

# From Pilots to Practice: Speeding the Movement of Successful Pilots to Effective Practice

**Peter Hussey, Richard Bankowitz, Michael Dinneen, David Kelleher, Karen Matsuoka,  
Joseph McCannon, Will Shrank, and Robert Saunders\***

April 23, 2013

*\*Participants in the Value Incentives Learning Collaborative  
of the IOM Roundtable on Value & Science-Driven Health Care*

---

*The views expressed in this discussion paper are those of the authors and not necessarily of the authors' organizations or of the Institute of Medicine. The paper is intended to help inform and stimulate discussion. It has not been subjected to the review procedures of the Institute of Medicine and is not a report of the Institute of Medicine or of the National Research Council.*

---

**INSTITUTE OF MEDICINE**  
OF THE NATIONAL ACADEMIES

**Advising the nation • Improving health**

## AUTHORS

**Peter Hussey**  
Policy Researcher  
RAND Corporation

**Richard Bankowitz**  
Chief Medical Officer  
Premier, Inc.

**Michael Dinneen**  
Director, Office of Strategy Management  
Department of Defense

**David Kelleher**  
President  
HealthCare Options, Inc.

**Karen Matsuoka**  
Director, Health Systems and Infrastructure  
Administration  
Maryland Department of Health and Mental  
Hygiene

**Joseph McCannon**  
Senior Advisor (*formerly*)  
Centers for Medicare & Medicaid Services

**Will Shrank**  
Director, Rapid Cycle Evaluation Group  
CMS Innovation Center

**Robert Saunders**  
Senior Program Officer  
Institute of Medicine

### The authors were assisted in their efforts by the following individuals:

**Katherine Baicker**  
Harvard University

**Susan DeVore**  
Premier Inc.

**Elliott Fisher**  
The Dartmouth Institute

**Elizabeth Johnston**  
Institute of Medicine

**Regina Julian**  
Military Health System

**Bob Kocher**  
Venrock

**Renee Mentnech**  
Centers for Medicare & Medicaid Services

**Bobby Milstein**  
ReThink Health

**Wynne Norton**  
University of Alabama–Birmingham

**Jon Perlin**  
Hospital Corporation of America

**Valerie Rohrbach**  
Institute of Medicine

**Gloria Sachdev**  
Purdue University

**Julia Sanders**  
Institute of Medicine

**Eric Schneider**  
RAND Corporation

**Stephen Shortell**  
University of California, Berkeley

**Alan Snell**  
St. Vincent Heal

**Suggested citation:** Hussey, P., Bankowitz, R., Dinneen, M., Kelleher, D., Matsuoka, K., McCannon, J., Shrank, W., and R. Saunders. 2013. *From Pilots to Practice: Speeding the movement of successful pilots to effective practice*. Discussion Paper, Institute of Medicine, Washington, DC.

<http://nam.edu/wp-content/uploads/2015/06/pilotstopractice>.

# **From Pilots to Practice: Speeding the Movement of Successful Pilots to Effective Practice**

Peter Hussey, Richard Bankowitz, Michael Dinneen, David Kelleher, Karen Matsuoka, Joseph McCannon, Will Shrank, and Robert Saunders<sup>1</sup>

Rising health care costs continue to stress budgets at all levels—family, employer, state, and national. At the same time, the results from health care are not commensurate with this level of investment. But there is cause for optimism that results can improve: the large number of individuals and organizations around the country who are involved in substantive efforts to improve care by piloting new care practices, care delivery models, payment methods, and other initiatives, thereby building an evidence base about what works and spreading that knowledge more broadly. To take best advantage of these activities, practical strategies are needed to accelerate and improve the planning, evaluation, scale-up, and spread of these initiatives. By improving the process for pilot projects, the potential becomes greater for large-scale improvements to the health care system—and the achievement of better care at lower cost.

## **THE IMPORTANCE OF PILOTS**

In recent years, there has been increasing concern about the rising cost of the health care system. In 2012, health care constituted 18 percent of the American economy, with expenditures of \$2.8 trillion (Martin et al., 2012). However, there are outstanding questions about the quality of care and patient health outcomes achieved from that substantial investment of resources. Medical errors remain too common, evidence is too rarely applied to health care, and quality remains uneven among different areas and populations (Classen et al., 2011; IOM, 1999, 2002; Landrigan et al., 2010; McGlynn et al., 2003; HHS, 2010). These deficits have highlighted the urgent need for effective, innovative approaches to improve health care.

Yet, there is optimism about the ability of the health care system to address these shortcomings. The system has entered a period of rapid experimentation with innovations designed to improve value—including new care practices, care delivery models, and payment methods. This experimentation is widespread, occurring throughout the health care system, and includes federal and state governments, hospitals, universities, physician groups, private companies, and others. Many of these experiments have been successful in improving the quality of health care, patient health outcomes, and overall value at specific sites, with specific groups of providers, or with specific patient groups. Nevertheless, the innovations piloted by these initiatives are rarely scaled up and spread.

This period of rapid innovation presents an opportunity to improve the health care system. The challenge is to rapidly improve understanding of which new ideas work, and under what circumstances, and use that information to scale up and spread successful approaches broadly. Our goal is to illustrate approaches to addressing this challenge. This paper examines practical lessons learned throughout the entire continuum of the pilot process—initial design,

---

<sup>1</sup> Working Group participants drawn from the Value Incentives Learning Collaborative of the IOM Roundtable on Value & Science-Driven Health Care.

planning, evaluation and learning, dissemination and implementation, and scale-up and spread. It also includes case studies of real pilots from inside and outside health care to demonstrate these key lessons. If these lessons are used widely to improve the process for pilot projects, the potential for large-scale improvement to the health care system becomes greater.

### **COMMON THEMES FROM CASE STUDIES**

Although there is substantial potential for pilot projects, the current process has several key challenges. To explore the challenges facing pilot projects in health care, we describe successful strategies for improving the pilot process. We identified these strategies by collecting a variety of case studies of pilot projects, including projects

- conducted by various organizations (e.g., state governments, large health systems, smaller health practices, and others);
- conducted by various industries (e.g., venture capital);
- using various evaluation methods (e.g., randomized controlled trials, observational studies, novel methods); and
- at various stages of their lifecycle (e.g., in planning, under evaluation, completed).

Based on our analysis of these case studies, several common themes emerged in developing, evaluating, and implementing successful pilots.

#### **Strategies for Successful Pilots**

- New evaluation methods, simulation and modeling tools, and evaluation techniques from other industries can be used to speed the assessment process and improve its generalizability.
- Starting quickly and improving a pilot project over time can ensure that any problems are addressed quickly, instead of at the conclusion of the pilot process.
- Pilots should anticipate how the project might spread and evaluate the initiative under a variety of real-world circumstances in order to be generalizable.
- Local leadership and culture, in addition to resources and data support, are vital to successfully implementing change.

#### **Progress Has Been Too Slow**

Given the enormity of the challenges the health care system faces in terms of cost and quality, the progress made by pilot initiatives has been too slow. In part, this is because the current process for developing, evaluating, and spreading pilot projects takes substantial time. Pilot projects can take many years, even more than a decade, to complete after initial conception, with lengthy approvals, site selections, data collection, and evaluations processes. Yet, after all this investment of time and resources, too few pilots result in transformative and sustained change.

