not meet other current eligibility guidelines. In addition, more procedures are now being performed at low-volume centers by less-experienced practitioners, and this may be associated with higher complication rates.

New Evidence Following Initial Release of COMPANION Results

Critiques of the COMPANION study that emerged after the release of the results generally did not question its methodology or the validity of its results but focused instead on its generalizability. Numerous editorials proposed eligibility criteria for CRT and ICD. Some authors contended that CRT appropriateness criteria should be extended to include patients with less-severe HF. Others argued that most of the benefit appeared to occur in the most-severe HF patients, and CRT-D use should be restricted to them. Still others suggested that CRT might also benefit patients with narrow QRS HF—or, in other words, patients without obvious conduction delays. Some follow-up analyses attempted to derive predictive factors for the treatment response to CRT. For example, patients with HF from

ischemic disease appeared to do less well than those with nonischemic HF. It was also suggested that higher B-type natriuretic peptide (BNP) levels at baseline might predict better outcomes or that functional cardiac MRI to diagnose midwall fibrosis was predictive. Finally, there were questions about cost-effectiveness—estimates ranged from \$7,000 to almost \$100,000 per QALY saved.

As a group, the three CRT-D trials demonstrated that CRT is appropriate for ICD candidates with severe HF. There was, however, a vigorous debate over the use of the ICD component for primary (not only secondary) prevention of sudden cardiac death in HF patients given both the significantly higher cost of CRT-D over CRT and the marginal gain in outcomes. Estimates of the cost of adding ICD to CRT were as high as \$171,538 per additional QALY (Feldman et al., 2005). Thus there were attempts to identify factors that predicted a treatment response from adding ICD to CRT. Some analyses suggested that a wider QRS or worse functional status on the NYHA scale predicted more benefit, while later studies found that these “predictive” factors were only weakly associated and that a broad range of patients benefited from CRT-D rather than CRT alone.

Key Barriers to Clinical-Practice Change

Uncertainty About the Generalizability of Results

Discussants reported that the COMPANION results generated uncertainty about which patients would receive the most benefit from CRT. COMPANION had enrolled predominantly NYHA class III patients, and experts were divided on whether the results might also apply to non-class III patients. Conflicting messages on this point may have blunted efforts to educate primary care providers about CRT. To date, no primary care professional society has produced guidelines regarding referral for CRT, and according to our discussants, primary-care providers have little knowledge of the cardiology-specialty guidelines. Some stakeholders believe that primary-care physicians may also hesitate to refer patients, because of incorrect views regarding the costs, risks, and side effects of CRT-D therapy. According to one estimate, only 45 percent of patients that meet guideline criteria are ever referred.

Use of CER Results to Treat Noncomparable Patients

It is not clear that the patients actually receiving CRT and CRT-D therapies are comparable to those treated in the CER studies. In fact, registry data indicate that HF patients in the community differ from the CER study population in at least two important respects: Registry patients are, on average, a decade older than those in trials, and they have many more comorbid conditions. For example, one in five patients in the American Heart Association’s Get With the Guidelines–Heart Failure registry (GWTG) who received CRT had a history of COPD, which has been associated with a 50-percent increased risk of death in HF patients (Piccini, Hernandez, et al., 2008). Patients with right-bundle branch block and atrial fibrillation also commonly receive CRT-D despite not being comparable to the study groups. Based on the limited available evidence, the outcomes of CRT in these noncomparable patients appear to differ from the COMPANION results. Despite this, the COMPANION results are frequently used as a rationale for treating this group. Guidelines leave these areas open, and as one discussant put it, “The fuzzier the guidelines are, the happier the medical device industry is.”

Financial Incentives Are Poorly Aligned with CER Evidence

There are no effective limitations on reimbursement by payers for CRT-D procedures. Stakeholders report that these procedures are more or less revenue-neutral for hospitals (especially CRT-D, with its longer operating-room time), so hospitals have no strong incentive to either encourage or discourage their use. The main financial driver is physician reimbursement. Without appropriateness criteria to guide reimbursement, some CRT-D devices have been implanted in patients without clear indications for the device. One critical driver of CRT-D use is Medicare reimbursement policy. CMS pays for ICD implantations but not CRT. (CMS approved ICDs for primary prevention of sudden cardiac death in 2003 but has yet to issue a national coverage determination for CRT.) Thus, despite the low marginal cost-effectiveness of CRT-D over CRT alone, virtually all implantations in Medicare-covered HF patients involve CRT-D devices. Many ICD implantations appear to be inappropriate: Recent studies reveal that use of CRT-D therapy is off-label in up to 30 percent of all implantations (Piccini, Hernandez, et al., 2008; Farmer, Kirkpatrick, et al., 2009; Fein, Wang, et al., 2010; Al-Khatib, Hellkamp, et al., 2011), while 70 percent of patients referred for CRT were not on OPT, as recommended by guidelines (Friedewald, Boehmer, et al., 2007). When medical therapy is maximized, up to one-third of patients may improve to a point at which they are no longer candidates for CRT, suggesting that another 20 percent or more of implantations may also be premature.

Lack of Patient Decision Aids

While CER-based guidelines regarding CRT-D are available for physicians (as discussed below), there are currently no decision aids for patients. Given the challenges of projecting and interpreting the risk of sudden death and the technical details of this largely preventive treatment, patients who do not have well-designed educational materials are unlikely to have much voice in the treatment decision. The experience with creating decision aids for coronary angioplasty suggests that effective tools could certainly also be produced for patients considering CRT therapy. The main implementation barrier is lack of incentives for physicians to offer decision aids to patients. Neither specialists who perform the procedure nor referring generalists have a triggering motivation to acquire decision aids or to encourage patients to use them if they are available. Limited research in other clinical areas suggests that decision aids may reduce the use of more-invasive treatment options. The most appropriate single point for employing decision aids in the pathway leading to CRT implantation is likely to be referring primary care physicians or cardiologists (rather than interventional cardiologists), because these physicians may have greater equipoise in deciding whether or not to recommend the procedure.

Limited CER Detailing Efforts

Stakeholders report that detailing of the results from COMPANION and similar studies was tailored to interventional cardiologists. Device manufacturers with a limited budget and sales force focused their resources on physicians most likely to actually perform CRT implantation procedures, to the exclusion of referring physicians, including general cardiologists and primary-care physicians. The device manufacturers may have relied on indirect marketing of the devices to these other physicians by interventional cardiologists. Academic detailing to specialists or generalists was limited and mostly resulted from informal “water-cooler” conversations with colleagues. The result appears to have been

limited awareness of the CER results among referring physicians and a dependence on referral to interventionists to evaluate whether patients were appropriate candidates for therapy.

The Importance of Referring Physicians in the Treatment Decision

Most stakeholders reported that HF patients, once referred to an electrophysiologist, were very likely to receive a device. However, many primary-care physicians and some general cardi-ologists are reportedly not familiar with the COMPANION CER results. This may result in failure to refer patients with appropriate indications. At the same time, referring physicians may use referral to an electrophysiologist as an opportunity for assessment when they are unsure whether the patient is an appropriate CRT candidate. As noted, many of these patients are not yet on optimal medications. Nevertheless, the assumption among many interventional cardiologists is that the referral was actually a decision that the patient needs the procedure and the referring physician will be disappointed if the patient does not receive one. As a consequence, many patients who are appropriate candidates for CRT therapy are apparently never referred (especially by primary-care providers), while others who may not be appropriate candidates undergo CRT implantations simply because they were referred for evaluation.

Lack of Clinical Decision Support Tools

Stakeholders report that clinical decision support tools would probably improve appropriate referrals of HF patients for CRT therapy. They note that primary-care physicians, in particular, would probably identify more patients for referral if they had such tools. They emphasize that decision support must be integrated into the flow of care, and that a “stand-alone” CRT tool is not likely to be used. As one discussant put it, “You’ve got to make it easy for physicians to do the right thing.” Integrating decision support into quality improvement programs that include feedback on a physician’s performance might provide a further incentive. For example, one discussant reported that two cardiology practices that implemented a “hard stop” and required physicians to identify whether each HF patient met criteria for use of CRT achieved 100-percent compliance, meaning that all of the patients referred for CRT met the guidelines.

Few Restrictions on Which Interventionists Can Perform Implantation of CRT

There is currently no binding restriction on which physicians or centers can offer CRT. Stakeholders report that the only real restriction is whether a hospital will credential a given physician to perform the procedure, and hospital credentialing criteria reportedly vary widely. The COMPANION study “hand picked” only high-volume electrophysiology centers to participate in the study group. Most procedures are not currently performed at such centers. This is significant because even for patients who meet the CER appropriateness criteria, implantations by non-electrophysiologists were associated with a higher risk of procedural complications and lower likelihood of receiving a CRT-D device (when indicated) than those for patients whose ICD was implanted by an electrophysiologist (Curtis, Luebbert, et al., 2009). Less-experienced operators and centers with low procedure volumes were also associated with more complications and poorer outcomes. Voluntary certification procedures (discussed below) may help alter this landscape.

Inattention to Cost-Effectiveness Results

Cost-effectiveness studies have raised significant questions regarding the relative cost of CRT therapy for some patients and even more questions about the marginal cost-benefit of adding ICD treatment to CRT for primary prevention of sudden cardiac death. Despite this, stakeholders reported that the decision to recommend CRT virtually never includes consideration of cost-effectiveness. Stakeholders report that until recently, doctors paid little attention to cost-effectiveness, and it still does not appear to significantly affect their clinical decisions regarding CRT.

Key Enablers of Clinical-Practice Change

Rapid Integration of CER Evidence into Guidelines

In 2005, following publication of COMPANION and related studies, the ACC/AHA guidelines were modified to include recommendations regarding CRT therapy. Dissemination of the guidelines was reportedly accelerated by publishing them on dedicated patient and physician webpages. The guidelines recommended use of CRT for patients with a left ventricular ejection fraction (LVEF) less than or equal to 35 percent, sinus rhythm, cardiac dyssynchrony (interpreted as a QRS duration greater than 120 ms), and NYHA functional class III or ambulatory class IV symptoms despite being on recommended OPT. They recommended that these patients should receive CRT unless contraindicated and assigned a level of evidence of A (McAlister, Ezekowitz, Dryden, et al., 2007). The guidelines do not address patient age or comorbid conditions, which might also impact the benefit of CRT therapy. They also do not address the question of CRT versus CRT-D, which, according to one discussant “is still a source of contention in the community.” Nevertheless, the guidelines do establish a basis upon which to assess appropriate use and have reportedly significantly increased referrals for CRT by cardiologists, especially HF specialists.

CME Activities

Several CME activities were specifically designed to help disseminate the CER findings on CRT. These include a CME roundtable of key opinion leaders at the American College of Cardiology that discussed the evidence for CRT use as well as lingering clinical questions and the AHA’s GWTG–Heart Failure program, which combined registry and educational activities. Similarly, the IMPROVE-HF registry included decision support tools, utilization reports, and other information on best practice. These were cited by stakeholders as helping disseminate the CER results, although their main impact was only among proceduralists and cardiologists specifically interested in HF management.

Growing Prominence of Clinical Registries

Several registries now collect data on HF patients and CRT use, and they have had a significant impact on dissemination of the CER findings. The IMPROVE-HF registry, launched around 2007, the first large, comprehensive registry for HF in outpatient settings, includes 167 U.S. outpatient cardiology practices. Similarly, the GWTG–Heart Failure registry collects data from 228 participating hospitals. The ACC and AHA also maintain the NCDR, which includes an ICD registry. COMPANION and other CER trials involving CRT were all specifically cited in the rationale for the IMPROVE-HF registry’s CRT performance measure (Fonarow, Yancy, et al., 2007).

These registries are also the only significant source of tracking data on HF patients' long-term functional and morbidity outcomes and complication rates (McAlister, Ezekowitz, et al., 2007). Very little other information is available regarding CRT use outside of clinical trials (Piccini, Hernandez, et al., 2008). One key finding from the IMPROVE-HF registry is that appropriate use of CRT increased 29.9 percent and use of ICDs increased 27.4 percent (although it should be noted that other studies have shown no such trend in appropriate use between 2005 and 2008) (Fonarow, Albert, et al., 2010).

Current registries have some limitations, including technical and incentive barriers to acquiring necessary information. Documentation of functional status occurred in only 58 percent of the cases in the IMPROVE HF registry, limiting the ability to determine other patients' suitability for CRT-D (Fonarow, Yancy, et al., 2008). Documentation of patients' QRS duration, likewise critical, was also often missing (Fonarow, Albert, et al., 2010). Similarly, the AHA registry cannot definitively address underuse of CRT, since it does not include information on potentially eligible patients (such as QRS duration or NYHA classification). Stakeholders also pointed out that the IMPROVE-HF registry is sponsored by a device manufacturer. Finally, while the registries have divergent goals and focus, they are effectively competing with each other for participants, which makes it difficult to get representative overall data. As one stakeholder put it, "We [still] have a long way to go."

Adoption of CER Results by Specialized Clinics

About 41 percent of the outpatient practices enrolled in the IMPROVE-HF registry were using dedicated HF clinics. These practices had higher utilization rates for CRT and HF education, but not other process measures (Albert, Fonarow, et al., 2010). Some healthcare systems have likewise established dedicated clinics for HF, and similarly, many hospitals maintain "HF units" that use standardized protocols to ensure the use of best-practices guidelines (Arnold and Gula, 2010). Stakeholders reported that physicians specializing in HF treatment are well acquainted with the COMPANION results and are the primary source of appropriate referrals for the procedure.

Professional Certification-Organization Activities

While certification is not compulsory, the Heart Rhythm Society has established a "path to competency" for implantation of CRT and CRT-D devices. Thus, there are now standards for who should undertake the procedure and how it should be performed. At present, however, there appear to be limited incentives to comply with the standards, which blunts their impact. This may be offset in the future by the increasing proportion of physicians trained during their residencies or fellowships to do CRT and CRT-D procedures.

Publication of High Rates of Off-Label and Inappropriate Use

Stakeholders suggest that recently published studies that document the frequent inappropriate use of CRT and ICD therapy (Fein, Wang, et al., 2010; Al-Khatib, Hellkamp, et al., 2011) have had an impact. Practitioners have reportedly reduced the frequency with which they perform procedures not covered by evidence, out of concern that payers such as Medicare will react by establishing reimbursement criteria, refusing to pay for off-label procedures, or even demanding refunds from physicians for already reimbursed procedures that were not performed with an appropriate indication. Interestingly, this use

of CER evidence to inform Medicare payment decisions is explicitly limited by legislation (Affordable Care Act, Section 6303).

Accountable-Care Organizations

Experts believe that the development of ACOs will reduce inappropriate CRT use through several mechanisms, including the better alignment of financial incentives with outcomes—i.e., no longer rewarding just the procedure, but rather the thought process and outcome. They are also placing more emphasis on shared decisionmaking by patients and their doctors. Finally, it is anticipated that they will improve evidence-based care through the use of quality improvement and clinical decision support tools.