## **Evidence Does Not Match Needs of Decision Makers**

When making decisions about which initiatives should be scaled up and spread widely, policy makers often lack the information they need about the effectiveness of different programs. To provide timely knowledge for policy decisions, the assessment process needs to be accelerated while retaining its rigor. Beyond their slow speed, current evaluations often do not provide evidence on how initiatives, which are often complex and consist of multiple components, work in real-world environments. Current evaluations often assess only whether an initiative succeeded in a specific context, and do not explicitly explain the mechanisms of how a particular program works or analyze the contextual factors (policy, cultural, and organizational) critical for its success (Berwick, 2008; Pawson and Tilley, 1997). Both types of information are important for highlighting where and when an initiative works. Without a focus on context, evaluation results deliver limited information about the potential for a pilot to be scaled up and spread to other settings and organizations.

## **New Models of Evaluation Are Needed**

To deliver the evidence that decision makers need, new models of evaluation are needed. Many current research designs, like traditional applications of randomized controlled trials, tend to focus on precise questions in well-controlled conditions in order to maximize the statistical validity of their results. Furthermore, randomized designs can be difficult to accomplish in a health care environment because of concerns about withholding potentially useful treatments or better care processes from patients and the difficulty of separating patients in control and intervention groups in a real-world health care delivery context (Alexander and Herald, 2009).

To be most useful, an evaluation model needs to balance several competing demands:

- meeting a certain standard of accuracy and generalizability;
- producing results as rapidly as possible to enable decisions; and
- making best use of resources, including time, effort, and other resources.

Various techniques now exist to balance those demands, ranging from mixed-methods studies with quantitative and qualitative components, effectiveness-implementation hybrid designs, realist evaluations, and pragmatic or practical clinical trials (Creswell et al., 2011; Curran et al., 2012; Fairall et al., 2012; Georgeu et al., 2012; Glasgow et al., 2005; Pawson and Tilley, 1997; Tunis et al., 2003). The particular research design needed for any given project depends on the questions to be addressed, the evidence to be generated, and the goals of the study.

## **Find Measures That Matter and Define Them Consistently**

All evaluations, regardless of the specific design, require the use of metrics that gauge important aspects of care, including clinical quality, cost, health outcomes, and patient experience. Often, the metrics that are chosen for evaluation are those that are easiest to measure, not necessarily those that are the most important. Different evaluations of similar interventions often use different measures, making comparisons of evidence difficult. Although sometimes it can be challenging to obtain the necessary data for high-priority measures, the move toward routine use

of electronic health records has expanded the quantity and quality of data potentially available for assessment.

Furthermore, the trend toward engaging patients throughout the health care system has promise in ensuring that pilot assessment is focused on measures that are meaningful for patients and consumers. Because patients are the focus of the health care system, these measures should assess aspects of care that are important to these stakeholders. This is necessary so that the measured changes lead to real improvements in care for patients and consumers.

Another important area for measure development is consistent measures of cost, resource use, and efficiency, particularly in assessing spillover effects and unintended effects. Such cost measures are essential to effectively gauging the extent to which different care and payment models are able to improve value. Indeed, the pressing need to bend the health care cost curve makes the development of such nationally consistent measures of cost and resource use a top priority.

Beyond cost measures, a common core set of measures that capture cost, quality, and health, defined consistently across pilot activities, can yield several benefits. A core set of measures can enable comparative analysis across pilots, allowing the identification and scaling of successful payment and delivery reform models. Moreover, successful models can be further analyzed to discover their common best practices, which can then be shared broadly across similar pilots.

### **Ensure That Programs Are Tried in a Wide Array of Real-World Settings**

The U.S. health care system is notable for its diversity, with organizations of various sizes (from large health care delivery systems to small practices with one or two clinicians), in various regions (from rural to urban), with various patient populations (from vulnerable to less vulnerable socioeconomic groups), and with differing missions (from academic to community health care organizations). Given this variety, initiatives that succeed in one context may not succeed in another or may require significant modifications to do so. Therefore, pilot projects should be tested in a variety of settings to ensure that the process and results can be replicated in diverse environments.

### **Planning Is Crucial to Success**

One reason for the slow pace of progress is logistical hurdles, regulatory barriers, or lengthy negotiations between different organizations and stakeholder groups. For example, failing to anticipate the need for regulatory approvals, such as those from institutional review boards (IRBs), can cause significant delays and increase the cost of a pilot. Similarly, pilots often may not acquire needed stakeholder buy-in or may underestimate the challenges in recruiting patients and obtaining data. Effective planning means foreseeing hurdles by identifying them in advance, allowing time for their likely occurrence, and developing strategies for overcoming them.

### **Design with the End in Mind**

Effective planning also means designing the pilot with the end in mind. For example, the Affordable Care Act authorizes the Centers for Medicare & Medicaid Services (CMS) Innovation Center to disseminate successful payment and delivery models without further congressional

approval as long as the models have been certified, by the CMS Office of the Actuary, to result in savings to the Medicare or Medicaid programs. Similarly, decision makers for other programs will consider the judgment of independent actuarial experts in whether the benefits of the model outweigh the costs of its implementation. With that in mind, the model should be designed to enable evaluations that produce the information needed by these actuaries or economists, such as the CMS actuaries and Office of Management and Budget economists.

Many pilots have the goal of improving value beyond a specific program, and seek to improve value across the health care system. Multipayer pilots can be an effective way of accomplishing this, but where this type of collaboration is not possible, tests should be implemented in such a way as to enable the evaluation of spillover effects and unintended consequences. This reinforces the importance of developing and collecting consistent measures to capture effects across the health system.

### **Start Quickly and Improve the Pilot over Time**

In the traditional form of pilot projects, the project is rolled out, maintains a fixed protocol during the pilot stage, and is assessed at the end. While this allows for easier statistical analysis, it also means that problems discovered during the pilot phase are not addressed until the conclusion of the pilot. Timely measurement during the pilot phase allows for fine-tuning of the project, can produce new insights, and allows stakeholders to learn quickly from failure (Gold et al., 2011). New statistical methods allow for the assessment of a project that changes over time, thereby providing the capacity for starting a project quickly and learning as it goes.

### **Several Barriers Exist to Dissemination, Broader Implementation, Scale-Up, and Spread**

Unless successful initiatives are disseminated and applied in regular practice, pilots have little value. Unfortunately, the reality is that many effective pilot projects are conducted each year, yet few become widely used and many others are used only in limited or superficial ways. Without a stronger focus on spread and scale-up, considerable effort will be spent on developing new ideas that improve care in a specific pilot situation, while the overall health care enterprise continues to confront shortfalls in care quality, cost, and health outcomes.

One major factor that affects the adoption of a particular initiative is the incentive structure of health care (Timbie et al., 2012). The incentives for much of health care currently reward the volume of services rather than the value or quality of the care. This discourages new ideas or initiatives that would reduce the quantity of care, even if they produce better outcomes for patients. In academic medicine, incentives for faculty and researchers, in terms of promotion, tenure, and career advancement, are for discovering and publishing new ideas. There are few rewards for individuals who focus on implementing or spreading ideas to a broad audience.

Previous studies of scale-up and spread have identified a number of factors that affect the use of new ideas in regular care, which include (Greenhalgh et al., 2004; McCannon et al., 2008)

- the *environment and context* in which scale-up will occur;
- the *evidence and foundation* behind the initiative;
- the *framing and communications* used;
- the *type of the initiative* to be spread; and
- the *strategies and method* for spreading information.

The number of factors in play highlights the complexity of scale-up and spread. This complexity is compounded by the fact that each factor has varying importance for clinicians, health care delivery organizations, payers, geographies, and patients. Furthermore, the process of scaling up new concepts is dynamic and organic. New practice patterns or interventions are not static once adopted; rather, organizations will adapt an initiative to the realities of their clinical practice and change it over time based on their needs. To adapt the innovation to other settings, the project leaders must identify which components of the initiative are fundamental and which can be altered to local conditions (Coburn, 2003; Greenhalgh et al., 2004).

### **Insufficient Tools for Scaling and Spreading Pilots**

Given the challenges involved in the scale-up and spread of new initiatives, greater attention is needed on sustainable methods for accomplishing this routinely for promising innovations. Unfortunately, scale-up and spread often receive little attention, due to a common assumption that good ideas will automatically spread due to their own merits. The standard process for spreading new pilot projects often relies solely on publishing them, often in peer-reviewed journals, and presenting them at conferences and other venues. Although this is an important first step in spreading new ideas, it is insufficient. There is too much information currently available for clinicians, health care delivery leaders, payers, and others to regularly apply to their practice. Overcoming this challenge requires implementing new methods, such as clinical decision support, to disseminate knowledge when it is useful to individuals and organizations that could implement it.

However, delivering more knowledge rarely changes clinical and organizational behaviors. Rather, new tools are needed to accelerate and promote the process of scale-up and spread throughout the nation. One effective technique is to create learning communities or collaboratives that share and communicate what works among their members. This has proved useful in several cases for rapidly disseminating the results of a particular initiative among a large number of stakeholders. Other tools include working with opinion leaders inside organizations, in-person feedback, and feedback on performance and practice variation (IOM, 2012). It is unknown which tools work best for which conditions, and many tools will need to be customized to local conditions.

Although tools are important, not all efforts to rapidly scale up initiatives may be well founded. Sometimes, it may take considerable time to fully understand the long-term effects of a particular project and evaluate whether it actually improved care. This caution highlights the need for a nuanced approach to scale-up and spread that continuously learns and improves over time.

### **KEY QUESTIONS FOR PILOTS**

Building on these themes, we identified several questions that all pilot projects should consider. They range from questions to ask during initial planning, to issues to consider in evaluation, to steps for promoting the scale-up and spread of the pilot project if it proves successful.

## Key Questions for Pilots

### *Planning and Starting Pilots*

- **Plan the pilot process.** What are the activities that will lead to success and how are the activities logically connected to the expected outcomes? What are the project's time frames for planning, data collection, and evaluation? (See cases 2,3)
- **Build the pilot on existing knowledge.** Will standardized protocols be employed in the pilot? If, so have you established the evidence basis for these protocols or will this be developed during the pilot? (See case 1)
- **Start as soon as possible and learn quickly.** What must be planned before starting, and what can be learned along the way? Is there an opportunity to phase the project so that it can adapt to early results?
- **Adapt the project to learn over time.** As the project is implemented, who will be responsible for early evaluation, modifying the project as necessary, and documenting what is actually implemented? If improvement goals are dynamic, what data will be used to identify the greatest opportunities for improvement? (See cases 1,2,9)

### *Accessing the Necessary Data*

- **Identify needs.** What infrastructure is needed for the project? How, by whom, and when will data be collected? Are any special tools, staffing, or training needed for data capture? (See cases 2,3,7,9)
- **Account for regulatory and organizational challenges.** What approvals are necessary to collect these data? Who will request approvals and how long does it normally take to obtain them? (See cases 2,3)

### *Assessing Success*

- **Identify measures that matter.** How is success defined in terms of important economic, clinical, or health impacts? What are the concrete, preferably quantitative, goals? Why are these goals important? (See cases 2,7,9)
- **Identify appropriate methods for assessing success.** Because different methods of evaluation have different strengths, have you identified what questions are most important to answer for this project and what evaluation techniques best answer those questions? Is there an opportunity to conduct a natural experiment? (See cases 1,4,5,6,7)
- **Evaluate how the pilot works in diverse environments.** How does the project assess success in diverse health care environments—different-sized organizations, different geographic areas, different technological capabilities? (See cases 4, 7)
- **Ensure that evaluation answers the needs of decision makers.** What information is needed by the individuals who will decide about the project's scale-up and spread? How have these stakeholders been involved, and what are their perceptions of the initiative? How does the project evaluation address their needs? (See cases 3,6,8)

### *Scale-Up and Spread*

- **Understand reasons the project might spread.** Why would other organizations or individuals want to adopt the piloted initiative? What gap or need does it fill? What is your theory about how change or adoption will occur? What needs to take place for this to happen? (See case 3, 9)
- **Outline the incentives and environmental factors that could promote adoption.** What incentives are most likely to stimulate interest in the initiative and change behavior? How should they be used? (See case 8)
- **Identify how the project can adapt.** How can the piloted initiative be adapted to local conditions or different health care settings? (See case 8)
- **Form a learning community.** Can a learning community be created to share best practices and lessons learned from the pilot? (See case 7)

## CASE STUDIES

The case studies in this section exemplify the range of challenges that pilots face. Yet, they also identify several new opportunities that exist, such as in gathering data and sharing information. Finally, these examples highlight several potential strategies that could be adopted to improve future pilot projects and the overall process for health care pilots.

### Case Study 1: Oregon Medicaid Experiment

In 2004, the state of Oregon closed enrollment for the Oregon Health Plan Standard, an initiative that provided coverage to low-income adults not eligible for the traditional Medicaid program, given budgetary pressures on the state. As these budgetary pressures eased in 2008, the state determined it could add 10,000 individuals to the program, and chose to do so by lottery to be as fair as possible. Approximately 90,000 people entered the waiting list during the 5-week sign-up window, 30,000 were selected to apply to the program, and approximately 10,000 enrolled (Finkelstein et al., 2011).

By making the policy change this way, the state achieved two objectives: (1) it expanded coverage equitably, and (2) it increased the knowledge base on the effects of expanding insurance coverage to additional low-income adults. Although there are several studies that compare insured and uninsured populations using a variety of health metrics, these studies all rely on observational data. Therefore, the studies' results were affected not only by whether an individual had insurance, but possibly by other personal characteristics such as income, employment, education, or even initial health. Many of these personal characteristics may not be easily measurable, making it difficult to control for them fully. A randomized trial can overcome these challenges, but such experiments are rarely completed due to infeasibility, cost, political challenges, or concerns about withholding available insurance coverage. The Oregon health insurance lottery, therefore, was a unique opportunity to learn more about the effects of insurance coverage and, more specifically, Medicaid coverage, in the context of a natural experiment (Finkelstein et al., 2011).

To understand the effects of this randomized controlled trial, the state worked closely with a team of researchers to assess the expansion and provide new knowledge on the effects of covering the previously uninsured. The researchers considered a number of outcomes, including access to health care, utilization, impact on household finances and debt, health behaviors, physical and mental health outcomes, effects on employment, and other measures of health and well-being. To examine these effects with adequate statistical power, the study used a range of data sources. These included administrative data on hospital discharges, mortality, and credit records. The researchers also gathered survey data by mail and telephone, conducted in-person interviews, and examined health screenings (Allen et al., 2010). This robust randomized natural experiment built on earlier observational studies examining the effects of Medicaid coverage on patient health and access (Oregon Office for Health Policy and Research, 2009).

The results from the first year of the expansion—although the analysis will be ongoing—found that those who received Medicaid coverage were more likely to receive health care services, obtain consistent primary care, and use preventive care. In addition, the participants experienced improved financial security, with fewer medical bills sent to a collections agency and fewer individuals undertaking new borrowing for health expenses. Furthermore, the participants reported improved health status and overall well-being. However, the researchers advise caution

in extrapolating the Oregon results, because policy changes depend on a wide variety of organizational, cultural, and population factors (Baicker and Finkelstein, 2011).

*Key Lessons Learned:*

- It is possible to embed evaluation into emerging policy initiatives through partnerships with researchers.
- It is possible to conduct a natural experiment through the phased implementation of a policy.
- Considering local needs and preferences is critical to success.
- Extrapolating results to other contexts must be done with caution.