Integrating Clinical Decision Support into EHRs

Many stakeholders noted that primary care doctors need triggers and an effective alert system more than cardiologists do. Developing and integrating clinical decision support tools with EHRs at the point of care is one approach. Alternatively, EHR systems could be used to mine registries to identify patients who would be candidates for CRT or an ICD. Neither has yet been implemented in practice. However, the technology does illustrate what one stakeholder described as “the kind of thing HIT can do when it’s closely coupled to a clinical need and solid science.”

Conclusions

Uptake of the COMPANION study results has been uneven. Recent estimates indicate that there is both significant underuse of CRT among potentially eligible HF patients and fairly frequent CRT-D use in patients who lack an indication based on current CER evidence, including the COMPANION finding of a marginal benefit of CRT-D over CRT alone for only one secondary outcome.

In contrast to the other CER case studies, COMPANION engendered relatively few controversies in the generation and interpretation phases. The results were fairly readily accepted, the main disputes being over the degree to which they could be generalized to HF patients who did not meet the original inclusion criteria. Formalization of the COMPANION results was relatively rapid; specialty-society guidelines were updated promptly, which promoted their uptake, at least among proceduralists and HF-management specialists. No primary-care specialty-society guidelines were issued, and in addition, the specialty-society guidelines left open the appropriateness of CRT-D for patients who did not meet inclusion criteria. This and other factors contributed to an ineffective dissemination phase.

This case study illustrates several potential strategies for improving the CER dissemination phase. Essentially all the COMPANION dissemination activities targeted interventional cardiologists and HF specialists rather than referring physicians. Specialty societies, industry, and other CME producers (such as registries) also directed their educational efforts toward those groups. Most primary-care providers (who manage many HF patients) are still unaware of the COMPANION results. In addition, those primary-care providers and general cardiologists who took an interest were confronted by conflicting and ambiguous guidelines. This generated considerable confusion and a reported reluctance to refer patients for CRT. Future CER dissemination should focus significant effort on providers further upstream

in the decision pathway and should deliver clear, unambiguous referral criteria. However, the COMPANION case is not merely a cautionary tale. HF registries have had a significant positive impact by publishing high-profile studies illustrating inappropriate ICD use. Similarly, recent limited experience shows that clinical decision support tools can be very effective at prompting appropriate referrals and discouraging inappropriate procedures, but only if they are integrated smoothly into providers' routine workflow.

In the implementation phase, imprecise guidelines and evidence-neutral reimbursement policies may contribute to the use of CRT-D for inappropriate indications. Reimbursement policies, particularly Medicare's, significantly favor CRT-D implantation over CRT alone, despite evidence that adding the ICD has a very high marginal cost relative to the benefits it confers. As shown in other studies as well, referral to an interventionist is also tantamount to ordering a procedure. This tendency is compounded by open guidelines that allow CER results to be cited as justifying use in patients who would not meet study inclusion criteria. While primary-care physicians and some general cardiologists fail to refer many potentially eligible patients, dedicated HF clinics have been achieving appropriate referrals and avoiding inappropriate ones. Such clinics may serve as a model for implementing CER results. Currently, patients are not generally equipped to participate as fully informed partners in clinical decisions, and decision aids are not readily available, but it is likely that such decision aids could significantly improve implementation if physicians were given appropriate incentives to use them.

Key findings from the COMPANION case study are given in Table 5.2.

Table 5.2**Key Findings from the COMPANION Case Study**

| Phase | Key Findings |
|----------------|--|
| Generation | <ul style="list-style-type: none"> • There were very few methodological issues with this CER study. |
| Interpretation | <ul style="list-style-type: none"> • Arguments about generalizability—i.e., which patients would benefit—muddled the main result: that CRT-D worked. • Multiple professional societies involved allowed different interpretations to persist. |
| Formalization | <ul style="list-style-type: none"> • Rapid integration of the CER into guidelines promoted uptake. • Professional-society guidelines leave open the appropriateness of CRT-D for patients who do not fit the study inclusion criteria. |
| Dissemination | <ul style="list-style-type: none"> • Specialty guidelines came only from cardiology societies and were not disseminated effectively to referring primary-care physicians. • Conflicting messages about the generalizability of the CER results left interested primary-care physicians confused and reluctant to refer patients. • Both CME and industry dissemination efforts focused only on interventionists, not the more-influential upstream referring physicians. • Clinical decision support tools appear to improve appropriate use but must be integrated into physicians' regular workflow to be accepted. • Registries have dramatically illuminated both significant underuse and high levels of premature and inappropriate ICD use. • Registries were also effective in disseminating appropriateness guidelines and other educational materials to participants. |
| Implementation | <ul style="list-style-type: none"> • Reimbursement significantly favors ICD (hence CRT-D) implantation, despite its high marginal cost benefit relative to CRT alone. • Referral to an interventionist is taken as an endorsement and very frequently results in the procedure. • CER results are often cited as a rationale for device implantation even in off-label, noncomparable patients. • CER adoption was best in dedicated clinics/units (here, HF clinics). • Lack of mandatory certification programs has allowed an increase in procedures done by less-skilled operators, resulting in more complications. • Implementing patient decision aids for HF treatments may improve outcomes, but both |

referring physicians and interventionalists need more incentives to use them.

- The development of ACOs and similar incentive concepts is likely to reduce inappropriate use by aligning practice with evidence.
-

CPOE Case-Study Report

Clinical-Practice Context

Over at least two decades, concerns have grown about the prevalence and impact of medication errors. It is estimated that more than 1 million serious medication errors occur annually in the United States, contributing to 7,000 annual medication-error-related deaths (Kuperman, Bobb, et al., 2007). On average, a medication error is estimated to add \$2,000 to the cost of hospitalization. These numbers suggest that medication errors account for roughly \$7.5 billion a year nationwide in hospital costs alone (Leapfrog Group, 2008). Medication (prescribing) errors occur for many reasons. Traditionally, physicians have relied on paper-based, handwritten prescribing, followed by the actions of nurses, pharmacists, and others who administer medications. Prescribing errors are introduced when physicians accidentally prescribe the wrong drug or the wrong dose, overlook drug-drug interactions, fail to note patient allergies, or provide orders with illegible handwriting. While most of these errors are caught and corrected before they can harm patients, those that reach patients can have devastating consequences. Medication errors are also associated with wasteful spending, including payments for the wrong medications, and substantial costs for treatment of severe adverse events.

CPOE has been proposed as a way to reduce medication errors. The first CPOE system (called a “medical information system”) was developed by Lockheed Corporation and implemented at El Camino Hospital in Mountain View, Calif., in 1971. In the decades following implementation of this prototype system, CPOE has evolved considerably. For example, early systems could handle prescriptions but not refills. Although CPOE systems became substantially more sophisticated and embedded as potentially usable applications within hospital information systems, only a handful of hospitals actually implemented CPOE during the 1980s (e.g., Wishard Memorial Hospital in Indianapolis, Ind., and LDS Hospital in Salt Lake City, Utah).

Skepticism about the utility of these systems, their costs, the costs of installation and training, and the willingness of providers to use them were frequently stated reasons for not implementing CPOE. Hospital executives may also have sensed that better technology would be developed in a short time and would render any acquisition and installation obsolete.

Comparative-Effectiveness Question

The principal question at the time of the design of the CER study was whether CPOE could reduce medication errors and medication-related adverse events among hospitalized patients more effectively than nurse-focused, pharmacist-focused, or team-based interventions. During the 1990s, some experts were not persuaded that traditional paper-based ordering systems were a significant problem or that computer-based ordering systems alone (e.g., for medications and lab tests) would reduce the rate of medication errors.

Study Design Characteristics

To address this question, Bates and colleagues developed a CER study that compared the effectiveness of CPOE alone with CPOE plus a team intervention. The study, conducted within six units at Brigham and Women's Hospital in Boston, Mass., used a pre/post design that compared rates of medication errors prior to CPOE adoption with the error rate during the ten months following CPOE adoption. Data were collected in the pre-CPOE period in six units of the hospital; the same six units were used in the post-CPOE period, along with two additional units included to increase study power.

While most CER studies of medications or devices have compared these items to one another, this study compared two quality improvement interventions (CPOE and CPOE plus a team-based intervention). In the CPOE intervention, physicians could select from a menu of medications defined by the hospital formulary, with default dose and dose ranges provided for each, as well as automatic checking for common drug allergies and drug-drug interactions. The CPOE application was developed in-house and was embedded in the existing hospital information system. The team intervention centered on pharmacy-specific process changes, including changing the role of the pharmacist, standardizing labeling of intravenous bags, and implementing a pharmacy communication log so that the nursing staff could better communicate with the pharmacy staff.

The primary study outcome was the rate of unintercepted serious medication errors (preventable and unintercepted potential adverse drug events [ADEs]). Secondary study outcomes included the numbers of errors in each stage (ordering, transcription, administration, and dispensing of drugs) and also within specific categories targeted by the interventions (wrong dose, errors in concentration of intravenous solutions, etc.). Case finding was accomplished by reporting of incidents by nurses and pharmacists, solicitation of incidents by a study investigator interacting with staff, and patient chart reviews by a study investigator.

Study Results

Table 6.1 presents the key results from the CPOE study. Between pre and post periods in the same hospital units, unintercepted serious medication errors decreased from 10.7 events per 1,000 patient-days to 4.86 events per 1,000 patient-days—a reduction of 55 percent—and unintercepted potential ADEs declined 84 percent, from 5.99 per 1,000 patient-days to 0.98 per 1,000 patient-days. The analysis indicated that the team intervention had no incremental benefit over the implementation of CPOE alone

(hence the pooling of the intervention arms and the analysis as a pre-intervention/post-intervention study). These results suggested that other hospitals should consider CPOE adoption the principal quality improvement intervention to reduce unintercepted serious medication errors.

Table 6.1

Results of the CPOE Study

| Outcome | Pre Rate (events/1,000 patient-days, mean) | Post Rate (events/1,000 patient-days, mean) | Difference (%) | p-value |
|--|---|--|-----------------------|----------------|
| Unintercepted serious medication errors ^a | 10.70 | 4.86 | -55 | 0.01 |
| Preventable ADEs | 4.69 | 3.88 | -17 | 0.37 |
| Unintercepted potential ADEs | 5.99 | 0.98 | -84 | 0.002 |
| All ADEs | 16.00 | 15.20 | -5 | 0.77 |
| Unpreventable ADEs | 11.30 | 11.30 | 0 | 0.99 |
| All potential ADEs | 11.70 | 3.38 | -71 | 0.02 |
| Intercepted potential ADEs | 5.67 | 2.40 | -58 | 0.15 |

^aPrimary outcome.

Additional Evidence Following the Initial Release of CPOE Results

Since Bates’s seminal study in 1998, several other investigators have examined the impact of CPOE adoption on the incidence of medication errors. Overall, these studies suggest that CPOE has benefits, but they also illustrate substantial barriers to its successful implementation and adoption. A systematic review by Reckmann and colleagues concluded that “the amount of evidence is very modest and the quality and generalizability of results is limited.” The review highlighted the variety of outcome measures used to evaluate CPOE systems, as well as differences in measure specifications for related outcomes (Reckmann, Westbrook, et al., 2009). This variability appears to have contributed to uncertainty about the overall impact of CPOE among hospitals and physician practices that were considering adoption. Some studies reported many challenges related to CPOE implementation in some settings. Some also reported instances of negative effects of CPOE (Kashani and Barold, 2005; Koppel, Metlay, et al., 2005; Sittig, Ash, et al., 2006). The evidence on CPOE implementation problems has

tended to indicate issues with system design and the methods used to introduce CPOE into various settings (Ash, Stavri, et al., 2003). Poorly designed systems, poorly conceived implementation plans, and inadequate integration of CPOE into physician workflow may have accounted for the limited effectiveness of CPOE in other studies.

Impact of the Trial on Clinical Practice

CPOE uptake by hospitals in the period following the publication of Bates's CER results has been slow at best. By 2008, only 7 percent of the hospitals responding to a Leapfrog Group survey had CPOE systems used by a majority of physicians to place medication orders.

Key Barriers to Clinical-Practice Change

A number of factors are thought to have slowed CPOE adoption, many of which are not unexpected given the problems of introducing a new and complex technology (Rogers, 1995; Doolan and Bates, 2002). The key barriers to CPOE adoption cited in the literature and in discussions with experts are summarized below.

Complexity of the CPOE Intervention

The advantage of CPOE over existing practices is difficult for users to perceive, and CPOE tends to be incompatible with the typical user's workflow. ADEs are rare and the value of averted medication errors may not be fully appreciated because they may not be directly observed. In addition, busy clinicians must invest time to learn the new workflow. Finally, CPOE systems have been difficult for hospitals to test on a trial basis because they are usually embedded in larger hospital information systems that cannot be easily or inexpensively changed (see technical challenges below).

Lack of a Strong Business Case for Adopting CPOE

The cost of installing CPOE systems is high, and the return on investment is unclear given current payment systems. As one expert noted, "He who pays is not he who gains." Payers are more likely than providers to reap the financial benefits from implementation of these systems, through the avoidance of costs associated with treating medication errors (Birkmeyer, Lee, et al., 2002; Poon, Blumenthal, et al., 2004). Without payers "coming to the table" with proposals to share these gains, providers have had little financial incentive to adopt CPOE (Poon, Blumenthal, et al., 2004). Hospital executives may also be averse to the reputational risk of a CPOE investment and installation failing, as happened quite publicly for one early-adopting hospital (Cedars Sinai Medical Center in 2002). Vendor assertions about CPOE value are often considered suspect and self-serving and appear to have a limited role in hospital and physician decisionmaking. Some experts also stressed that financial decisions, particularly those using a return-on-investment (ROI) framework, tend to focus on relatively short time frames and not the years it may take to begin to see the ROI from CPOE adoption (if implementation is successful). As a capital-intensive investment, CPOE adoption may lose out to other projects offering a higher immediate ROI, such as new imaging or surgical facilities, even if physicians are in favor of it.

Technical Challenges of Integrating CPOE into Current Systems

CPOE is not a stand-alone product; it is typically integrated into EHR or clinical information systems, and it must also be compatible with other technologies such as pharmacy bar-code systems. Hospitals have too often attempted to integrate CPOE into legacy health-information systems designed to support administrative functions. These older systems, unlike EHRs, may be ill-equipped to handle the addition of a CPOE application, because key clinical data may be stored in separate systems using different technologies and thus are not exchangeable with the CPOE application. One expert explained that when hospitals attempt to absorb and build a CPOE application into a health-information system, “technology controls behavior, rather than behavior determining the needed technology.” Moreover, the increasing complexity of CPOE applications, including sophisticated add-ons such as clinical decision support, places further demands on existing health-information systems.