**Case Study 2: St. Vincent Health and Central Indiana Beacon Collaborative**

Under the Central Indiana Beacon Collaborative, St. Vincent Health led a project to reduce the rates at which patients are readmitted to the hospital within 30 days after discharge at 13 hospitals. Focusing on their two most common conditions for readmissions, congestive heart failure and chronic obstructive pulmonary disease, the project designed a program of home monitoring for patients with these conditions using telemonitoring with videoconference support. Once designed, the project leaders sought to evaluate its effectiveness at reducing readmission rates using a randomized controlled research trial with 3,000 patients. To prepare for the evaluation, project leaders selected, purchased, distributed, and deployed telemonitoring equipment; arranged for patient monitoring; and sought regulatory approvals from the various institutions involved.

One early challenge the study faced was obtaining IRB approval, because each institution used different processes and applications that required multiple follow-ups. The approval process took more than 6 months to complete and delayed the start of the project. A second challenge was ensuring that the necessary evaluation data were available, especially as patients are not always readmitted to the same hospital. The project now seeks to integrate data from their regional health information exchange to complete the existing hospital-based data sources. The study also expended considerable energy to enroll the necessary number of patients, to satisfy lengthy (IRB-approved) protocols, and to deal with time constraints on hospital case managers. Some of the strategies used to overcome these challenges were integrating the enrollment process with discharge and other workflow processes, simplifying patient enrollment forms for ease of understanding, and engaging physicians in understanding the importance of the trial.

Finally, by examining early results, the sponsors recognized that enrollment in the inpatient randomized control trial would not fully make use of all of the pilot's telemonitoring resources. They reacted by adapting the program to include complex ambulatory patients, who are also at risk for hospitalization, using a pre-post study design. These outpatient results proved decisive in the parent organization's decision to maintain the program after the pilot phase and to spread this program to other geographic areas and patient populations.

### *Key Lessons Learned:*

- Identify and secure the necessary regulatory approvals as soon as possible to avoid delays.
- Specify in advance what data you will need to evaluate the project and how these data will be captured.
- Engaging patients and clinicians is critical to ensure adequate patient recruitment.
- Use early program results to improve the pilot over time in ways that maximize its value and increase the potential for the intervention to be implemented at scale.

### **Case Study 3: Quality Health First (QHF) Program**

Seeking to improve chronic disease care and preventive health services, the Indiana Health Information Exchange launched the QHF program. The program focuses on improving primary care through quality measurement, the provision of alerts and reminders to individual providers, public reporting of results at the practice-site level, and a multipayer pay-for-performance program. As part of its work, it has implemented a pilot program, supported by a Beacon Community grant, focused on improving clinical outcomes for patients with diabetes. The goal of the project is to increase the number of patients whose blood sugar levels are controlled. After cataloguing the care models currently in use in the community, the project identified three promising diabetes care models (represented by six programs)—a clinical pharmacist model, a registered nurse/registered dietician model, and a registered nurse case manager model.

The impact of each care model was initially analyzed based on self-reported data for small cohorts of patients with poor blood sugar control (HgbA1c  $\geq$ 9.0 percent). Patients were enrolled in the study and evaluation is currently under way, using observational data to compare the models against a control group. These data will then be analyzed using an economic model to estimate the value provided to payers by each clinical model. Because successful spread of the intervention depends upon reimbursement from health insurers, the program obtained input on its evaluation tool from the two largest carriers in the community.

One key challenge in this pilot was accessing the necessary data, due to barriers in executing the required data-use agreements, because data-use agreements are particularly sensitive when clinical information (such as current HgbA1c results) is used in the study. The pilot also discovered that developing a data collection tool for six different programs was more challenging than expected. The challenges involved in designing and implementing processes, tools, and permissions in order to obtain data slowed both the initial implementation and the evaluation of the project. These challenges highlight several lessons for future pilots. First, pilots should identify in advance whether data-use agreements will be necessary and include time to negotiate these agreements. Second, pilots should develop the tools and processes necessary to rapidly collect evaluation data as early as possible to prevent delays. Third, projects should involve the stakeholders responsible for making decisions about scaling up the pilot. This involvement helps to ensure that the project produces the evidence needed to support decisions about implementing or scaling up the initiative. For the QHF project, those stakeholders are the insurance carriers that will be asked to provide reimbursement to sustain successful programs. Finally, it is important to assess each participating organization's capacity to perform the tasks

assigned to it in a timely manner. This applies to administrative tasks such as data collection and to clinical capacity for treating patients.

*Key Lessons Learned:*

- Obtaining data-use agreements can be a lengthy process.
- If scale-up or spread depends on outside parties, it is important to anticipate their information needs.
- Specify in advance who, using what methods, will collect what data.
- Assure that each participating organization has the capacity to perform the functions required of them by the pilot.

**Case Study 4: Military Health System Patient-Centered Medical Homes Initiative**

The Military Health System (MHS) provides health insurance and direct health care services for almost 10 million eligible beneficiaries. Almost 4 million beneficiaries are enrolled in Military Treatment Facilities (MTFs) in the MHS's direct care system, which consists of 56 medical centers/hospitals and 364 medical clinics. There is great diversity in the patient populations served by the MHS (ranging from young adults to retirees) and in service locations (including rural, urban, domestic, international, and maritime sites). There is also diversity in the context of its operations, with some facilities providing routine care for individuals stationed in the United States and others treating wounded individuals in combat theaters.

Due to concerns about decreasing patient satisfaction and rising costs in the direct care system, the MHS explored several strategies and models for improving primary care. One early pilot site, the Bethesda National Naval Medical Center (now the Walter Reed National Military Medical Center), adopted a patient-centered medical home (PCMH) model and saw significant performance improvements in access, care continuity, and preventive and chronic care management. Based on the success of this example and of other MHS sites that had adopted the PCMH model of care, the MHS leadership decided in 2008 to spread the PCMH model to all 435 primary care practices in the system. To support the PCMH strategy, the MHS increased staffing at its primary care clinics in return for expected performance improvements, including clinical transformation, increased satisfaction, continuity, access, and private-sector care recapture.

In 2009, the MHS codified its PCMH implementation strategy in a Department of Defense/Health Affairs policy; the Uniformed Services (Army, Navy, and Air Force) followed with formal implementation instructions in 2010. In order to ensure consistent implementation of PCMH principles across all branches, MHS leadership also made the decision that all primary care practices should be formally recognized by the National Committee for Quality Assurance (NCQA) at one of the higher recognition levels (Level 2 or 3). By the end of 2011, 48 practices had achieved NCQA recognition. The MHS accelerated PCMH implementation and NCQA recognition in 2012, and, by the end of the year, more than 160 primary care practices serving almost 50 percent of all direct care beneficiaries had achieved formal recognition. Of the number of formally recognized primary care practices, more than 90 percent were recognized at Level 3, the NCQA's highest level. Due to accelerated implementation, almost all MTF beneficiaries are expected to be enrolled in NCQA-recognized PCMHs by the close of 2014.

Several elements have been added to support practices in this transition, including sustainment guidance, training, integration of other services, including behavioral health, and an enhanced digital infrastructure. In 2013, the MHS expects to implement a nurse advice line, providing an array of services including appointments to an MTF PCMH. The MHS also has initiated several specialty care pilots, which are designed to increase shared patient accountability and system implementation. In the future, the MHS will provide real-time practice patterns and actionable patient information directly to providers and further integrate other types of health care services into the MHS's PCMH model in order to support the MHS's transition to an integrated delivery system and an accountable care organization (ACO.)