These technical challenges have created an “inefficient market” for CPOE, according to one expert. With many technologies, a dominant vendor emerges or multiple vendors compete, but they collectively reach consensus on standardized protocols (e.g., standard Internet protocols). CPOE systems, however, continue to lack interoperability, even with modern EHRs. While the CPOE market may mature and improve with time, substantial forces are working against interoperability and the broader sharing of information. In particular, lack of interoperability promotes sustainable business for individual vendors, as clients typically remain “captives” of that vendor if they want to make enhancements. Furthermore, hospital financial incentives tend to impede sharing of clinical data with competing institutions; in a fee-for-service environment, maximizing treatment volume is more important than realizing efficiencies through information sharing.

Until very recently, CPOE technology has not been a user-friendly quality improvement tool. CPOE systems are difficult to use, and vendor products differ sufficiently that clinicians who learn to use one cannot necessarily transfer that knowledge to use of another. CPOE products require substantial investments of staff effort and resources to set up and customize them to suit the local practice context (e.g., to define medication formularies or ensure that local pharmacies are represented in electronic prescribing databases). These factors form a substantial barrier, particularly for smaller hospitals or outpatient practices.

Clinician and Organizational Resistance

Many physicians and other clinicians hold negative views about CPOE, lack incentives to adopt the technology, and may strongly resist doing so. Past negative experiences with prototype EHR systems, a perceived negative impact of CPOE on workflow, and inadequate time for training may also create resistance. Completing all the steps to enter an order may be perceived as more cumbersome than giving a verbal order or handwriting one on a chart. Computing tasks may also be perceived by physicians as work that should be done by nurses and other staff. Such expectations can produce tension and conflict between professionals as workflow changes are implemented. The aversion to such changes among employees and organization managers is strengthened by highly publicized failed EHR implementation attempts (Morrissey, 2004). Despite the potential for net benefit, many clinicians tend to perceive successful adoption as unlikely (Doolan and Bates, 2002).

Limited Generalizability of CER Evidence from Quality Improvement Studies

Evidence of success of CPOE adoption at one hospital, no matter how meticulously documented and communicated, may fail to convince professionals and managers that implementation will be successful at other hospitals. The organizational and practice contexts that may enable implementation in some settings may not exist in others. Hospitals that lack health-information systems and those with legacy systems that cannot accommodate the addition of CPOE are unlikely to perceive CPOE implementation as an achievable goal. CPOE applications studied in seminal CER studies were designed and implemented within systems that were largely home-grown. While CER cannot and probably should not assess the effectiveness of an intervention in all imaginable settings, limited generalizability tends to weaken the persuasive power of CER results.

Lack of Knowledge and Guidance About Implementation

CER studies on quality improvement approaches may include descriptions of an intervention, but they rarely include detailed descriptions of how the intervention was implemented. Even hospital managers eager to implement CPOE may find that the implementation process is disruptive and protracted. The required coordination and cooperation among relatively independent professionals increases the risk that implementation will fail. While specific implementation guides for CPOE do exist, a technically skilled workforce, including talented HIT professionals, is necessary to ensure successful implementation (Ash, Stavri, et al., 2003). Learning by doing and trial and error are frequently part of CPOE adoption, suggesting that implementation guidance is only somewhat helpful.

Mistrust of Vendors

There is a widespread sense among hospital managers that the business case presented by vendors lacks credibility, especially since vendors have a strong financial incentive to peddle their own wares. The commitment of vendors to the long-term success of CPOE implementation is perceived to be low. Many lack a track record of successful implementation, and instances of vendors selling systems that fail to meet functionality needs also make hospital leaders and clinical professionals wary of newly developed and unproven systems.

Key Enablers of Clinical-Practice Change

Financial Incentives That Improve the Business Case for Implementation

If financial incentives are better aligned to produce an ROI, hospitals are more likely to adopt CPOE. In general, experts believe that the financial incentives available through the Health Information Technology for Economic and Clinical Health Act (HITECH) to adopt and become meaningful users of EHRs, combined with imminent financial penalties for failing to do so, have had a positive effect on adoption. While success is difficult to gauge at this point, well-designed mandates and financial incentives seem to be somewhat effective. Many experts believe that the HITECH incentives are persuading some hospitals to adopt, because, for a very large medical center, millions of dollars may be at stake. Efforts to develop a framework for sharing the financial benefits of CPOE (as well as the cost) between providers and payers could help speed adoption.

Some experts are skeptical, however, noting that some hospitals, even if penalized for not having CPOE, will still be reluctant to adopt it in the near future. There is ongoing concern among hospital executives, CPOE system vendors, and physicians that mandates and incentives should be introduced on a timeline that allows providers to deploy resources most effectively. Meaningful-use requirements may have the paradoxical effect of slowing progress, in the view of some experts. For example, meaningful use has led some CPOE vendors to put everything on hold for 18 months to await the release of the new requirements instead of continuing to work with hospitals to improve existing systems.

The importance of financial incentives in accelerating CPOE adoption is illustrated by two cases that demonstrate how adoption can occur when such incentives are well aligned with an organization's interests: The Kaiser-Permanente system and the VA health system are integrated health delivery systems, so both the costs and the cost savings from implementing CPOE accrue to them. Other incentives to adopt CPOE, aside from those incorporated in HITECH, may also be increasing in settings that have been traditionally organized around the fee-for-service payment system, as these organizations adapt preemptively to payment reform efforts such as expected-value-based purchasing and ACOs. Whether these new payment models will accelerate adoption of CPOE in nonintegrated health systems remains uncertain.

The Importance of Mission and Commitment of Local-Organization Leadership

The results of CER studies on CPOE may help strengthen local organizational leaders' commitment to adopt CPOE and integrate it into their organizations' missions. Published literature suggests that when hospitals make adoption a priority, implementation is almost always successful (Ash, Stavri, et al., 2003). For example, when a hospital's mission is to make patient safety a top priority and CPOE is stated as being a crucial part of fulfilling that mission, CPOE adoption and successful implementation appear likely. However, few hospitals may be able (or may want) to formulate or articulate such a mission. Successful CPOE adoption has been found to be strongly associated with—and nearly always requires—strong leadership, a long-term commitment of resources, involvement of physicians (including engagement of physician champions), leveraging of relevant staff (e.g., young physicians), and assuring responsiveness of the information technology (IT) department to address problems and complaints quickly.

Publicity May Motivate Change

The impact of publicity about the problem of medication errors on adoption of CPOE is uncertain, but the IOM report *To Err is Human* clearly fostered urgency about the need to reduce medical errors more than a decade ago. Advocacy by CPOE vendors, employers, and employer coalitions such as the Leapfrog Group appears to be increasing. Policymakers have also advocated for change. An op-ed in the *Washington Post* in response to the IOM report cited CPOE as the way to save “tens of thousands of lives every year,” as well as “huge amounts of money” (Gingrich, 2000); the piece called for Congress to consider passing a bill “requiring that within three years every doctor's prescription and every patient's record be computerized.” The impact of any specific dissemination activity or publicity effort on technology adoption is difficult to estimate, but collectively, mass-media publicity efforts are thought to

motivate change in a positive, although perhaps incremental, fashion (Lasalandra, 1998; Burling, 2001; Johnson, 2001; Finley, 2009).

Learning from CPOE-Champion Hospitals and Failed Implementations

One early CPOE adopter used its experience with an implementation effort that initially failed to develop a simple tool that other hospitals could use to estimate the benefits of adoption based on their own specific data. These efforts may help to address the generalizability problem raised in the peer-reviewed literature and may provide a more unbiased estimate of the potential ROI than that provided by vendors. Stakeholders mentioned that large, champion hospitals often work with vendors to improve CPOE content and the quality of off-the-shelf systems. Such efforts may be very beneficial to small hospitals if they result in better vendor products that require substantially less configuration and modification. The magnitude of the efforts is hard to quantify, but it is believed that they are positive. Failed CPOE implementations may also provide lessons learned for hospital or physician leaders considering adoption.

Identifying Options for Making Quality Improvement Interventions More Affordable

The high cost of CPOE adoption was consistently cited as a barrier. Financial incentives like those created by HITECH may help to offset the costs of acquiring CPOE systems, but they may paradoxically undercut price competition among vendors. Payers might find it beneficial to bear some of the providers' CPOE acquisition costs. Alternatively, they might share some of the gains that could accrue from more efficient care. Malpractice-insurance discounts for hospitals with CPOE were cited by one expert as a potential way to offset the costs of adoption.

Developing Standard Interventions and Implementation Strategies

Early negative experiences with CPOE implementation (such as overly sensitive drug-drug interaction alerts) might have been averted if a national database or set of guidelines had been developed. Instead, hospitals have been allowed to customize quality improvement applications, particularly alerts. In theory, this can be dangerous, because patients receive treatment in multiple settings in which the CPOE systems may be using different clinical logic or alerts. Standards for interoperability might be beneficial, particularly given some of the disincentives for information sharing. More information on vendors' track records and their commitment to the CPOE market, along with evidence of successful long-term relationships with hospitals could reduce hospitals' uncertainty about investment risk. Development of system standards is a core focus area of the Office of the National Coordinator for Health IT, and standards for CPOE may be an area for further development in the future.

Simplification of Quality Improvement Interventions

One stakeholder told us that CPOE systems were so complicated that vendors often did not know their true capabilities—sometimes forgetting that their own systems contained certain functionalities. While certification processes have helped ensure product quality, it appears that hospital decisionmakers still face considerable uncertainty when trying to compare and select CPOE vendors and systems. Reducing the complexity of these systems to reduce the time needed to learn to use them and to enable adequate piloting of systems prior to making an investment could encourage their adoption.

Disseminating Information on Implementation in Addition to CER Results

The general consensus among experts is that CER results on the benefits of CPOE are persuasive. Documenting those benefits in terms of reduced medication errors, dollars saved, or improvements in staff workflows would be beneficial and could incrementally encourage adoption. It is less clear whether sufficient evidence exists regarding successful implementation strategies in different settings. Producing more evidence about implementation may overcome providers' ambivalence about the risks associated with it.

Improving Dissemination of Implementation Examples Through Channels Other Than Peer-Reviewed Publications

Some of the greatest CPOE success stories remain unpublished. Furthermore, one expert doubted whether academic publications even reach the relevant decisionmakers. While the effect of dissemination through the peer-reviewed literature is unclear, efforts to encourage champion hospitals to publish data and lead in promoting change are clearly important. Vendors may also have a role in describing CPOE successes and failures, although such a role raises concerns regarding vendors' conflicts of interest.

Conclusions

The adoption of a new medication or device that does not require significant changes to the workflow of clinicians and staff but instead funnels through the existing workflow is relatively easy, but the adoption of quality improvement strategies faces additional barriers because of the significant changes to organization, financing, and staff work that may be required. These barriers may neutralize the impact of even outstanding CER evidence.

The CPOE case study provides a number of lessons about translating evidence on CER involving delivery systems interventions into new practices. We selected CPOE as an example of a delivery-system intervention for which there is published CER evidence, but it is worth noting that CPOE not only has features that are typical of many such interventions but also has some that may be easier to implement. First, CPOE is both a new technology and a new set of workflow requirements, making it complex and requiring substantial up-front investment, as well as coordination, communication, and long-run commitments from numerous stakeholders with potentially conflicting goals. The staff involved directly in implementation of CPOE must undertake nontrivial workflow changes; as with other technology-based quality improvement interventions, effective use of CPOE requires dramatic changes in individual process and social interaction with peers.

Second, CPOE is a variable technology with evolving features and functionalities across a wide range of hospitals and vendors, depending on their needs and existing HIT capabilities. This poses challenges for end users (particularly hospital executives) who seek to use CER evidence for decisionmaking. Our case study suggests that these individuals often struggle to conceptualize the intervention and consequently may find it difficult to assess the applicability of the results to their own setting.

Third, the financial investment required to implement CPOE is large, and key leaders must have clear reasons and plans for implementation to overcome resistance from staff. Financial incentives appear to improve the business case for such a large investment. The experience of early adopters suggests that organizational factors and mission can be significant enablers even when financial incentives are not aligned.

Fourth, the range of target stakeholders for CER results concerning CPOE is broader than that of the typical users of other types of CER studies. Stakeholders include hospital executives, technology vendors, physicians, pharmacists, and other clinical staff. This may increase the complexity of messaging to effectively and consistently disseminate the CER results.

This case study illustrates that quality improvement interventions based on CER evidence (particularly those that improve patient safety) may benefit from some combination of strong mandates, systematic standards, and financial incentives that improve the business case for implementation.

Key findings from the CPOE case study are given in Table 6.2.

Table 6.2**Key Findings from the Bates CPOE Case Study**

| Phase | Key Findings |
|-------------------------------|--|
| Generation/ interpretation | <ul style="list-style-type: none">• Published studies tend to document unique CPOE implementations, which are perceived as not being generalizable. |
| Formalization | <ul style="list-style-type: none">• Adoption has been retarded by the significant tension between the needs for uniform standards/interoperability and local adaptability.• Vendors prefer a captive market, a disincentive to standardization. |
| Dissemination | <ul style="list-style-type: none">• High-profile reports and popular media coverage have helped promote public awareness of medication errors and have shown that CPOE is a demonstrated solution.• Advocacy by health-care purchasers, CPOE vendors, and patient safety groups also promotes awareness.• There are still too few well-documented examples of <i>how</i> to successfully implement CPOE.• Failed implementations have discouraged potential adopters, but lessons learned have led to the development of champion hospitals and implementation tools.• The usual peer-reviewed-publication channels may not reach CPOE decisionmakers. |
| Implementation | <ul style="list-style-type: none">• CPOE is complex, may be incompatible with legacy systems, and will overturn current provider workflow; therefore, adoption requires careful coordination between stakeholders.• Clinicians tend to skeptically resist adoption unless they perceive direct benefits to them or their patients but are usually enthusiastic users after implementation occurs—especially if quality improvement data show real improvements in patient safety.• Implementation of very expensive technology is more likely if the entity that bears the implementation cost also derives financial benefits.• Meaningful-use mandates and penalties for failure to implement HIT will very likely also drive implementation. |

Factors Influencing the Translation of CER Research into New Clinical Practices: A Synthesis of Themes from the Case Studies

As the preceding chapters illustrate, a myriad of factors influence whether CER translation into clinical practice is ultimately successful. However, our synthesis suggests that some of these factors are “root causes,” i.e., they are fundamental and may represent high-leverage points for action to improve adoption of a new practice. We have identified at least five root causes of the failure of CER to change clinical practice. They manifest themselves in somewhat different ways across the case studies, appear to explain the strategies of many stakeholders who have an interest in CER, and typically exert their effects over multiple phases of the CER translation process. In the following, we describe the root causes in order of importance and describe the various mechanisms through which they impede the adoption of new clinical practices. We also describe the limitations of our study. In Chapter Eight we discuss the policy implications of our findings and suggest opportunities for future research that would either address gaps in our analysis or that would build directly on the lessons learned from this study.