Measurement continues to be a major focus as the pilot process moves forward. The Tri-Service PCMH Advisory Board overseeing the primary care transformation includes a subgroup focused on performance measures. The measures subgroup recommends targets, monitors performance, recommends new measures, and identifies best practices from top-performing PCMHs. In analyzing progress, the NCQA PCMH recognition standards have played a critical role, as these are a recognized set of metrics for assessing progress in multiple components of the PCMH model. Beyond these standards, other metrics that have been assessed include continuity of care, utilization of emergency services, access to acute and routine care, per capita costs, and patient and staff satisfaction. In the future, the program intends to develop additional measures; these measures will include primary care, specialty care utilization, lab, radiology and pharmacy utilization, health outcomes, admissions, and integrated behavioral health measures. Performance measure results are reviewed not only by senior MHS leadership but by the armed services, MTFs, and PCMHs. Furthermore, the MHS PCMH criteria include the requirement that all PCMHs monitor, publicly post, and advance key performance measures.

#### **Patient-Centered Medical Home Evaluators' Collaborative**

Another initiative to improve the measurement of medical homes is the Commonwealth Fund-sponsored Patient-Centered Medical Home Evaluators' Collaborative, started in 2009. The Evaluators' Collaborative aims to establish a standard assessment process and core set of measures through consensus and to share the consensus results broadly. This has become more critical as the PCMH model has become popular, with more than 90 commercial health plans, 42 states, and 3 federal initiatives testing the model. The Evaluators' Collaborative recently released its recommended core set of measures for assessing cost, utilization, and clinical quality outcomes (Crabtree et al., 2011; Rosenthal et al., 2012). Although several elements of the medical home models have been shown to improve quality and reduce cost, few evaluations have been published that assess its overall impact. This lack of data signals an urgent need to increase the evidence base for the model, which can indicate its overall effectiveness as well as provide lessons on what elements provide the greatest effect.

There are multiple lessons to be learned from the MHS PCMH project. One lesson is that the support of senior leadership was critical in focusing attention and providing resources for the program's success. Furthermore, collaboration across the three military health branches (Army, Air Force, and Navy) was also crucial to drive standardization and consistent implementation. This collaboration was achieved through a PCMH Advisory Board composed of stakeholders across the whole MHS that met regularly. In addition, PCMH implementation was enhanced by soliciting feedback from the PCMH teams at the MTFs. Finally, transparency was important in

motivating improvement. This transparency was achieved by providing all clinics with the status of Tri-Service PMCH Advisory Board activities as well as by the requirement that all clinics monitor and post their performance measures, allowing for results and outcomes to be visible.

*Key Lessons Learned:*

- Support of senior leadership is important for driving change.
- Transparency of performance is required to motivate improvement and measure progress.
- A culture of collaboration allows for sharing of best practices and consistent implementation.
- Standardized metrics allow for comparisons across sites and the identification of successful strategies.

### **Case Study 5: REDUCE-MRSA Trial**

MRSA (methicillin-resistant *Staphylococcus aureus*) is a common hospital-acquired infection that can result in death and morbidity for patients. Beyond the impact on patient health, these infections can increase the cost of care with intensive treatments and additional services. Yet, it is unknown what strategy is most effective in preventing such infections in hospitals. Three different interventions are in common use: active screening and isolation (usual care), active screening and decolonization of MRSA carriers, and universal decolonization without regard for MRSA status. Evidence supports the effectiveness of all three interventions in reducing infection, and all three approaches had been previously implemented by hospitals, but there was no comparative evidence on which was intervention most effective.

To identify the most effective of these three strategies, the REDUCE-MRSA trial (Randomized Evaluation of Decolonization versus Universal Clearance to Eliminate MRSA) used a novel evaluation approach (a cluster-randomized trial) to examine different ways of preventing MRSA infections in hospital intensive care units (ICUs) (Platt et al., 2010). The pilot was conducted between September 2009 and September 2011 in 42 Hospital Corporation of America (HCA) hospitals, comprising approximately 70 intensive care units. The trial took advantage of existing personnel, procedures, infrastructure, and information systems in order to perform an evaluation under usual practice conditions with lower costs. Data have been collected and are being analyzed now.

The leaders of the pilot chose to use a cluster-randomized design because of its efficiency and low cost (Platt et al., 2010). Unlike other types of randomized controlled trials, this type of evaluation randomizes providers, clinics, and organizations instead of individual patients to treatment and control groups. This type of evaluation has several benefits—its results have strong statistical validity due to randomization, it provides evidence on how interventions work in real-world health care settings, and it has lower cost and time requirements. In its application to the REDUCE-MRSA trial, hospitals were randomly assigned to one of the three infection control interventions after being stratified into three groups based on their ICU patient volume.

The use of a cluster-randomized design was credited by the pilot team with achieving several key efficiencies. Overall, the cost of conducting an evaluation of the pilot (as opposed to implementing the interventions without randomization) was less than \$2 million. Features of the pilot that were identified as leading to efficiencies included support of the system's leaders; streamlined implementation; the ability to use existing resources, personnel, and policies; col-

lecting data through routine care; and a data infrastructure that allowed centralized access to the data needed for the study (Platt et al., 2010). In addition, a history of collaboration among the members of the study team enabled rapid design and implementation of the pilot.

There are also several limitations to cluster-randomized trials. The assignment of groups to interventions without individual consent raises ethical questions about participation in research. Cluster-randomized evaluations may be less costly than individual-level randomized trials, but they are likely more costly and logistically complex than simple observational studies. However, this type of design may be one of the most efficient ways to balance robust evaluation design and the cost and feasibility of evaluations of the effectiveness of health care delivery interventions.

Another pilot program at HCA underscores the role of organizational factors in a pilot's success. In the HCA "39 Weeks" pilot, 27 hospitals in the HCA system in 14 states participated in observing the effect of elective early delivery at different gestational ages (after 37 weeks). The initial trial found that elective delivery before 39 weeks increased neonatal morbidity (Clark et al., 2009). After this result was found, another trial was conducted on three potential strategies for reducing such elective deliveries at the pilot facilities, with sites chosen for geographic and demographic representativeness. The three strategies were (1) a hard stop (no elective deliveries allowed except for special cases), (2) a soft stop (physicians were allowed to order an elective delivery, but such cases were reviewed retrospectively), and (3) an education-only campaign. The medical staffs of each facility were allowed to choose their preferred method after being informed of the system's intent to restrict this practice based on patient safety considerations. That trial found that elective early delivery was reduced from 10 percent to 4 percent of overall deliveries and that the rate of neonatal ICU admissions fell by 16 percent. The greatest improvement was found when elective deliveries before 39 weeks were simply not allowed. This trial highlights the relative ineffectiveness of education alone in changing physician behavior (Clark et al., 2010). In implementing a large-scale change, HCA found a number of organizational factors that were necessary for change: executive support, a business case, effective communication strategies, and personal and institutional openness to change.

#### *Key Lessons Learned:*

- New forms of evaluation, such as cluster-randomized trials, can be less costly and more feasible to implement while continuing to provide the benefits of high-quality evidence.
- Cluster-randomized trials may provide a good balance of robustness and efficiency for health care delivery pilots.
- Organizational factors, such as leadership and culture, influence whether a trial succeeds.

### **Case Study 6: Venture Capital**

Other industries have also developed methods for assessing the success of different ideas under uncertainty. One industry that conducts many such evaluations of new ideas is the venture capital sector. To review a potential venture investment, firms complete a due diligence process to understand all aspects of the company and the product or service they intend to produce. This selection process must be done in the context of a large number of proposals considered every year. For example, one firm receives 800 plans every year, seriously examines 20 to 25 of those plans, and closes 2 to 3 deals that year.