Root Causes of Failure of CER to Change Clinical Practice

1. Financial incentives are primary drivers of the adoption of new clinical practices, whether or not these practices are supported by the CER evidence. CER results that threaten the financial interests of a stakeholder will be challenged at all phases of the translation process.

The most fundamental determinant of successful CER translation is the extent to which the economics of adopting a new clinical practice are favorable to providers and patients. Not surprisingly, the marginal profitability of a procedure smooths the way for implementation of a supportive CER finding and acts as a strong deterrent to the implementation of one that suggests that a more expensive therapy may not be superior to a less expensive one. Our case studies of the comparative effectiveness of interventional and noninterventional procedures highlight the perverse consequences of fee-for-service reimbursement in the face of CER evidence showing little or no marginal benefit. Rates of reimbursement for PCI, CRT-D, and spinal stenosis surgery are orders of magnitude higher than those for alternative treatments, and many discussants noted that once patients were referred to interventional specialists, even if only for consultation, there was a high likelihood that they would receive an invasive procedure. In the words of one HF physician, interventionists are “mainly looking for contraindications [at that point].”

CATIE is a notable exception, in that prescribers typically have little financial stake in the choice of antipsychotics, so change in the implementation phase could not have been driven by financial considerations. But the lack of practice change following COURAGE and SPORT despite seemingly equivalent effectiveness of treatments and the modest benefit of surgery, respectively, illustrate the tendency to prescribe a more costly treatment when reimbursement policy is not well aligned with the CER evidence base. The high rates of inappropriate use of CRT-D and complex spinal surgery after the SPORT results were published reinforce this general conclusion.

While the role of financial incentives in the utilization of expensive therapies is fairly well known and may by itself explain the failure of implementation of a CER-based practice, our case studies highlight their role in influencing more than just the implementation phase. Financial incentives may have a greater influence on the adoption of new clinical practices than CER evidence in the following ways:

- Stakeholders with a financial interest in the outcome of a CER study may seek to influence the study design to increase the odds that its results will favor them, or they may initiate efforts to critique and thus undermine potentially unfavorable CER studies at the time participants are being enrolled. According to several physician discussants, the pharmaceutical and device industries, respectively, developed campaigns to raise doubts about the design of the CATIE and SPORT trials even before their findings were known. Critiques of CER study design by interested stakeholders may peak at the time results are released to maximize the likelihood that the studies will be viewed as methodologically weak. Many physician discussants believe these messages are not likely to be a critical factor in the way the average practitioner interprets the studies, but their impact on the formalization of CER evidence through guidelines, quality measures, and clinical decision support may be important, and this may have downstream effects on local implementation.
- The interpretation of CER results through a dynamic scientific debate among stakeholders appears to be influenced by financial incentives of the participants. In the case of CATIE, industry sponsorship of key opinion leaders and aggressive detailing practices reflected a high degree of pharmaceutical-industry effort to influence the interpretation of the results in a manner that was favorable to its financial interests. Subspecialty societies with and without potential financial gains publicly disagreed about the interpretation of COURAGE and SPORT results.
- The formalization of guidelines and quality measures based on CER evidence may be influenced in subtle ways by financing. Historically, most guideline and quality measure development activities have been supported by professional societies, industry, or some combination of the two. Conflicting subspecialty guidelines are a well-known problem, although this may begin to change in the wake of the recent IOM recommendations for improving the consistency and credibility of clinical-practice guidelines (Institute of Medicine, 2011). Professionals have few financial incentives to facilitate the development of performance measures unless they will be paid based on the measure results. Discussants noted that guideline recommendations for spinal surgery tracked closely with the mission of the sponsoring professional society and were likely to conflict with one another. Moreover, some guidelines may not change in a timely way

in response to CER evidence. For example, the APA took the position that formulary restrictions and other policies of payers should not be altered by CATIE results, instead choosing to defend prescriber autonomy. The prescribing guidelines were eventually modified, but other formalization activities such as the modification of quality measures were slow or nonexistent.

- Advocacy is often directed against payers in public forums and may have the effect of limiting their willingness to formalize CER evidence into coverage policies or utilization review policies. All of our payer discussants indicated that their organizations are averse to restricting coverage for treatments that are well established. Rather, they selectively pursue areas in which practice variation is extensive, where evidence clearly does not support the practice, and where risks to patients are unambiguous. Two examples of areas targeted by payers are the use of antipsychotics in pediatric populations and lumbar fusion surgery. Payers appear to be less willing to pursue prior authorization policies to avoid facing protracted and costly challenges by professional societies. In addition, payers who have narrow service areas may be more concerned with preserving good public relations, and public payers and publicly funded organizations like the VA may face additional political pressures to preserve generous coverage for their beneficiaries. When payers do implement prior authorization policies, they may rely on industry-standard medical-necessity criteria, which tend to be rather permissive.
- The dissemination of new practices based on results is an expensive undertaking. Detailing and advertising of new procedures and treatments are in the interest of the industry involved, but counterdetailing is rarely supported unless payers or the public choose to back it. Indeed, the lack of financing for dissemination activities to support CER-based practices may be an important impediment to practice change. Provider demand for clinical decision support tools and patient decision aids may drive vendors' efforts to develop these tools, but with the exception of meaningful-use standards, there is little financial incentive for physicians to use them.

Despite the seemingly powerful influence of financial incentives favoring both the status quo and an accelerating panoply of new procedures, recent trends, including the emergence of new models of provider organization and payment, provide some basis for optimism that CER evidence can be more influential in the future. Physicians continue to abandon private, independent practice in favor of health-system affiliation and increasing reliance on salary-based payment. This trend may dampen their exposure to perverse financial incentives associated with fee-for-service payment. Physicians working within a larger organizational context may be more likely to use performance measurement and feedback, patient decision aids, clinical decision support tools, and registries, all of which have the potential to increase responsiveness to CER. Discussants commented that such tools had the potential to shed light on inappropriate utilization or other quality concerns that might prompt involvement by payers or policymakers. Value-based and episode-based payment approaches may also encourage more cost-effective care, particularly if these models encourage quality improvement.

Payers indicated to us that they are actively engaged in horizon-scanning for CER evidence that may form the basis of policies before practices become widespread in the community. While our case studies focused on well-established practices, newer CER may focus on less-widespread practices and may involve less-controversial treatment comparisons. Activities that curtail the influence of financial interests in each of the phases of CER translation (such as the recent IOM report calling for greater transparency and integrity of the guideline development process) may reduce the countervailing forces that work to undermine or neutralize even the best CER evidence.

2. Even the best CER studies may fail to produce an unambiguous “winner,” so it may be difficult to achieve a consensus interpretation of the results.

The promise of CER lies in its potential to increase the use of the safest and most effective treatments and decrease the use of ineffective treatments or those for which an equivalent, but less expensive treatment exists. CER studies that produce clear “winners” (showing unambiguously that one treatment is better than another or that two treatments have essentially equivalent effectiveness) should be more likely to change practice because their results are difficult to challenge. For example, the Women’s Health Initiative demonstrated conclusively that hormone replacement therapy increased the risk of clotting and MI. However, our case studies suggest that even among the best-designed and -conducted CER studies, ambiguous results are likely to be common. Moreover, persuading stakeholders about treatment equivalence may be much more difficult than persuading them of the superiority of one treatment over another, because equivalence is generally defined across multiple outcomes, while claims of superiority are often based on a statistically significant difference on a single outcome.

SPORT is a typical example of a CER study that was plagued by methodological concerns and failed to produce a clear winner. Given ambiguous results, professional societies interpreted the results differently, leading to contradictory guidelines. CATIE and COURAGE, two well-executed CER studies, found comparable effectiveness of alternative treatments, yet their impact on practice has been muted because of limitations in their designs. Debate about the meaning of the trials and their applicability beyond the included populations has continued for years. The CPOE trial produced unambiguous improvements in safety, but the generalizability of the results was questioned. COMPANION (as well as the subsequent CARE-HF trial) demonstrated a relatively unambiguous significant survival benefit of CRT, but an unclear benefit of CRT-D vis-à-vis CRT due to a positive finding on a single secondary outcome.

Many factors increase the risk of an ambiguous result from a CER study. Comparison of two somewhat effective treatments may result in smaller differences in expected effects than would result from a placebo-controlled study. The use of active comparison groups, more-inclusive patient populations, and real-world practice settings may either level the playing field, introduce greater statistical noise, or do both. Stakeholders may differentially weight the importance of study end points (e.g., clinical effectiveness, side effects, safety), contributing to different interpretations of the results. Meanwhile, stakeholders may differ in their equipoise for recommending treatments and therefore may view the same effect size differently. All of these factors complicate the interpretation of findings. In the CATIE case study, some stakeholders focused on the equivalent effectiveness outcomes between conventional

and atypical antipsychotics to argue for step therapy, while others cited heterogeneity in benefits and harms of individual drugs to argue for maintaining open access to all antipsychotics. The nonsurgical spine community (and several spine-surgeon discussants) found the benefits of spinal stenosis surgery to be marginal, while most surgeons found that the results confirmed prior, smaller studies demonstrating the superiority of surgery.

CER studies that produce ambiguous results open the door to selective interpretation, which tends to undermine consensus that could facilitate guideline updates by professional societies or the formation of coverage policies by payers. In the case of CATIE, SAMHSA, the entity responsible for translating mental health research into practice, pursued a neutral policy of “greater patient engagement” because it viewed the CATIE results as “mixed.” While the American College of Cardiology typically drafts consensus statements following key trials, it failed to do so following COURAGE. Several payers indicated that they avoid focusing on clinical areas that lack clear winners to ensure that providers have enough flexibility to use clinical judgment when making treatment decisions. Payers are more likely to focus on off-label uses (e.g., antipsychotics for kids), practices out of the mainstream (e.g., dosing limits that are multiples of recommended limits), and procedures that display substantial variation in rates (e.g., spinal fusion surgery). In many states, payers are permitted by law to use cost-effectiveness data for coverage decisions as long as treatments can be considered equivalent, but operational definitions of equivalence are incomplete.

In cases where one treatment is unambiguously harmful, clinical practice has been known to change rapidly, and our findings confirm to some extent that general rule. Anecdotal reports suggest that the use of olanzapine, which had the most severe metabolic side effects (and greatest efficacy) among the medications assessed in CATIE, may have declined in the post-CATIE period, indicating that physicians may have perceived that the metabolic risks exceeded the benefits. The use of performance measurement, including mortality and complication rates, both within and outside of registries may also help provide data on treatment harms in the future and may influence providers’ decisions to recommend PCI, CRT, and spinal stenosis surgery or payers’ willingness to implement coverage policies. Thus, adverse event data from registries may help to identify more unambiguous losers over time.

While ambiguity may lead to incomplete use of CER results and limit the potentially attainable change in clinical practice, discussants emphasized that the lack of winners does not invariably mean that the CER fails to have an impact on clinical practice. Many indicated that identifying winners is not the goal of CER; the goal is to generate information to help physicians and patients arrive at satisfactory treatment decisions. Discussants indicated that CATIE, COURAGE, and SPORT reassured physicians that using moderate-dose conventional antipsychotics or less-aggressive therapies can confer equivalent or nearly equivalent benefits to patients. However, only an empirical study can confirm this hypothesis.

3. Cognitive biases play an important role in stakeholder interpretation of CER evidence and may be a formidable barrier in all phases of CER translation.

Our case studies suggest that at least three cognitive biases may influence the way physicians and other stakeholders interpret new CER evidence. First, confirmation bias, the tendency for a stakeholder to embrace evidence that confirms preconceived notions of treatment effectiveness and reject evidence to the contrary, may reinforce established practice patterns. Confirmation bias may be more prevalent in a CER context than in other types of clinical research, because treatments have often been in use for years and providers may be emotionally invested in them. In the years preceding CATIE, psychiatrists had been exposed to overwhelmingly positive research and messaging about atypical antipsychotics, which contributed to skepticism about CATIE's unexpected findings. Both CATIE and COURAGE demonstrate how confirmation bias may have led stakeholders to dismiss CER results that challenged widely held treatment paradigms and, instead, to criticize the studies on methodological grounds. In our case studies, confirmation bias and financial incentives reinforced one another (with the possible exception of CATIE for prescribers), making it difficult to separate the magnitude of each effect. Stronger study designs (particularly emphasizing the generalizability of findings) and careful monitoring of study conduct (particularly to prevent crossover) may preempt these critiques and counteract the influence of confirmation bias.

A second cognitive bias is the belief that aggressive intervention (even for low marginal benefit) is better than inaction. This belief appears to be a potent driver of the use of PCI in patients with stable angina (the so-called "oculostenotic reflex"). It may be reinforced by perverse financial incentives and by providers' perceived risk of malpractice liability if they decline to recommend the procedure. Payer discussants noted that physicians perceive that they "are not protected if there is an adverse outcome, even if you follow the evidence." While patient preferences may also drive the use of more-aggressive treatment, providers may be less likely to attempt to convince a patient about a different course of treatment than the one favored by the patient because these situations may heighten their perceived exposure to a malpractice suit. More complete data on treatment harms or heterogeneity of benefits may promote greater equipoise among both physicians and patients. In addition, the use of utilization review with feedback to providers—particularly outlying physicians—can be effective in helping them recognize aberrant utilization. However, payers typically do not have all the elements within their administrative data systems to be able to measure appropriate utilization.

A third cognitive bias (at least in the United States) is the widespread tendency to perceive new technologies as superior to older technologies. Stakeholders who have financial interests in new technologies may also disseminate messages that reinforce this bias. For example, atypical antipsychotics were often marketed as "second-generation" medications, implying that they offered improvement over conventional antipsychotics. But, as one discussant noted, "We didn't call beta-blockers 'second-generation anti-hypertensives.'" Lengthy CER studies, such as COURAGE, are open to the criticism that technology has advanced, and this argument may carry weight with the average practitioner who observes the rapid change in medical technology first-hand. This is particularly important for the interpretation of CER studies involving devices. Adaptive study designs could provide the flexibility to allow the evolution of treatments through the course of a trial, but such designs were only utilized in one of the trials we studied.

Strategies for mitigating cognitive biases are available, but their effectiveness is not completely clear. Enhancing the transparency of stakeholder positions by using approaches that foster explicit formal decisionmaking processes could mitigate cognitive biases in a policymaking context. Disclosure of financial and intellectual conflicts of interest is another strategy used by the IOM and others, and regulating detailing and direct-to-consumer advertising may also be effective. These options are discussed in greater detail in Chapter Eight.

4. The questions posed by a CER study and its design may not adequately address the needs of end users or focus adequately on the decisionmaking opportunities with the greatest potential to influence clinical practice.