To complete their evaluation process, the firm considers a wide variety of factors, including

- Validity of the idea
- Management
- Financial model
- Sales and reimbursement projections
- Clinical, scientific, and technical validity
- Operations capacity
- Market and competitive analysis
- Financial and accounting
- Regulatory and legal landscape
- Intellectual property
- Exit scenarios
- Capital structure
- Governance

Although not all of these factors are directly applicable to the pilot evaluation process in health care, many of them are. This case study underscores that an idea, product, or service with excellent technical or scientific credentials may not be feasible to put into practice due to inadequate leadership, limited implementation capacity, poor culture, or other environmental factors. In the end, there is an acknowledgment that new initiatives require a leap of faith, with intangible factors playing a critical component (Suennen, 2011; Suennen et al., 2011).

There are several other core principles from venture capital that may be applicable to the health care sector. Given the number of factors firms consider in their evaluation decisions, one prerequisite for investing in an idea is a demonstrated proof of concept. Once that has been demonstrated, other common practices to ensure success are supporting pilots with data and ongoing feedback to help with course-correction; tying funding to achieved milestones; and tracking progress on those milestones through consistently defined performance metrics. Overall, there is an interest in ensuring that metrics and goals are clear but leave room for flexibility on implementation in order to spur innovation. Similarly, there is an interest in providing design flexibility, as overly rigid specifications could create artificially high barriers to entry or inflate implementation costs to a point that limits spread and dissemination. For example, not limiting which types of providers can participate in ACOs can give ACOs flexibility to utilize a range of health professionals to meet the specific health needs of their communities.

#### *Key Lessons Learned:*

- There are lessons about evaluation to be learned from other industries.
- Many factors influence the spread of new ideas, including culture, leadership, operations, and environmental factors.
- Quantitative and qualitative approaches are needed to fully understand an idea's feasibility.
- Providing clear goals along with substantial flexibility in design and implementation can spur innovation.
- Using consistently defined performance measures can enable rapid-cycle comparative analysis of which interventions perform best and facilitate broader dissemination by lowering implementation costs.
- Monitoring performance on an ongoing basis, rather than relying on post-hoc evaluations, enables course corrections for pilots that are not performing to expectation. Pilots require timely and ongoing data feeds to enable this type of continuous quality improvement.

## Case Study 7: QUEST Program

Launched in 2008, the QUEST program, sponsored by Premier and assisted through a strategic partnership with the Institute for Healthcare Improvement, was designed to improve hospital performance along six dimensions—(1) adherence to evidence-based care, (2) mortality, (3) efficiency, (4) reducing patient harm, (5) improving the patient experience, and (6) lowering costs—and later expanded to include reducing readmissions. The program began with a cohort of 156 hospitals as charter members, and sought to improve performance at these hospitals using standardized performance measures, transparency among all participants, and learning collaboratives to share best practices. During the first 3 years of the program, the charter hospitals were able to reduce their median risk-adjusted mortality from 0.94 to 0.64—an absolute reduction of 30 percent—and were simultaneously able to reduce their median costs by more than \$1,500 per case-mix-adjusted discharge on an inflation-adjusted basis. Since its inception, the program has grown to include almost 350 hospitals.

Several evaluations are under way. One aspect of the assessment used a realist evaluation framework to examine the factors that allowed some hospitals to make improvements, the activities top-performing hospitals conducted to make improvements, and how the QUEST project contributed to performance improvement. Realist evaluation relies on the fundamental principle that an intervention interacting with a local context is what leads to change. This type of evaluation focuses on two questions—how the intervention brings about change, and what works for whom under what circumstances. For this project, the realist evaluation used interviews with participating hospitals (especially top-performing hospitals and those that have achieved rapid improvement) in addition to site visits with hospitals that had rapidly improved in QUEST performance metrics.

The QUEST evaluation found that some hospitals made substantial improvements on three of the metrics (evidence-based care, mortality, and efficiency), while others were stable or made more modest improvements on those dimensions. The assessment found that rapid improvement was associated with the following organizational features:

- the presence of leadership champions who were visibly involved and active in removing barriers,
- aligning improvement projects with mission and goals,
- engaging staff,
- using a specific improvement method,
- using timely and reliable data,
- using a specific execution framework, and
- taking advantage of a learning community.

In addition, the evaluation found that top-performing hospitals used the following mechanisms to improve performance:

- using data to identify improvement opportunities,
- using evidence to improve performance and standardize care,
- holding staff accountable, and
- providing education and feedback to staff.

These findings underline the importance of organizational factors in spreading and scaling up pilot projects (Van Citters and Nelson, 2011).

The launch and execution of this collaborative afforded additional lessons for implementation and scaling. First, an enhanced data infrastructure is a prerequisite for improvement. In this project, an existing data collection infrastructure was available for most metrics because participants were already using a set of performance improvement applications developed by Premier. Where this was not adequate, such as in measuring costs, participants rapidly agreed on a standardized mechanism for collection. Second, metrics need to be defined early by the participants. For this project, participants focused on defining metrics from the very beginning, including definitions of the key performance metrics, risk adjustment (if any), and any exclusions. Where no standard metrics existed, such as for harm, the participants agreed to start working on any known problems within their institutions while metrics were being developed. In addition to aiding evaluation, a standardized set of metrics facilitated the addition of new members to the cohort by providing a clear definition of success. Third, success was defined in absolute rather than relative terms. Participants decided together what constituted a top performance threshold, and these thresholds remained fixed and not relative. This discouraged a tournament mentality and encouraged active collaboration among the sites. Finally, transparency is an important tool for improvement. Participants also agreed to transparency of the data, including key performance metrics and underlying data, thus expediting the identification of pockets of excellence in each domain (such as in sepsis mortality). The cohort also agreed to rapid dissemination of best practice via an Internet portal, weekly webcasts, and structured collaborative execution structures (such as rapid performance “sprints”).

#### *Key Lessons Learned:*

- An evaluative framework that includes the institution’s context as well as the intervention (a realist framework) can uncover structural or cultural elements critical for success.
- Data transparency and best-practice sharing are important for accelerating changes in practice patterns. Transparency promotes competition, while fixed, nonrelative thresholds create a collaborative environment in which all can succeed.
- Scale-up is facilitated if performance metrics and definitions are standardized and agreed upon in advance, if definitions of top performance are known, and if a data collection infrastructure can be rapidly developed using existing components.
- Notwithstanding the above, for advancement to take place as rapidly as possible, participants should not delay improvement activities while awaiting consensus on metrics or methodologies.

### **Case Study 8: ReThink Health Dynamics**

Simulation modeling offers the potential to speed pilot projects by exposing the conditions under which various interventions may be most effective and by estimating the likely short- and long-term consequences. This mode of interactive learning is used by ReThink Health Dynamics, a project of the ReThink Health alliance supported by the Fannie E. Rippel Foundation, the California HealthCare Foundation, and other allies. The goal is to simulate at a local level the health and economic effects of different regional reform initiatives. As with all models, the value in studying simulated scenarios is not in the numerical projections per se, but rather in the ability

to compare various scenarios in terms of their relative direction, timing, and magnitude of effects (Sternman, 2006). Simulated results also help diverse stakeholders to view, and interact with, the health system's potential responses to different interventions, which in turn gives them greater foresight into the interconnections and likely effects of major policy options.