Our case studies suggest that multiple end users have potentially unrealistic expectations of CER studies. Even successfully designed and executed studies appear to disappoint some stakeholders. This occurs for a number of reasons. First, there is an unavoidable tension in the design of CER studies between the goals of producing treatment effectiveness data for selected populations (which would support personalized medicine) and producing general estimates of treatment effectiveness for a broadly inclusive population (which would support generalizable results). COURAGE and SPORT were designed to produce general population estimates (the second category of studies) and have been criticized for producing insufficiently detailed data to inform clinical treatment decisions for specific patients. While these two studies provide useful data on the average course of patients who pursued less-aggressive therapy, they provide limited data to enable the tailoring of treatment for individual patients. In SPORT, a sizable percentage of patients did not respond clinically to spine surgery, suggesting significant heterogeneity in benefits. Similarly, CRT may not have been disseminated widely because clinicians noted that a substantial population did not respond to it, but they lacked information that would allow them to predict which patients would respond. Assessing important causes of heterogeneity up front, if known, and conducting pre-planned analyses may be one way to deal with this tension; however, not all CER studies will be powered to detect such effects. And while registries may provide greater power to develop prediction rules, they lack the randomization of the original CER studies.

CER studies may not be designed with a comprehensive or explicit understanding of the beliefs and concerns of clinical practitioners. For example, CATIE focused on the relative effectiveness of different types of antipsychotic medications, but it had neither adequate power nor sufficient follow-up time to evaluate differences in safety—particularly the incidence of tardive dyskinesia. In retrospect, it was apparent that side effects were a chief concern for many practitioners, so some might argue that the question posed (i.e., relative effectiveness) was less compelling than questions about potential side effects and personalization of therapy. In this light, the trial design may not have been optimized to inform clinical practice. Similarly, hospital executives, who control decisionmaking regarding delivery interventions like CPOE, did not have relevant information on the applicability of the CER results to their particular settings.

Finally, many CER studies focus on a single step in a clinical decision algorithm. This is scientifically necessary, to provide clarity about the specific question the CER study is designed to address. However, the matching of scientific questions to high-leverage clinical decision points is critically important if the goal is to optimize the quality and costs of clinical practice. Head-to-head comparisons of treatments may help providers select appropriate treatments in the latter stages of clinical decision algorithms, and this is clearly important. However, upstream decisions about whether to pursue diagnostic tests or procedures may have a greater impact on patient outcomes and the overall value of care, especially if the downstream treatment decisions involve equivalent effectiveness of alternative treatments. Our discussants noted that the design of COURAGE focused on a downstream decision point at which interventionists were not at equipoise. CER that develops data for clinical-prediction rules and diagnostic algorithms may overcome this problem. For example, some current trials are focusing on the role of stress-testing to better guide the decision about whether to refer a patient to angiography. Similarly, diagnostic imaging has emerged as a key step in pathways leading to both PCI and spinal stenosis surgery. If providers such as primary-care physicians are responsible for upstream decisions to refer, and if both providers and patients have weaker incentives than interventionists have to choose an intervention, their decisionmaking may be more readily influenced by evidence.

5. Clinical decision support and patient decision aids can help to align clinical practice with CER evidence, but they are not widely used.

The use of clinical decision support in many areas of healthcare is limited, and the use of patient decision aids is suboptimal. Our SPORT, COURAGE, and COMPANION case studies confirm this. Furthermore, perverse financial incentives and lack of accountability for implementation have limited the production and dissemination of both of these tools. Limited adoption of CPOE is a prime example of the larger problem of inadequate health-information infrastructure that can support more consistent and effective use of CER to inform clinical decisions. For example, the absence of EHR-based algorithms that would identify potentially eligible patients and alert physicians to consider referral for CRT is an important missed opportunity leading to underuse of that technology. In general, decision support tools to promote evidence-based diagnostic testing and appropriate referral to specialists are uncommon, although both may lead to the use of treatments that are better aligned with CER evidence. According to one of our discussants who specializes in CER, one of the purposes of decision support is to increase the confidence of primary-care physicians in managing patients who do not require referral to specialists.

The consequences of limited patient empowerment in medical decisionmaking are significant. Many patients are unaware of the relative benefits and harms of treatment, and in some cases they greatly overestimate benefits and underestimate harms. For example, patients undergoing PCI for stable angina often mistakenly assume that the procedure reduces their risk of future heart attacks, which is not supported by the COURAGE findings or any other body of evidence. Well-informed patients who make treatment decisions according to their preferences may serve as a counterweight to providers who lack equipoise, particularly in the case of PCI and spinal stenosis surgery. Moreover, patient satisfaction with care can be significantly enhanced when decision aids are used. Although numerous medical publishing

companies are developing products tailored specifically for patients, and professional societies have developed patient-oriented websites, both appear to be only weakly integrated with clinical processes, and their effectiveness is unclear. In contrast, direct-to-consumer advertising has become more prevalent and may create or reinforce misconceptions about treatments. The incorporation of CER evidence into direct-to-consumer advertising is an area that requires further study.

Even if incentives were better aligned and providers were held accountable for their use, it is unclear how to best incorporate clinical decision support and patient decision aids into clinical practice. Both sets of tools must be seamlessly embedded in clinical workflows. Physician discussants noted that primary-care physicians operate under extreme time pressures and re-source constraints and may have limited willingness or even opportunities to integrate these tools. While many payers are beginning to experiment with shared-decisionmaking initiatives, progress has been slow. Most tools are based on first-generation technologies that tend to overremind providers and disrupt workflow. Limited HIT infrastructure may continue to pose problems for the implementation of clinical decision support in the coming years unless CMS's EHR Incentive program succeeds in transforming the HIT capacity nationwide. Finally, providers need significant training to develop competence in helping patients successfully use shared-decisionmaking tools.

Limitations of This Study

We used an exploratory case-study methodology, examining five carefully selected CER studies of the recent past to identify some of the potential root causes for the failure of CER to change clinical practice. We anticipated that our results would inform the development of subsequent qualitative or quantitative studies using more-rigorous designs. Given the need to inform the current portfolio of federally funded CER, our priorities were to distill the key lessons from our case studies, using expert opinion from a broad range of stakeholders in a relatively short time frame. Because of these factors, a number of limitations regarding our study's scope, methodology, and generalizability should be kept in mind when interpreting our results.

Scope

We struck a specific balance between depth of content and breadth of inclusion of case studies, because our primary objective was to identify and synthesize themes across cases. Since we had a limited sample of expert discussions and stakeholder perspectives for each case study, we emphasized the early phases of the CER translation process, limiting discussions about the local implementation phase, on which there is a large existing literature. As a result, we had only a limited number of discussions with primary-care providers, particularly those who practice in nonacademic settings. Community-based providers might have perspectives that differ considerably from those of the opinion leaders with whom we tended to speak. In particular, we might not have identified the full range of implementation barriers that individual physicians confront when making treatment decisions on the basis of CER evidence. A larger sample of end users from a diversity of practice settings could identify additional barriers and enablers. In cases where we could undertake only a small number of discussions per stakeholder group, the information provided by each stakeholder might not have been representative of the larger group.

In addition, we did not have time for discussions with some stakeholder groups, including popular media and individual patients.

Despite considerable effort and multiple invitations, the device industry and, to a lesser extent, the pharmaceutical industry were underrepresented in our sample of discussants. While we were able to identify potential discussants, most of them declined to participate, citing concerns about the sensitivity of the information they might be asked to share. The perspectives of the device industry would have been useful for the COMPANION case study and, to a lesser extent, for SPORT and COURAGE. Our retrospective study design also proved challenging in this regard, as general industry turnover meant that many potential discussants were not employed by the industry in a suitable capacity at the time the results of studies such as CATIE were released and so were unable to provide insight into the strategies and dissemination activities surrounding the CER translation process. This challenge will confront any retrospective examination of CER dissemination and will be true of many potential discussants.

While we covered a broad range of topics, we did not include a case study on diagnostic imaging, which might have added value given the role of diagnostic imaging in determining subsequent utilization of medical care. We also might have considered CER studies that compared behavioral interventions (e.g., dietary therapy for patients with diabetes); however, we believed that behavioral interventions would involve too many unique factors that would limit our ability to draw conclusions over multiple case studies. Finally, we might have expanded our scope to specifically include one or more “success stories,” to better draw contrasts between factors that impeded and facilitated translation. We ultimately chose not to use this criterion to select topics because of the difficulty of identifying a sufficiently large sample of unambiguous success stories. Instead, we prioritized topics in a way that ensured diversity of types of treatment comparisons.

Case-Study Methodology

Because of the formative goals of our study, we elected not to use a formal qualitative research methodology that included the coding of themes with the use of specialized analytic software. Biased interpretations of the data by the research team were mitigated by requiring a minimum of three investigators to be present for each discussion. In addition, we held a debriefing immediately following each discussion to identify key themes, and we drafted case-study-specific summaries to refine the presentation of themes and to arrive at a consensus interpretation. Because we did not record each discussion, our analysis was based on examination of the transcripts generated by the dedicated note takers. Although we were not always able to derive exact quotes from discussants, we used email follow-up for areas that needed clarification. While we were unable to verify the veracity of statements or double-check the accuracy of quantitative estimates, our root-cause analysis drew primarily on themes that were mentioned repeatedly by stakeholders. Future studies might use more-structured interviews to capture similar types of information from a greater number of discussants.

Because our case studies dated from as early as 1998, stakeholders’ recollections of specific events may have been subject to recall bias. We believe this bias, if present, was likely to be mitigated by the fact that most stakeholders were intimately involved with the relevant studies. Moreover, the crosscutting discussions we held with certain stakeholders, including payers and representatives of integrated

delivery systems, did not require detailed knowledge of each case study. We might have used greater numbers of *group* discussions to guard against recall bias introduced by any individual stakeholder or to ensure that each discussion more accurately reflected the perspective of the larger stakeholder group. It was logistically challenging to schedule group discussions, so we prioritized individual discussions with senior managers or key opinion leaders who we assumed would be most informed about a broad array of issues.

The environmental scans we conducted for each case-study topic were helpful in structuring discussions. However, we were unable to integrate findings from the grey literature, which contains highly redundant information (e.g., similar press releases were picked up by dozens of trade publications) and has few identifiable authoritative sources. Because we were thus unable to conduct an efficient search of the grey literature, we may have missed an opportunity to gain greater insight into additional perspectives—particularly those of insurers and device and pharmaceutical manufacturers.

Generalizability

Because of the limited scope of topics we could address and the limited number of discussions we were able to hold, some of our findings regarding the root causes of failed CER translation or the facilitators of practice change may not be generalizable to other topics or to a broader range of practice settings. For example, small health plans may face a different set of barriers to the use of CER to guide policymaking than those faced by payers that operate nationally. Community physicians may have different interpretations of the meaning of these studies and may also face a different set of implementation challenges. Because of the extensive heterogeneity within each stakeholder group, our findings may be significantly more nuanced than we have presented them to be.

The CPOE case study, in particular, may not provide results that generalize to the broader set of delivery-system interventions. CPOE is a heterogeneous intervention that can be simple or complex. It may be integrated into EHRs or stand alone, and its successful adoption is much more highly context-dependent than are the subjects of our other case studies. Our findings from the CPOE case study may be more generalizable to informatics topics than to the broader set of delivery-system interventions, which have substantial heterogeneity. As identified by the IOM, this is a high priority for future CER, and we emphasize that while our findings concerning adoption of CPOE were relatively negative, CER may have considerable potential to bring about changes in practice for the class more broadly.

Policy Implications of the Case-Study Results

The federal investment in CER has many objectives. In contrast to the traditional research agenda defined by a peer-review process in which research priorities are based primarily on the novelty and scientific interest of the topic, the CER agenda is supposed to be set by a broader set of stakeholders. Providing better evidence to support the decisions of clinicians and patients, thereby making healthcare delivery more effective and efficient, is the overarching goal envisioned by the architects of CER.

Translation of research into practice is not a new concern. Since the dawn of scientific medicine, physicians have noted the difficulty of changing established patterns of practice on the basis of scientific evidence. In at least the past two decades, many efforts, including evidence-based medicine, quality measurement, quality improvement, and electronic clinical decision support, to name a few, have focused on improving the manner in which healthcare is delivered, overcoming what Eisenberg referred to as “voltage drops” along the pathway from scientific knowledge to clinical practice (Eisenberg and Power, 2000). More recently, implementation science has come into its own as a field of study.

The case studies in this report led us to conclude that the translation of CER research into practice comprises five phases: generation, interpretation, formalization, dissemination, and implementation. Under some circumstances, CER evidence can move through these five phases relatively quickly, but in many cases, the evidence languishes or is never implemented at all. Our case studies suggest that the pathway through the five phases is arduous and that there is a high risk that CER evidence will fail to influence practice.

Chapter Seven summarized our analysis of the root causes of the failure to translate CER evidence into clinical practice. While these are formidable, they are not insurmountable. In this chapter, we reflect on the policy implications of the case studies and identify policy options that can address the root causes of failure and promote more effective translation of CER evidence into local practice settings. We begin with a description of the infrastructure that would enable effective translation, and we then suggest policy options that could be deployed in the following areas: governance, standards, financing, professionalism, marketing and education, and research and evaluation. The list of options is not intended to be an exhaustive or detailed policy prescription. Rather, we believe that our case studies have identified some of the high-leverage opportunities that policymakers, including members of PCORI, may find useful as they chart a course forward.

A CER-Enabling Infrastructure

Maximizing the impact of CER will require the effective delivery of evidence to numerous audiences. For the most part, the five phases of CER translation are not actively managed or supported in a coherent way. Instead, evidence is published, then a variety of stakeholders take actions that serve their interests and motives. The dynamic interplay among these stakeholders can be unpredictable and sometimes haphazard, producing suboptimal implementation of new clinical practices and potentially contributing to poor quality and high costs of care. While larger payers and integrated systems have developed the infrastructure necessary for translation (including committees to appraise and interpret CER evidence, guideline-development processes, and strategies to develop and implement performance measures), they do so at a cost that is prohibitive to most health plans and providers. It is unlikely (and probably undesirable) that any group in the United States will manage the CER translation process centrally. However, a CER-enabling infrastructure could remove some of the perverse incentives of the current approach and reduce the number of missed opportunities to implement new and effective clinical practices.

What is a CER-enabling infrastructure? We envision it as a set of policymaking bodies, policies, and funded activities that achieve three aims: (1) enabling generation of CER that is highly relevant to decisionmakers, (2) enabling more-effective translation of CER results into practice, and (3) enabling more-effective evaluation of the impact of translation activities. This infrastructure is already a work in progress embodied in PCORI and federal agencies that play a role in CER. It exists in a wide array of federal and state agencies, regulators, professional societies, academic institutions, and healthcare-delivery organizations, but it is not sufficiently organized to optimize the implementation of new clinical practices. Similarly, the enabling infrastructure for *evaluation* of the impact of CER translation activities exists in federal agencies, academic institutions, and research organizations, but it has failed to produce robust, generalizable evidence on best practices to facilitate translation of CER into practice.

The policy options outlined below are not intended to create a command-and-control infrastructure. Changes in the translation process, such as reengineering financial incentives, must be carried out by a diverse set of stakeholders in both the public and private sectors, and these changes must address a remarkable diversity of payment arrangements. These are policy options that could bring greater coherence and transparency to the process of CER translation, achieve greater balance of the influence of participating stakeholders, and enhance the voice of the public and patients whose health outcomes depend on effective, safe, and affordable care. Enacting some or all of these policy options could be expected to modify the financial and other incentives that shape clinical decisionmaking over time, so that decisions could be increasingly based on evidence.

Policy Options

Governance

Create a transparent governance mechanism for oversight of the CER translation process. Financial incentives play a major role in many or all of the phases of CER translation into practice. Such incentives are unavoidable given the structure of the U.S. economy, and financial conflicts of interest may also be

unavoidable, but the consequences of those conflicts of interest are not inevitable. While much of the funding for CER comes from public sources, much of the support for translation activities comes from private sources. Transitioning all CER translation activities to public funding is unrealistic, although some of these activities, including the development of guidelines, quality measures, and the nation's health IT infrastructure, have been publicly funded to a degree.

To avoid the potential for financial incentives to derail CER translation, especially when new practices challenge vested interests, a governance mechanism that monitors the translation process and documents activities could maximize the incorporation of CER results into clinical practice in a timely way. This governance mechanism could take many forms but would most likely serve as an advisory committee. Its stakeholder membership and its relationship to federal agencies, PCORI, and other organizations would have to be determined. For instance, it might work closely with PCORI and be modeled along the lines of the PCORI methodology committee. Going one step further, this entity could develop specific policy recommendations that would address the root causes of failure of CER translation. Given the diversity of involved stakeholders, it seems unlikely that this body would have the authority to enact policies, but it could play a convening role, bringing together key stakeholders who wield the authority to enact policies and document their activities. By representing key stakeholder groups and ensuring a transparent consensus-development process, this body could have substantial influence.

The case studies suggest some especially important opportunities for this governance structure that may help to mitigate financial conflicts of interest. The following opportunities would increase transparency.

1. *Include patient or consumer representatives.* As the recipients of healthcare, patients are the key beneficiaries of the clinical practices guided by CER results. Ultimately, members of the public pay for healthcare, as well as the investment in CER. Patient and consumer representatives should be involved early in decisionmaking concerning CER (e.g., the questions to be answered by CER, the types of patients to be included in CER studies, and appropriate outcome measures), and they should also be involved in guiding the other phases of translation, including interpretation of results, formalization, dissemination, and implementation. Such engagement may require a public investment, and patient and consumer representatives may require training to enable their full and active participation in governance committees. The goal of such engagement would be to ensure that the CER translation process is well informed by the end-user perspective and to provide a counterbalance to other stakeholder interests—particularly those of stakeholders with opportunities for financial gain.
2. *Include public and private payer and purchaser representatives.* As stewards of the financing of healthcare and representatives of their member or customer interests, payers and purchasers have a strong interest in the translation of CER evidence into high-value clinical practices. Payers tend to focus on specific clinical topics and face common challenges in

- using CER evidence. Because they ultimately have a central role in formalization, dissemination, and implementation of policies based on CER results, it is crucial to include their perspectives in the governance process for CER translation. Inclusion of public payers is especially important, as they serve distinct populations with higher health risks and vulnerability. The interests of public and private payers often differ in the current U.S. healthcare system. Payers and purchasers may be able to identify the specific opportunities for high-value care that would be especially amenable to CER and for which modification of payment or coverage policies could be especially effective in optimizing clinical practice.
3. *Enable and support public-comment opportunities.* Vigorous solicitation of public comment with verification and full disclosure of potential conflicts of interest (both financial and intellectual) can enhance the credibility of CER. An entity and mechanisms to oversee compliance with disclosure policies would need to be established. Use of electronic media to enable structured rating and voting exercises (with appropriate safeguards to prevent manipulation) could create a transparent system to aggregate input from a broad range of stakeholders on CER priorities and concerns about the use of this information. These data might also help in arriving at consensus interpretations of CER evidence and its implications for practice.
 4. *Institute strong policies on disclosure and management of potential conflicts of interest.* A perception of CER evidence as biased by financial interests can undermine its credibility and impede the take-up of new practices. Conflicts of interest cannot be eliminated, but the IOM experience suggests that they can be disclosed and managed effectively. The governance of the financing of CER projects will undoubtedly be subject to conflict-of-interest policies. Our case studies suggest that the governance and oversight of all the phases of the CER translation process could benefit from such policies, which might be patterned on the recent IOM recommendations for achieving trustworthy guidelines. While most current approaches focus primarily on financial conflicts of interest, disclosure could also include potential institutional and intellectual conflicts of interest. The optimal strategy for monitoring and enforcing disclosure could be specified by a body such as PCORI.
 5. *Use the governance mechanism to generate a prospective public record of stakeholder expectations.* Documenting the positions of relevant stakeholders at the outset of CER studies with respect to the objectives of the study and the parameters around which the results should be interpreted would create a public record of expectations of each stakeholder in the CER process. While it would be non-binding, public documentation can create both a record of expectations that can be drawn upon during the interpretation phase and an institutional memory in organizations that have relatively high rates of executive turnover. Such a record could discourage post-hoc efforts to undermine the credibility of studies that produce results contrary to the interests of specific stakeholders.

Standards

Support and enhance creation of standards for CER generation and translation. Standards can enhance transparency and produce consistency of interpretation. Efforts to create standards for the design and conduct of CER are already in progress through activities of federal agencies and PCORI. However,

standardization, including guidelines and quality measurement, can be helpful at every phase of the CER translation process. At least two other opportunities for standardization, described below, hold promise.

1. *Incorporate data elements that are critical to translation activities into the CER registry.* A registry of federally funded CER studies is currently being developed by ASPE. Similar to the government-sponsored clinical-trials registry, the CER registry will create a common format for the presentation of key characteristics and results of each study that will enable the public to easily retrieve information about proposed, in-progress, and completed CER. This registry will include a variety of important facts about CER studies that could assist in consistency of interpretation of results:
 - Patterns of use of interventions relevant to the CER studies at the time they are initiated
 - Anticipatory commentary, including prioritized lists of the methodological strengths and weaknesses of proposed CER studies
 - Inclusion and exclusion criteria for patients in a CER study
 - Prespecified definitions of equivalence and nonequivalence of treatments
 - Prespecified thresholds for clinical action based on CER results (e.g., effectiveness, cost, and safety) and the specific actions that would be recommended under different result scenarios.

Creating explicit standards for CER study objectives, design, sampling, and causes of heterogeneity in sampled populations may be useful to guide the forming of a consensus interpretation of CER study results. Predefining thresholds for action may reduce post-hoc reframing of questions and implications by stakeholders to serve their own interests.

2. *Encourage development of standardized electronic clinical data systems (clinical registries).* CER studies can be expensive and time-consuming. Standardized electronic clinical registries are increasingly providing opportunities to conduct low-cost, rapid CER. Apart from generating new research, registries can provide data for longitudinal tracking systems to evaluate the impact of CER translation activities on clinical-practice patterns. Engaging clinicians and patients in the use of clinical-data registries may also increase the credibility of CER results by enhancing the perceived trustworthiness of the data for end users, overcoming some of the traditional skepticism about administrative data and claims data generated for billing purposes. A common finding from our case studies was that feedback to individual physicians on utilization patterns was often a potent motivator to change practice—particularly when the data were of high quality.

Registries can support and enhance the creation of standards, because the design decisions necessary to build a clinical registry (e.g., specification of clinical concepts, data elements, measurement approaches) usually lead directly to standardization. Furthermore, generating consensus on design decisions frequently raises questions about priorities and methods, and it identifies additional opportunities for CER. Development of registries and provider participation

in them may be stimulated through the CMS EHR Incentive Program and the clinical-data needs of ACOs. However, additional incentives may be required to advance these efforts.

Financing

Encourage public financing of CER translation and promote the use of CER evidence in payment programs. The case studies suggest two policy options related to CER financing.

1. *Provide direct and indirect public support for formalization of CER evidence and the dissemination of CER-based clinical practices.* In line with the emphasis on public, patient, and consumer representation in the governance of CER, public financing could be provided for translation of high-profile CER evidence into guidelines, quality measures, and clinical decision support tools. Forcing guideline and measure development to rely upon industry funding tends to skew the formalization process toward generation of CER with profit-making potential for industry rather than serving the public's interest in promoting high-value healthcare. Public financing need not replace the selective and targeted funding of such activities by industry or professional interests but would play a complementary role.

Likewise, public financing for dissemination activities and quality improvement initiatives that are based on high-quality evidence (e.g., counterdetailing and other efforts) could supplement the substantial private support for activities. This public financing could focus on counteracting marketing approaches of industry interests that might otherwise undermine CER evidence or selectively promote less-effective or more-costly practices. The public activities need not be broad-based but might use empirical analysis to identify utilization patterns that are at odds with the CER evidence to enable targeted counterdetailing programs.

2. *Promote the use of CER-based clinical practices through payment policy and incentive programs directed toward providers and patients (e.g., value-based purchasing).* New payment models could reduce perverse fee-for-service incentives that may drive the use of procedures or services that have weak evidence of effectiveness. Through reliance on guidelines, quality measures, and clinical decision support, these models can incorporate CER evidence. Public payers have a special role given their accountability to the public. Encouraging payers such as Medicaid and Medicare to use prudent purchasing strategies based on CER evidence can assure that the financial incentives they (and patients) face are well aligned to support evidence-based clinical practices and discourage practices that are not evidence-based. Value-based insurance design, an approach that may be gaining favor among insurers, could begin simply with clinical areas in which the CER evidence is strongest and could then evolve into other areas.

Professionalism

Supporting professional consensus across the phases of CER translation. Two policy options could encourage development of consensus.

1. *Foster and support a broad vision of professionalism in the governance of CER translation.* With the increasing subspecialization of medicine and surgery, subspecialty professional societies have proliferated. Conflicting professional-society guidelines are one manifestation of this trend. Differences in the interpretation, formalization, and dissemination of CER evidence and lobbying by subspecialty societies may impede effective changes in clinical practice.

Broadly constituted professional committees may be capable of producing balanced, consensus interpretations of CER results, resolving differences of opinion about their interpretation in a transparent manner. To date, organizations involved with guideline and performance-measure development have only slowly begun to diversify their expert panels to accommodate a more diverse mix of specialties; some have resisted this process. To achieve a broad vision of professionalism, steering committees and advisory committees could include generalist representatives in addition to subspecialists, and the balance between the numbers of generalist and subspecialist professionals could be set so that no single perspective could dominate the policymaking of these groups. Assuring disclosure of conflicts of interest and managing these conflicts effectively could neutralize external influences.

Other steps might reverse the tendency for conflict among professionals. As noted above, encouraging the participation of professionals and patients in clinical registries might enhance professional and public trust in CER results produced from such registries. Engaging subspecialty societies in the construction of shared clinical registries and health-information exchange might counteract the tendency to focus on narrowly defined subspecialty interests.

2. *Provide training for professionals on the role of cognitive biases in diagnostic and treatment decisionmaking.* As described in Chapter Seven, cognitive biases may contribute to the failure of CER to affect clinical practice. Cognitive biases may be difficult or impossible to change, however, since they are “hard-wired” to some extent. However, it is possible to compensate for some cognitive biases by altering the decisionmaking context. For example, the framing of the description of expected gains and losses related to an intervention can affect the likelihood that the intervention will be recommended and accepted. If CER evidence informs this framing, implementation of a CER-based clinical practice may be more likely to succeed.

Enabling professionals to recognize the circumstances under which decisions and clinical recommendations are prone to cognitive biases may reduce the influence of factors other than CER evidence. Helping professionals identify these circumstances could increase the alignment of clinical recommendations and decisions with CER evidence. Training in the nature, role, and impact of cognitive biases could be incorporated into the curricula of professional schools and clinical training programs. Such curricula might also include education on specific methods for modifying the decisionmaking context to compensate for cognitive biases.

Education and Marketing

Promote demand for CER-based clinical services through public education and marketing. Marketing campaigns, including detailing to clinicians and direct-to-consumer advertising, are generally aimed at exploiting cognitive biases, and this may impede the take-up of CER-based clinical services. Recent initiatives by some states and organizations have been undertaken to reduce the influence of detailing of physicians. Regulation of the claims permitted in direct-to-consumer advertising may also play an important role in mitigating bias in decisionmaking. Two options are suggested.

1. *Promote patient demand for CER-based clinical services through shared decisionmaking using formal patient decision aids.* Stimulating patient demand can be a powerful adjunct to other approaches for increasing the use of CER-based clinical practices. CER results and CER-based clinical practices are disseminated through a variety of channels, including clinicians' recommendations, mass-media publications and broadcasts, direct-to-consumer advertising, and the Internet. The challenge for patients (and for clinicians) is to separate recommendations grounded in CER from non-evidence-based claims.

Shared decisionmaking involving the use of formal decision aids could ensure that patients receive the best available evidence about alternative tests and treatments in a usable form, but decision aids have not been incorporated routinely into clinical practice. Many demonstration projects and studies are under way to identify the optimal methods and for timing the delivery of high-quality information to patients. Federal and state legislation is beginning to encourage the use of patient decision aids through quality-measurement and payment-incentive programs. Efforts to assure that CER evidence informs decision aids and that decision aids are used routinely in practice could provide the needed impetus to the uptake of CER-based clinical practices. Decision aids could also counter the messages that promote suboptimal clinical practices.

2. *Support "social marketing" for high-profile CER results to counteract the effects of industry-sponsored detailing and direct-to-consumer advertising.* Patient demand for services can undermine the implementation of CER-based clinical practices. Social marketing—the application of marketing techniques to promote behavioral change—has the potential to increase awareness of and demand for evidence-based healthcare services. Social marketing campaigns following the release of new CER findings could stimulate patients to engage their providers in discussions about the meaning of the results for their own care, leading to the use of treatments that may be better aligned with the CER evidence. In particular, social marketing may counteract the unintended consequences of direct-to-consumer marketing and complement academic detailing programs that may be costly in certain practice contexts.

Social media, such as Facebook and Twitter, may be effective tools for enhancing social marketing. In the same way that they have been used increasingly to enhance recruitment of

volunteers for clinical trials, social media could play an important role in facilitating translation of CER into practice.