Building upon the HealthBound policy simulation model developed by the Centers for Disease Control and Prevention, the ReThink Health model represents several key features of a local health system (Hirsch et al., 2012; Milstein et al., 2011). Its simulations account for the demographic composition of the local population (such as age distribution, economic status, and insurance coverage); social, environmental, and economic conditions (such as behavioral policies, environmental hazards, and crime), as well as patterns in health care delivery and costs (such as adequacy of preventive and chronic care, sufficiency of primary care providers, total cost of care, and health care inflation). When testing intervention options, the model estimates the likely impacts over time on many performance metrics, including disease burden, access to care, utilization of care in different venues, premature mortality, health care costs, worker productivity, equity among subgroups, and return on investment. Many initiatives may be simulated individually and in combination, including those that address risk reduction, care improvement, workforce capacity, cost saving, value improvement, financial incentives, and changes in funding mechanisms. A particular strength of this tool is its ability to represent funding structures that involve capturing and reinvesting health care cost savings (Fisher et al., 2009; Magnan et al., 2012).

Beginning in 2011, the ReThink Health Dynamics model was developed and piloted in five regions: Pueblo, Colorado; Manchester, New Hampshire; Contra Costa County, California; Alameda County, California; and Whatcom County, Washington). It has since expanded to represent other regions, including Atlanta, Georgia, Morristown, New Jersey, and the United States overall. Users have identified common potential pitfalls that threaten to undermine policy initiatives, including unsustainable program financing, capacity bottlenecks (such as lack of primary care providers), provider pushback, biases in using short time horizons to evaluate outcomes, and the entrenched nature of health inequities (Hirsch et al., 2012).

For example, planners in several cities have studied scenarios involving greater use of evidence-based guidelines and the establishment of medical homes. Using the model, they were able to see how these initiatives intensify the demand on primary care providers and could backfire to produce greater costs in situations where there is insufficient capacity to support both initiatives at the same time. In those circumstances, the added demand may leave people, especially those at the lower end of the socioeconomic spectrum, with less access to primary care, and force them to seek care in higher-cost settings, such as emergency rooms. This example highlights the importance of combining multiple complementary interventions to improve the chances of success.

A formative evaluation of the initial pilot sites found that several leadership capabilities are needed for ReThink Health modeling to work most effectively: local leaders must be able to convene and coordinate stakeholders, gather the necessary data, develop ownership of the model, engage organizational champions, and support these capabilities with the needed resources.

*Key Lessons Learned:*

- Simulation modeling can improve the success of pilot initiatives by uncovering unanticipated consequences and potential constraints.
- Local leadership, culture, stewardship, and organized action are vital to successfully implement high-leverage policies.
- All interventions have trade-offs; therefore, a suite of complementary strategies is often needed to sustain significant impacts over time.

**Case Study 9: Community Care of North Carolina**

Community Care of North Carolina (CCNC) is a statewide medical home and care management program designed to improve quality while lowering costs for the North Carolina Medicaid program. It currently provides a medical home to 1.1 million individuals in all 100 North Carolina counties; it further offers resources to 4,500 participating primary care providers to help them care for their Medicaid patients. These resources include partnerships with local community providers like health systems, hospitals, and health departments, and memberships in regional community care networks across the state. The regional networks provide care management support through local care managers, pharmacists, psychiatrists, and medical directors to improve local health care delivery (CCNC, 2012).

After the state has identified priorities, common core metrics are selected to track progress on those priorities, such as appropriate pharmacological therapy for asthmatics and blood pressure control for those with cardiovascular or ischemic vascular disease (CCNC, 2013). Community care networks have flexibility in how they proceed with their quality-improvement efforts but measure their performance in a consistent way to enable comparative performance monitoring against their peers. Those networks that are outperforming their peers then voluntarily share best-practice solutions so that effective interventions can be spread to other networks. At the same time, CCNC's Informatics Center sends the networks real-time data on hospitalizations, emergency room visits, and provider referrals, along with claims-based data on cost, utilization, and information on a provider's panel of patients (e.g., the conditions they have and the medications they take).

As a result of these efforts, CCNC has improved quality and cost for the Medicaid program. CCNC performs in the top 10 percent on common measures for diabetes, asthma, and heart disease compared with commercial managed care, as defined by Healthcare Effectiveness Data and Information Set measures. At the same time, an estimated \$700 million in savings has been achieved by CCNC for the state Medicaid program since 2006 (CCNC, 2012). CCNC is now attempting to scale up their program by expanding beyond the Medicaid population to demonstrate whether the model can produce similar results in the Medicare population and in the commercially insured population.

*Key Lessons Learned:*

- Clear goals with flexibility on means can promote innovation and help to identify best practices.
- A set of common, consistently defined performance measures can improve understanding of the goals of the initiative, demonstrate progress toward its aims, and build provider

confidence in the reliability and validity of comparative performance information. The use of consistently defined measures also enables rapid-cycle accurate comparisons to quickly identify top-performing interventions that are candidates for widespread dissemination.

- Building in feedback loops that allow health care professionals to learn about their performance relative to their peers and to share best practices is an effective way to improve quality and spread promising interventions.
- Pilots need to be supported with timely, user-friendly data feeds to continuously monitor their performance and be given the flexibility to change their course to improve. The traditional use of post-hoc evaluations at the conclusion of a project is not sufficient.
- Multipayer efforts are critical for demonstrating the ability for any intervention to produce system-wide improvements in cost and quality.

## SUMMARY

These case studies underscore the range of challenges currently facing pilot projects. The evidence produced by pilot projects often does not provide the information that decision makers need; the process can be slow; and pilots too rarely result in transformative change throughout the system. Yet, these cases are also a source of optimism that such challenges can be overcome. Several sites have successfully experimented with new models of evaluation that show how an intervention functions in a variety of real-world environments; other sites have developed new methods for scaling up and spreading successful initiatives; still others have demonstrated strategies for minimizing logistical and regulatory hurdles. Further progress will depend on embedding these concepts in the pilot project framework of public and private payers, regulators, delivery organizations, and other groups so that these become the norm, not the exception. Achieving this goal provides an important tool for tackling the key national health care challenges of cost, quality, and innovation—in short, for creating a health care system that learns.

## REFERENCES

- Alexander, J. A., and L. R. Herald. 2009. What can we learn from quality improvement research? A critical review of research methods. *Medical Care Research and Review* 66(3): 235-271.
- Allen, H., K. Baicker, A. Finkelstein, S. Taubman, B. J. Wright, and Oregon Health Study Group. 2010. What the Oregon Health Study can tell us about expanding Medicaid. *Health Affairs (Millwood)* 29(8):1498-1506.
- Baicker, K., and A. Finkelstein. 2011. The effects of Medicaid coverage—learning from the Oregon experiment. *New England Journal of Medicine* 365(8):683-685.
- Berwick, D. M. 2008. The science of improvement. *JAMA* 299(10):1182-1184.
- CCNC (Community Care of North Carolina). 2012. *Community Care of North Carolina: 2012 overview*. <https://www.communitycarenc.org/media/related-downloads/overview-of-ccnc.pptx> (accessed February 15, 2013).
- . 2013. *Performance measures: Metrics to stimulate quality improvement*. <https://www.communitycarenc.org/quality-improvement/performance-measures> (accessed February 15, 2013).
- Clark, S. L., D. R. Frye, J. A. Meyers, M. A. Belfort, G. A. Dildy, S. Kofford, J. Englebright, and J. A. Perlin. 2010. Reduction in elective delivery at <39 weeks of gestation: Comparative effectiveness of 3 approaches to change and the impact on neonatal intensive care admission and stillbirth. *American Journal of Obstetrics & Gynecology* 203(5):449e441-446.

- Clark, S. L., D. D. Miller, M. A. Belfort, G. A. Dildy, D. K. Frye, and J. A. Meyers. 2009. Neonatal and maternal outcomes associated with elective term delivery. *American Journal of Obstetrics & Gynecology* 200(2):15e151-154.
- Classen, D. C., R. Resar, F. Griffin, F. Federico, T. Frankel, N. Kimmel, J. C. Whittington, A. Frankel, A. Seger, and B. C. James. 2011. "Global trigger tool" shows that adverse events in hospitals may be ten times greater than previously measured. *Health Affairs* 30(4):581-589.
- Coburn, C. E. 2003. Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher* 32(6):3-12.
- Crabtree, B. F., S. M. Chase, C. G. Wise, G. D. Schiff, L. A. Schmidt, J. R. Goyzueta, R. A. Malouin, S. M. Payne, M. T. Quinn, P. A. Nutting, W. L. Miller, and C. R. Jaen. 2011. Evaluation of patient centered medical home practice transformation initiatives. *Medical Care* 49(1):10-16.
- Creswell, J. W., A. C. Klassen, V. L. P. Clark, and K. C. Smith. 2011. *Best practices for mixed methods research in the health sciences*. [http://obssr.od.nih.gov/mixed\\_methods\\_research](http://obssr.od.nih.gov/mixed_methods_research) (accessed October 30, 2012).
- Curran, G. M., M. Bauer, B. Mittman, J. M. Pyne, and C. Stetler. 2012. Effectiveness-implementation hybrid designs: Combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical Care* 50(3):217-226.
- Fairall, L., M. O. Bachmann, C. Lombard, V. Timmerman, K. Uebel, M. Zwarenstein, A. Boulle, D. Georgeu, C. J. Colvin, S. Lewin, G. Faris, R. Cornick, B. Draper, M. Tshabalala, E. Kotze, C. van Vuuren, D. Steyn, R. Chapman, and E. Bateman. 2012. Task shifting of antiretroviral treatment from doctors to primary-care nurses in South Africa (STRETCH): A pragmatic, parallel, cluster-randomised trial. *Lancet* 380(9845):889-898.
- Finkelstein, A., S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. P. Newhouse, H. Allen, and K. Baicker. 2011. *The Oregon health insurance experiment: Evidence from the first year*. Cambridge, MA: National Bureau of Economic Research.
- Fisher, E. S., M. B. McClellan, J. Bertko, S. M. Lieberman, J. J. Lee, J. L. Lewis, and J. S. Skinner. 2009. Fostering accountable health care: Moving forward in Medicare. *Health Affairs (Millwood)* 28(2):w219-w231.
- Georgeu, D., C. J. Colvin, S. Lewin, L. Fairall, M. O. Bachmann, K. Uebel, M. Zwarenstein, B. Draper, and E. D. Bateman. 2012. Implementing nurse-initiated and managed antiretroviral treatment (NIMART) in South Africa: A qualitative process evaluation of the stretch trial. *Implementation Science* 7:66.
- Glasgow, R. E., D. J. Magid, A. Beck, D. Ritzwoller, and P. A. Estabrooks. 2005. Practical clinical trials for translating research to practice: Design and measurement recommendations. *Medical Care* 43(6):551-557.
- Gold, M., D. Helms, and S. Guterman. 2011. Identifying, monitoring, and assessing promising innovations: Using evaluation to support rapid-cycle change. *Issue Brief (Commonwealth Fund)* 12:1-12.
- Greenhalgh, T., G. Robert, F. Macfarlane, P. Bate, and O. Kyriakidou. 2004. Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly* 82(4):581-629.
- HHS (U.S. Department of Health and Human Services). 2010. *Adverse events in hospitals: National incidence among medicare beneficiaries*. Washington, DC: HHS, Office of Inspector General.
- Hirsch, G., J. Homer, B. Milstein, L. Scherrer, C. Ingersoll, L. Landy, J. Sterman, and E. Fisher. 2012. *ReThink Health Dynamics: Understanding and influencing local health system change*. Paper read at 30th International Conference of the System Dynamics Society, St. Gallen, Switzerland.
- IOM (Institute of Medicine). 1999. *To err is human: Building a safer health system*. Washington, DC: National Academy Press.
- . 2002. *Unequal treatment: Confronting racial and ethnic disparities in health care*. Washington, DC: The National Academies Press.
- . 2012. *Best care at lower cost: The path to continuously learning health care in america*. Washington, DC: The National Academies Press.
- Landrigan, C. P., G. J. Parry, C. B. Bones, A. D. Hackbarth, D. A. Goldmann, and P. J. Sharek. 2010. Temporal trends in rates of patient harm resulting from medical care. *New England Journal of Medicine* 363(22):2124-2134.
- Magnan, S., E. Fisher, D. Kindig, G. Isham, D. Wood, M. Eustis, C. Backstrom, and S. Leitz. 2012. Achieving accountability for health and health care. *Minnesota Medicine* 95(11):37-39.
- Martin, A. B., D. Lassman, B. Washington, and A. Catlin. 2012. Growth in US health spending remained slow in 2010; health share of gross domestic product was unchanged from 2009. *Health Affairs (Millwood)* 31(1):208-219.
- McCannon, C., M. Schall, and R. Perla. 2008. *Planning for scale: A guide for designing large-scale improvement initiatives*. Cambridge, MA: Institute for Healthcare Improvement.

- McGlynn, E. A., S. M. Asch, J. Adams, J. Keeseey, J. Hicks, A. DeCristofaro, and E. A. Kerr. 2003. The quality of health care delivered to adults in the United States. *New England Journal of Medicine* 348(26):2635-2645.
- Milstein, B., J. Homer, P. Briss, D. Burton, and T. Pechacek. 2011. Why behavioral and environmental interventions are needed to improve health at lower cost. *Health Affairs (Millwood)* 30(5):823-832.
- Oregon Office for Health Policy and Research. 2009. *Trends in Oregon's healthcare market and the Oregon Health Plan: A report to the 75th Legislative Assembly*. [http://www.oregon.gov/oha/OHPR/RSCH/docs/Trends/2009\\_LegisTrendsReport.pdf](http://www.oregon.gov/oha/OHPR/RSCH/docs/Trends/2009_LegisTrendsReport.pdf) (accessed March 12, 2013).
- Pawson, R., and N. Tilley. 1997. *Realistic evaluation*. Thousand Oaks, CA: Sage.
- Platt, R., S. U. Takvorian, E. Septimus, J. Hickok, J. Moody, J. Perlin, J. A. Jernigan, K. Kleinman, and S. S. Huang. 2010. Cluster randomized trials in comparative effectiveness research: Randomizing hospitals to test methods for prevention of healthcare-associated infections. *Medical Care* 48(6 Suppl):S52-S57.
- Rosenthal, M. B., M. K. Abrams, A. Bitton, and Patient-Centered Medical Home Evaluators' Collaborative. 2012. *Recommended core measures for evaluating the patient-centered medical home: Cost, utilization, and clinical quality*. New York: The Commonwealth Fund.
- Sterman, J. D. 2006. Learning from evidence in a complex world. *American Journal of Public Health* 96(3):505-514.
- Suennen, L. 2011. *How doing a venture deal is like a relationship*. <http://www.slideshare.net/lisasuennen/how-venture-capital-is-like-a-relationship> (accessed July 19, 2012).
- Suennen, L., W. Rosenzweig, and C. Indig. 2011. Perspectives from the field. *Stanford Social Innovation Review* Fall(Suppl):10-11.
- Timbie, J. W., E. C. Schneider, K. Van Busum, and D. S. Fox. 2012. Five reasons that many comparative effectiveness studies fail to change patient care and clinical practice. *Health Affairs (Millwood)* 31(10):2168-2175.
- Tunis, S. R., D. B. Stryer, and C. M. Clancy. 2003. Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 290(12):1624-1632.
- Van Citters, A. D., and E. C. Nelson. 2011. *Quality and performance improvement among Quest charter member hospitals: A realist evaluation to understand factors that drive and restrain improvement*. Charlotte, NC: Premier.