Research and Evaluation

Further research on each of the phases of translation and the interactions of stakeholders may help to optimize the CER research portfolio and its translation into effective clinical practice. Our case studies were retrospective and limited in number, so our conclusions are necessarily preliminary. Prospective evaluation of the effectiveness of CER results in changing clinical practice could help to refine the strategies used to interpret, formalize, disseminate, and implement CER-based clinical practices. The potential research and evaluation agenda is vast, and outlining it in detail is beyond the scope of this report. However, our results suggest a set of potentially high-priority areas for investment.

1. Support of research to identify the gaps in clinical decisionmaking that are the highest-priority topics for end users of CER. Understanding the needs of potential end users of CER could be an important goal of a research portfolio. Information about the gaps in clinical decisionmaking that could be topics of CER could be gathered in a variety of ways. Stakeholders involved in governance could be one source of input, but they may not provide a complete picture of the high-leverage opportunities for CER.

Other approaches include studying the potential and actual responses of key stakeholders (such as professionals, payers, and the public) to plausible CER study outcomes to obtain data on the factors that interact to yield successful changes in practice. Using these factors to model and anticipate the effects of stakeholder responses to CER results in each of the phases of CER translation could help to identify topics that are feasible targets for change. In addition, research on end-user needs could foster the selection of approaches for dissemination of CER results tailored to the expectations of these stakeholders.

Encouraging the design of CER studies that examine key clinical diagnostic decision points that are upstream from the use of drugs or procedures may also increase the relevance of CER to payers, professionals, and the public.

2. Promoting integration of the CER registry with clinical registries to support evaluation of the impact of CER studies and the factors associated with successful translation. Prospective evaluation of the impact of CER can be strengthened by reliable estimates of current clinical-practice patterns. Our results suggest that practice patterns, particularly rates of use of alternative treatments for specific indications and variations in those rates, may be incompletely understood. Without credible baseline estimates, it may be difficult to study the changes in clinical practice that follow publication of CER results.

Furthermore, clinical-decision algorithms often include contingencies such that the downstream treatments are promoted or impeded by decisions that occur earlier in the diagnostic algorithms that clinicians use. CER results pertinent to these downstream treatment choices can be helpful but may require accompanying research about diagnostic choices that occur upstream. Our COURAGE case study suggested that future CER focusing on the role of stress-testing within clinical decision algorithms may help to better tailor the use of downstream treatments such as angiography and PCI. Clinical registries

that can provide longitudinal data on patients may enable more-complex studies and increase the relevance of results for end users.

Finally, the linking of the CER registry with clinical registries that enable the study of impact could assist policymakers in refining knowledge of the factors associated with successful translation of CER results from different types of CER studies.

3. Support of projects that develop unbiased and efficient methods for formalization of CER results. At present, guidelines, performance measures, and clinical decision support tools are the primary approaches to formalizing CER evidence into clinical-practice recommendations and tools that can be used to disseminate CER results and CER-based practices. Methods for developing and refining these tools and assuring that they are unbiased are still a work in progress. Support for research and demonstration projects that develop and study new methods for formalization could lead to more effective, efficient, and unbiased tools.

Research on formalization might focus on methods for assuring transparency through the use of formal panel protocols, criteria development, explicit rating algorithms, and statistical methods for weighting the priority of evidence and the tools that are produced. Development of new HIT applications that could accelerate the use of these methods and reduce the cost of producing such tools could be a key priority.

4. Support of projects that enhance the utility of CER results by demonstrating and evaluating models for the use of decision aids by clinicians and patients. Effective clinical practice is built on effective clinical decisionmaking. Shared decisionmaking is inherently difficult to implement because of incomplete information and information asymmetries between professionals and patients. One of the major contributions of CER is the provision of information professionals and patients can use to arrive at optimal diagnostic and treatment decisions. But financial incentives, marketing, and prior beliefs complicate the task of introducing CER evidence.

Our results suggest that designing more-effective decision aids, training professionals to use them, and strategies for embedding them in routine practice have all proven challenging. Research and evaluation projects that lead to better decision aids and more effective use of them could increase the demand for and impact of CER.

5. Support of research into the ways in which CER evidence is used by integrated delivery systems. While a small number of experts from integrated delivery systems participated in our discussions, a more systematic assessment of the ways in which these organizations use CER evidence merits further study. Because integrated systems routinely evaluate evidence with the explicit goal of developing clinical-guidance tools, they may have unique perspectives on which CER topics are likely to have the greatest impact on clinical practice. In addition, lessons learned by integrated systems involving the translation of CER evidence into practice could provide insight into the development of policies that are useful outside of these systems. For example, we learned that integrated systems are more likely to have their own guideline-development processes, sophisticated HIT systems, and greater levels of patient trust—all of which are conducive to CER translation. Future studies might engage these organizations to elicit best

practices in CER translation and evaluate which strategies may be transferrable to nonintegrated delivery settings.

Concluding Remarks

The United States is making a sizable investment in CER in the hope that the results will influence the decisions of clinicians and patients and optimize the quality and costs of care. Attention to the root causes of ineffective CER translation seems critical. The policy options that emerged from our case studies of five prominent CER projects appear to us to have the greatest potential to address key factors that may lead to failure of CER to influence clinical practice. These results underscore the complexity of introducing CER evidence. Despite that complexity, it appears that a relatively finite number of policy changes, carefully selected, could enhance the impact of CER in the future. Taken together, these policy options suggest a number of different paths forward, and exploring multiple approaches may be appropriate given the large number of factors that impede CER translation. As the ARRA-financed CER portfolio begins to produce new evidence, a number of opportunities will exist to experiment with these approaches.

References

- Abraham, W. T., W. G. Fisher, et al. (2002). "Cardiac resynchronization in chronic heart failure." *N Engl J Med* 346(24): 1845–1853.
- Al-Khatib, S. M., A. Hellkamp, et al. (2011). "Non-evidence-based ICD implantations in the United States." *JAMA* 305(1): 43–49.
- Albert, N. M., G. C. Fonarow, et al. (2010). "Influence of dedicated heart failure clinics on delivery of recommended therapies in outpatient cardiology practices: Findings from the Registry to Improve the Use of Evidence-Based Heart Failure Therapies in the Outpatient Setting (IMPROVE HF)." *Am Heart J* 159(2): 238–244.
- Arnold, J. M. O., and L. J. Gula (2010). "The art and science of heart failure: Predicting the unpredictable." *J Am Coll Cardiol* 55(17): 1811–1813.
- Ash, J. S., P. Z. Stavri, et al. (2003). "A consensus statement on considerations for a successful CPOE implementation." *J Am Med Inform Assoc* 10(3): 229–234.
- Atlas, S. J., R. A. Deyo, et al. (1996). "The Maine Lumbar Spine Study, Part III. 1-year outcomes of surgical and nonsurgical management of lumbar spinal stenosis." *Spine (Phila Pa 1976)* 21(15): 1787–1794; discussion 1794–1785.
- Atlas, S. J., R. B. Keller, et al. (2000). "Surgical and nonsurgical management of lumbar spinal stenosis: Four-year outcomes from the Maine Lumbar Spine Study." *Spine (Phila Pa 1976)* 25(5): 556–562.
- Aversano, T., L. T. Aversano, E. Passamani, et al. (2002). "Thrombolytic therapy vs primary percutaneous coronary intervention for myocardial infarction in patients presenting to hospitals without on-site cardiac surgery: a randomized controlled trial." *JAMA* 287(15): 1943–1951.
- Bangalore, S., and F. H. Messerli (2007). "Is there an ischemic threshold beyond which percutaneous coronary intervention is beneficial in the Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation (COURAGE) trial?" *Am J Cardiol* 100(9): 1495.
- Bates, D. W., et al. (1998). "Effect of computerized physician order entry and a team intervention on prevention of serious medical errors." *JAMA* 280: 1311–1316.
- Bhatt, D. L. (2007). "Interpreting the COURAGE trial: Is medical therapy as good as PCI stable angina? Two views." *Cleve Clin J Med* 74(9): 618ff.
- Birkmeyer, C. M., J. Lee, et al. (2002). "Will electronic order entry reduce health care costs?" *Eff Clin Pract* 5(2): 67–74.

- Boden, W. E. (2007). "Interpreting the COURAGE trial: It takes COURAGE to alter our belief system." *Cleve Clin J Med* 74(9): 623–625, 629–633.
- Boden, W. E. (2008). "Management of chronic coronary disease: Is the pendulum returning to equipoise?" *Am J Cardiol* 101(10A): 69D–74D.
- Boden, W. E., R. A. O'Rourke, et al. (2007). "The evolving pattern of symptomatic coronary artery disease in the United States and Canada: Baseline characteristics of the Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation (COURAGE) trial." *Am J Cardiol* 99(2): 208–212.
- Boden, W. E., R. A. O'Rourke, et al. (2007). "Optimal medical therapy with or without PCI for stable coronary disease." *N Engl J Med* 356(15): 1503–1516.
- Borden, W. B., R. F. Redberg, et al. (2011). "Patterns and intensity of medical therapy in patients undergoing percutaneous coronary intervention." *JAMA* 305(18): 1882–1889.
- Bristow, M., L. A. Saxon, J. Boehmer, S. Krueger, D. A. Kass, T. De Marco, P. Carson, L. DiCarlo, D. DeMets, B. G. White, D. W. DeVries, and A. M. Feldman (2004). "Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure." *N Engl J Med* 350(21): 2140–2150.
- Buchanan, R. W., J. Kreyenbuhl, et al. (2010). "The 2009 schizophrenia PORT psychopharmacological treatment recommendations and summary statements." *Schizophr Bull* 36(1): 71–93.
- Budde, T. (2006). "AICD treatment in 2004—State of the art." *Eur J Med Res* 11(10): 432–438.
- Burkhardt, J. D., and B. L. Wilkoff (2007). "Interventional electrophysiology and cardiac resynchronization therapy delivering electrical therapies for heart failure." *Circulation* 115(16): 2208–2220.
- Burling, S. (2001). Philadelphia-area hospitals launch effort to reduce drug errors. *The Philadelphia Inquirer*.
- Calverley, P., R. Pauwels, J. Vestbo, et al. (2003). "Combined salmeterol and fluticasone in the treatment of chronic obstructive pulmonary disease: a randomised controlled trial." *Lancet* 361(9356): 449–456.
- Carpenter, W. T., and R. W. Buchanan (2008). "Lessons to take home from CATIE." *Psychiat Serv* 59(5): 523–525.
- Carreyrou, J., and T. McGinty (2011). Hospital bars surgeon from operating room: Medical board in Oregon separately investigates doctor who stood out for high rate of multiple spinal procedures. *Wall Street Journal*.
- Cheng, A., Y. F. Wang, et al. (2010). "Acute lead dislodgements and in-hospital mortality in patients enrolled in the National Cardiovascular Data Registry Implantable Cardioverter Defibrillator Registry." *J Am Coll Cardiol* 56(20): 1651–1656.
- Chou, R., and M. Helfand (2005). "Challenges in systematic reviews that assess treatment harms." *Ann Intern Med* 142(12 Pt 2): 1090–1099.
- Covell, N. H., M. T. Finnerty, et al. (2008). "Implications of CATIE for mental health services researchers." *Psychiat Serv* 59(5): 526–529.

Curtis, J. P., J. J. Luebbert, et al. (2009). "Association of physician certification and outcomes among patients receiving an implantable cardioverter-defibrillator." *JAMA* 301(16): 1661–1670.

Dettling, M., and I.-G. Angheliescu (2006). "Effectiveness of antipsychotic drugs in patients with chronic schizophrenia: Comment." *New Engl J Med* 354(3).

Deyo, R. A., D. T. Gray, et al. (2005). "United States trends in lumbar fusion surgery for degenerative conditions." *Spine (Phila Pa 1976)* 30(12): 1441–1445; discussion 1446–1447.

Deyo, R. A., S. K. Mirza, et al. (2010). "Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults." *JAMA* 303(13): 1259–1265.

Deyo, R. A., A. Nachemson, et al (2004). "Spinal-fusion surgery: The case for restraint." *N Engl J Med* 350(7): 722–726.

Diamond, G. A., and S. Kaul (2007). "COURAGE under fire: On the management of stable coronary disease." *J Am Coll Cardiol* 50(16): 1604–1609.

Dixon, L., D. Perkins, et al. (2009). *Guideline Watch (September 2009): Practice Guideline for the Treatment of Patients with Schizophrenia*. A. P. Association.

Doolan, D. F., and D. W. Bates (2002). "Computerized physician order entry systems in hospitals: Mandates and incentives." *Health Aff (Millwood)* 21(4): 180–188.

Duckworth, K., and M. J. Fitzpatrick (2008). "NAMI perspective on CATIE: Policy and research implications." *Psychiat Serv* 59(5): 537–539.

Einstein, A. J., M. J. Henzlova, and S. Rajagopalan. "Estimating risk of cancer associated with radiation exposure from 64-slice computed tomography coronary angiography." *JAMA* 298(3): 317–323.

Eisenberg, J. M., and E. J. Power (2000). "Transforming insurance coverage into quality health care: Voltage drops from potential to delivered quality." *JAMA* 284(16): 2100–2107.

Farmer, S. A., J. N. Kirkpatrick, et al. (2009). "Ethnic and racial disparities in cardiac resynchronization therapy." *Heart Rhythm* 6(3): 325–331.

Federal Coordinating Council for Comparative Effectiveness Research (2009). *Report to the President and the Congress*. U.S. Department of Health and Human Services. Washington, DC.

Fein, A. S., Y. F. Wang, et al. (2010). "Prevalence and predictors of off-label use of cardiac resynchronization therapy in patients enrolled in the National Cardiovascular Data Registry Implantable Cardiac Defibrillator Registry." *J Am Coll Cardiol* 56(10): 766–773.

Feldman A. M., L. G. de Lissovoy, M. R. Bristow, et al. (2005). "Cost effectiveness of cardiac resynchronization therapy in the Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure (COMPANION) trial." *J Am Coll Cardiol* 46(12): 2311–2321.

Finley, D. (2009). Hospitals slow to adopt technology that averts errors. Dead by mistake: A special report. *San Francisco Chronicle*.

Fonarow, G. C., N. M. Albert, et al. (2010). "Improving evidence-based care for heart failure in outpatient cardiology practices: Primary results of the registry to Improve the Use of Evidence-Based Heart Failure Therapies in the Outpatient Setting (IMPROVE HF)." *Circulation* 122(6): 585–U114.

Fonarow, G. C., C. W. Yancy, et al. (2008). "Heart failure care in the outpatient cardiology practice setting: Findings from IMPROVE HF." *Circ Heart Fail* 1(2): 98–106.

Fonarow, G. C., C. W. Yancy, et al. (2007). "Improving the use of evidence-based heart failure therapies in the outpatient setting: The IMPROVE HF performance improvement registry." *Am Heart J* 154(1): 12–38.

Freidson, E. (1970). *Profession of Medicine: A Study of the Sociology of Applied Knowledge*. New York, NY, Harper and Row.

Friedewald, V. E., J. P. Boehmer, et al. (2007). "The editor's roundtable: Cardiac resynchronization therapy." *Am J Cardiol* 100(7): 1145–1152.

Frye, R. L., P. August, et al. (2009). "A randomized trial of therapies for type 2 diabetes and coronary artery disease." *N Engl J Med* 360(24): 2503–2515.

Gibson, J. N., and G. Waddell (2005). "Surgery for degenerative lumbar spondylosis: Updated Cochrane Review." *Spine (Phila Pa 1976)* 30(20): 2312–2320.

Gingrich, N. (2000). High-tech cure for medical mistakes. *The Washington Post*.

Greenhalgh, T., G. Robert, et al. (2004). "Diffusion of innovations in service organizations: Systematic review and recommendations." *Milbank Q* 82(4): 581–629.

Haig, A. J., and C. C. Tomkins (2010). "Diagnosis and management of lumbar spinal stenosis." *JAMA* 303(1): 71–72.

Harris, G. (2010). More doctors giving up private practices. *New York Times*.

Hiddemann, W., M. Kneba, M. Dreyling, et al. (2005). "Frontline therapy with rituximab added to the combination of cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP) significantly improves the outcome for patients with advanced-stage follicular lymphoma compared with therapy with CHOP alone: Results of a prospective randomized study of the German Low-Grade Lymphoma Study Group." *Blood* 106(12): 3725–3732.

Higgins, S. L., J. D. Hummel, et al. (2003). "Cardiac resynchronization therapy for the treatment of heart failure in patients with intraventricular conduction delay and malignant ventricular tachyarrhythmias." *J Am Coll Cardiol* 42(8): 1454–1459.

Hochman, J. S., G. A. Lamas, et al. (2006). "Coronary intervention for persistent occlusion after myocardial infarction." *N Engl J Med* 355(23): 2395–2407.

Institute of Medicine (2011). *Clinical Practice Guidelines We Can Trust*. Washington, DC.

Johnson, L. (2001). Technology keeps prescriptions accurate; electronic devices, computers used to ensure medicine dispensed correctly. *South Bend Tribune*.

Kashani, A., and S. S. Barold (2005). "Significance of QRS complex duration in patients with heart failure." *J Am Coll Cardiol* 46(12): 2183–2192.

Katritsis, D. G., and J. P. Ioannidis (2005). "Percutaneous coronary intervention versus conservative therapy in nonacute coronary artery disease: A meta-analysis." *Circulation* 111(22): 2906–2912.

Kereiakes, D. J., P. S. Teirstein, et al. (2007). "The truth and consequences of the COURAGE trial." *J Am Coll Cardiol* 50(16): 1598–1603.

Kirtane, A. J., and D. J. Cohen (2008). "When is better not good enough? Insights from the COURAGE economic study." *Circ Cardiovasc Qual Outcomes* 1(1): 4–6.

Kocher, R., and N. R. Sahni (2011). "Hospitals' race to employ physicians: The logic behind a money-losing proposition." *N Engl J Med* 364(19): 1790–1793.

Koppel, R., J. P. Metlay, et al. (2005). "Role of computerized physician order entry systems in facilitating medication errors." *JAMA* 293(10): 1197–1203.

Kovacs, M. J., M. Rodger, D. R. Anderson, et al. (2003). "Comparison of 10-mg and 5-mg warfarin initiation nomograms together with low-molecular-weight heparin for outpatient treatment of acute venous thromboembolism: A randomized, double-blind, controlled trial." *Ann Intern Med* 138(9):714–719.

Kuperman, G. J., A. Bobb, et al. (2007). "Medication-related clinical decision support in computerized provider order entry systems: A review." *J Am Med Inform Assoc* 14(1): 29–40.

Lasalandra, M. (1998). Computers help to get "write" prescriptions. *The Boston Herald*.

Leon, A. R., W. T. Abraham, et al. (2005). "Safety of transvenous cardiac resynchronization system implantation in patients with chronic heart failure: Combined results of over 2,000 patients from a multicenter study program." *J Am Coll Cardiol* 46(12): 2348–2356.

Leapfrog Group (2008). "Factsheet: Computerized physician order entry." As of October 28, 2011: http://www.leapfroggroup.org/media/file/Leapfrog-Computer_Physician_Order_Entry_Fact_Sheet.pdf

Leapfrog Group (2008). *Leapfrog Hospital Survey Results*. Washington, DC.

Leucht, S., K. Wahlbeck, et al. (2003). "New generation antipsychotics versus low-potency conventional antipsychotics: A systematic review and meta-analysis." *Lancet* 361(9369): 1581–1589.

Lewis, S., and J. Lieberman (2008). "CATIE and CutLASS: Can we handle the truth?" *Brit J Psychiat* 192(3): 161–163.

Lieberman, J. A., T. S. Stroup, J. P. McEvoy, et al. (2005). "Effectiveness of antipsychotic drugs in patients with chronic schizophrenia." *N Engl J Med* 353(12): 1209–1223.

Lin, G. A., R. A. Dudley, et al. (2007). "Cardiologists' use of percutaneous coronary interventions for stable coronary artery disease." *Arch Intern Med* 167(15): 1604–1609.

Lindholm, L. H., B. Carlberg, and O. Samuelsson (2005). "Should beta blockers remain first choice in the treatment of primary hypertension? A meta-analysis." *Lancet* 366(9496): 1545–1553.

Lindholt, J. S., H. T. Sorensen, J. B. Michel, H. F. Thomsen, and E. W. Henneberg (2008). "Low-dose aspirin may prevent growth and later surgical repair of medium-sized abdominal aortic aneurysms." *Vasc Endovasc Surg* 42(4): 329–334.

Malmivaara, A., P. Slati, et al. (2007). "Surgical or nonoperative treatment for lumbar spinal stenosis? A randomized controlled trial." *Spine (Phila Pa 1976)* 32(1): 1–8.

Mas, J-L., G. Chatellier, B. Beyssen, et al. (2006). "Endarterectomy versus stenting in patients with symptomatic severe carotid stenosis." *N Engl J Med* 355(16): 1660–1671.

McAlister, F. A., J. Ezekowitz, D. M. Dryden, et al. (2007). Cardiac resynchronization therapy and implantable cardiac defibrillators in left ventricular systolic dysfunction. *Evidence Reports/Technology Assessments*. Rockville, MD, Agency for Healthcare Research and Quality.

McAlister, F. A., J. Ezekowitz, et al. (2007). "Cardiac resynchronization therapy for patients with left ventricular systolic dysfunction: A systematic review." *JAMA* 297(22): 2502–2514.

McFalls, E. O., H. B. Ward, T. E. Moritz, et al. (2004). "Coronary-artery revascularization before elective major vascular surgery." *N Engl J Med* 351(27): 2795–2804.

Mehra, M. R., C. W. Yancy, et al. (2009). "Evidence of clinical practice heterogeneity in the use of implantable cardioverter-defibrillators in heart failure and post-myocardial infarction left ventricular dysfunction: Findings from IMPROVE HF." *Heart Rhythm* 6(12): 1727–1734.

Miller, R. (2011). "Delaware insurance regulators find faults in insurers' imaging preauthorization process." heartwire (website). As of November 2, 2011: <http://www.theheart.org/article/1213741.do>

Mississippi Chapter of the ACC (2011). "New reports find RBMs aren't so appropriate!" As of April 24, 2011: <http://www.msacc.org/bbs/bbs2.html>

Morrissey, J. (2004). "Harmonic divergence: Cedars-Sinai joins others in holding off on CPOE." *Mod Healthcare* 34(8): 16.

National Conference of State Legislatures (2010). "Marketing and direct-to-consumer advertising (DTCA) of pharmaceuticals." As of March 24, 2011: <http://www.ncsl.org/default.aspx?tabid=14461#MN93>

National Quality Forum (NQF) (2009). *Safe Practices for Better Healthcare—2009 Update: A Consensus Report*. Washington, D.C.

Neale, T. (2011). "Major PCI Trial May Have Changed Practice." Medpage Today. As of April 24, 2011: <http://www.medpagetoday.com/Cardiology/PCI/26091>

O'Gara, P. T. (2010). "The COURAGE (Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation) trial: Can we deliver on its promise?" *J Am Coll Cardiol* 55(13): 1359–1361.

Oostenbrink, J. B., M. P. Rutten-van Molken, M. J. Al, J. A. Van Noord, and W. Vincken(2004). "One-year cost-effectiveness of tiotropium versus ipratropium to treat chronic obstructive pulmonary disease." *Eur Respir J* 23(2): 241–249.

O'Rourke, R. A. (2008). "Optimal medical therapy is a proven option for chronic stable angina." *J Am Coll Cardiol* 52(11): 905–907.

Owens, D. C. (2008). "What CATIE did: Some thoughts on implications deep and wide." *Psychiat Serv* 59(5): 530–533.

Owens, D. K., A. Qaseem, et al. (2011). "High-value, cost-conscious health care: Concepts for clinicians to evaluate the benefits, harms, and costs of medical interventions." *Ann Intern Med* 154(3): 174–180.

Parks, J. J., A. Q. Radke, et al. (2008). "Impact of the CATIE findings on state mental health policy." *Psychiat Serv* 59(5): 534–536.

Patel, M. R., G. J. Dehmer, et al. (2009). "ACCF/SCAI/STS/AATS/AHA/ASNC 2009 Appropriateness Criteria for Coronary Revascularization: A report by the American College of Cardiology Foundation Appropriateness Criteria Task Force, Society for Cardiovascular Angiography and Interventions, Society of Thoracic Surgeons, American Association for Thoracic Surgery, American Heart Association, and the American Society of Nuclear Cardiology Endorsed by the American Society of Echocardiography, the Heart Failure Society of America, and the Society of Cardiovascular Computed Tomography." *J Am Coll Cardiol* 53(6): 530–553.

Pedersen, T. R., O. Faergeman, J. J. Kastelein, et al. (2005). "High-dose atorvastatin vs usual-dose simvastatin for secondary prevention after myocardial infarction: The IDEAL study: A randomized controlled trial." *JAMA* 294(19): 2437–2445.

Peterson, E. D., and J. S. Rumsfeld (2008). "Finding the courage to reconsider medical therapy for stable angina." *N Engl J Med* 359(7): 751–753.

Piccini, J. P., A. F. Hernandez, et al. (2008). "Use of cardiac resynchronization therapy in patients hospitalized with heart failure." *Circulation* 118(9): 926–933.

Pickhardt, P. J., J. R. Choi, I. Hwang, et al. (2003). "Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults." *N Engl J Med* 349(23): 2191–2200.

Polinski, J. M., P. S. Wang, et al. (2007). "Medicaid's prior authorization program and access to atypical antipsychotic medications." *Health Aff (Millwood)* 26(3): 750–760.

Poole, J. E., M. J. Gleva, et al. (2010). "Complication rates associated with pacemaker or implantable cardioverter-defibrillator generator replacements and upgrade procedures: Results from the REPLACE registry." *Circulation* 122(16): 1553–U1543.

Poon, E. G., D. Blumenthal, et al. (2004). "Overcoming barriers to adopting and implementing computerized physician order entry systems in U.S. hospitals." *Health Aff (Millwood)* 23(4): 184–190.

Prasad, A., C. Rihal, et al. (2008). "The COURAGE trial in perspective." *Catheteriz and Cardiovasc Interven* 72(1): 54–59.

Reckmann, M. H., J. I. Westbrook, et al. (2009). "Does computerized provider order entry reduce prescribing errors for hospital inpatients? A systematic review." *J Am Med Inform Assoc* 16(5): 613–623.

Redberg, R. F. (2011). Squandering Medicare's money. *The New York Times*.

Reynolds, M. R., D. J. Cohen, et al. (2006). "The frequency and incremental cost of major complications among Medicare beneficiaries receiving implantable cardioverter-defibrillators." *J Am Coll Cardiol* 47(12): 2493–2497.

- Rogers, E. M. (1995). *The Diffusion of Innovations*. New York, Free Press.
- Rosenheck, R., D. Perlick, et al. (2003). "Effectiveness and cost of olanzapine and haloperidol in the treatment of schizophrenia: A randomized controlled trial." *JAMA* 290(20): 2693–2702.
- Rosenheck, R. A., D. L. Leslie, et al. (2008). "Second-generation antipsychotics: Cost-effectiveness, policy options, and political decision making." *Psychiat Serv* 59(5): 515–520.
- Rosenheck, R. A., and M. J. Sernyak (2009). "Developing a policy for second-generation antipsychotic drugs." *Health Aff (Millwood)* 28(5): w782–w793.
- Shaw, L. J., D. S. Berman, et al. (2008). "Optimal medical therapy with or without percutaneous coronary intervention to reduce ischemic burden: Results from the Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation (COURAGE) trial nuclear substudy." *Circulation* 117(10): 1283–1291.
- Sittig, D. F., J. S. Ash, et al. (2006). "Lessons from 'Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system.'" *Pediatrics* 118(2): 797–801.
- Tommaso, C. L. (2008). "One year perspective on COURAGE." *Catheter Cardiovasc Interv* 72(3): 426–429.
- Tosteson, A. N., J. S. Skinner, et al. (2008). "The cost effectiveness of surgical versus nonoperative treatment for lumbar disc herniation over two years: Evidence from the Spine Patient Outcomes Research Trial (SPORT)." *Spine* 33(19): 2108–2115.
- Weinstein, J. N., J. D. Lurie, et al. (2007). "Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis." *N Engl J Med* 356(22): 2257–2270.
- Weinstein, J. N., J. D. Lurie, et al. (2006). "Surgical vs nonoperative treatment for lumbar disk herniation: The Spine Patient Outcomes Research Trial (SPORT) observational cohort." *JAMA* 296(20): 2451–2459.
- Weinstein, J. N., T. D. Tosteson, et al. (2008). "Surgical versus nonsurgical therapy for lumbar spinal stenosis." *N Engl J Med* 358(8): 794–810.
- Weintraub, W. S., W. E. Boden, et al. (2008). "Cost-effectiveness of percutaneous coronary intervention in optimally treated stable coronary patients." *Circ-Cardiovas Qual and Outcomes* 1(1): 12–U33.
- Weintraub, W. S., J. A. Spertus, et al. (2008). "Effect of PCI on quality of life in patients with stable coronary disease." *N Engl J Med* 359(7): 677–687.
- Woods, S. W., H. Morgenstern, et al. (2010). "Incidence of tardive dyskinesia with atypical versus conventional antipsychotic medications: A prospective cohort study." *J Clin Psychiat* 71(4): 463–474.
- Young, J. B., W. T. Abraham, et al. (2003). "Combined cardiac resynchronization and implantable cardioversion defibrillation in advanced chronic heart failure: The MIRACLE ICD Trial." *JAMA* 289(20): 2685–2694.