

Explaining Michigan: Developing an Ex Post Theory of a Quality Improvement Program

MARY DIXON-WOODS, CHARLES L. BOSK, EMMA LOUISE AVELING, CHRISTINE A. GOESCHEL, and PETER J. PRONOVOST

University of Leicester; University of Pennsylvania; Johns Hopkins University

Context: Understanding how and why programs work—not simply whether they work—is crucial. Good theory is indispensable to advancing the science of improvement. We argue for the usefulness of ex post theorization of programs.

Methods: We propose an approach, located within the broad family of theory-oriented methods, for developing ex post theories of interventional programs. We use this approach to develop an ex post theory of the Michigan Intensive Care Unit (ICU) project, which attracted international attention by successfully reducing rates of central venous catheter bloodstream infections (CVC-BSIs). The procedure used to develop the ex post theory was (1) identify program leaders' initial theory of change and learning from running the program; (2) enhance this with new information in the form of theoretical contributions from social scientists; (3) synthesize prior and new information to produce an updated theory.

Findings: The Michigan project achieved its effects by (1) generating isomorphic pressures for ICUs to join the program and conform to its requirements; (2) creating a densely networked community with strong horizontal links that exerted normative pressures on members; (3) reframing CVC-BSIs as a social problem and addressing it through a professional movement combining “grassroots” features with a vertically integrating program structure; (4) using several interventions that functioned in different ways to shape a culture of

Address correspondence to: Mary Dixon-Woods, Department of Health Sciences, Adrian Building, University of Leicester, LE1 7AH, UK (email: md11@le.ac.uk).

commitment to doing better in practice; (5) harnessing data on infection rates as a disciplinary force; and (6) using “hard edges.”

Conclusions: Updating program theory in the light of experience from program implementation is essential to improving programs’ generalizability and transferability, although it is not a substitute for concurrent evaluative fieldwork. Future iterations of programs based on the Michigan project, and improvement science more generally, may benefit from the updated theory present here.

Keywords: Patient safety, quality improvement, evaluation science, program theory, health care–acquired infections.

Why We Need to Understand How and Why Programs Work

THE EVIDENCE-BASED PRACTICE MOVEMENT MEANS THAT THE need to assess the effectiveness of health care interventions is not in dispute. Recognizing the importance of understanding how and why interventions work has come more slowly, but without a clear account of what an interventional program comprises, how its activities are linked to its outcomes, and how context and program interact, its operation remains a black box (Stame 2004). Expanding successful first iterations of programs without understanding the social processes and mechanisms that produced the outcomes means that other implementers of the programs will not know what they need to do to make the program work and where they should direct their efforts and resources. Then, when the program does not achieve the desired results, it is difficult to know whether this was due to faulty theory (the wrong thing was done), flawed implementation (the correct thing was done, but in the wrong way), or some combination of both.

This is an especially important problem in the emerging field of quality improvement (QI) in health care. The impact of QI interventions tends to be startlingly variable across apparently similar settings, and the positive effects of early iterations of interventions are often difficult to replicate (Lomas 2005). Four of the explanations for this phenomenon are as follows:

1. The intervention did not work in the first place (the reported improvement was really caused by something else).
2. The intervention in (at least some) new contexts does not properly replicate the original intervention.
3. Contextual effects mean that the intervention cannot succeed consistently.
4. Some combination of 2 and 3 may be responsible, as the two may interact.

In this article, we are especially interested in the second explanation. QI studies often are remarkably poor at describing exactly what a program comprises, and particularly at describing the program and its activities in a way that they can easily be reproduced (Shojania and Grimshaw 2005). Such studies often are equally poor at describing the theoretical basis of their interventions (e.g., what is the means by which this intervention might reasonably be expected to achieve the hoped-for effects?). This failure to produce good accounts of what the program involved (what activities were undertaken) and to explain how the program achieved its effects (what mechanisms were at work) leads to a problem that might be termed “cargo cult quality improvement.” Cargo cult science was famously described by Richard Feynman in a 1974 commencement address at Caltech:

In the South Seas there is a Cargo Cult of people. During the war they saw airplanes land with lots of good materials, and they want the same thing to happen now. So they've arranged to make things like runways, to put fires along the runways, to make a wooden hut for a man to sit in, with two wooden pieces on his head like headphones and bars of bamboo sticking out like antennas—he's the controller—and they wait for airplanes to land. They're doing everything right. The form is perfect. It looks exactly the way it looked before. But it doesn't work. No airplanes land. So I call these things Cargo Cult Science, because they follow all the apparent precepts and forms of scientific investigation, but they're missing something essential, because the planes don't land. (Feynman 1999, 208)

When QI initiatives are implemented without a proper understanding of what they involve and how they work, they similarly risk becoming distorted imitations that succeed only in reproducing the superficial outer appearance but not the mechanisms (or set of mechanisms) that

produced the outcomes in the first instance. In new contexts, not having a well-explicated program theory risks the program's failure.

What do we mean by "program theory"? Despite the ubiquity of terms like *theory*, *program theory*, and *theory of change*, there is little consensus on how they should be defined (Donaldson 2007). Here, we see a program theory as an explanation of why the effects observed in a program are likely to have occurred. Because it is specific to a particular program and operates in relation to the program as whole, program theory is distinct from other types of theory, such as those operating at a higher level of generality (e.g., theories of behavior change) or those explaining the science behind individual components (e.g., why chlorhexidine is an effective infection control measure).

A big challenge for those trying to replicate a QI program is that the theory explaining why a program worked (or did not work) may not be discovered by inspecting the program's protocols or reading the formal, protocol-driven study reports. Programs almost never proceed as planned. Instead, they, and the assumptions that guide them and shape their actions, usually change over time as the programs progress. For instance, program leaders must respond to human agency and institutional contexts that affect the feasibility of their plans, and they need to figure out ways of doing things that they may not have thought of at the start. Ideally, the evolution of the program and the lessons learned are captured through evaluative fieldwork conducted in real time as the program is carried out. With some exceptions (e.g., Benning et al. 2011), such studies remain rare in QI, but the need to both describe the program and understand how it works remains just as pressing, particularly when apparently successful programs are rolled out in new contexts.

In this article, we first outline an approach, located within the broad family of theory-oriented evaluation methods, for developing ex post theory by using program leaders' experience and social scientists' input. Second, we describe an application of this approach to the Michigan Keystone Project, which reported a dramatic average reduction in central venous catheter bloodstream infections (CVC-BSIs) in more than one hundred participating intensive care units (ICUs) (Pronovost et al. 2006).

The Michigan project is an especially important candidate for ex post theory building for several reasons. It is one of the few QI programs that had impressive results; it sustained its effects and showed that these were associated with reduced mortality; and it attracted immense

international interest. Encouraged by the World Health Organization (WHO), ICUs around the world are trying to “match Michigan” by adopting and implementing this program. At the same time, however, popular accounts of the program have often been simplistic and partial, and have perpetuated the myth that the program’s achievements can be traced to a “simple checklist” rather than a complex social intervention (Bosk et al. 2009). Such accounts risk causing the cargo cult problem by misleading those seeking to implement the program in new contexts into thinking that all that they need to do is introduce a checklist. A more sophisticated theory of how the program really worked may help avoid these false assumptions, prevent mistakes in implementing the program, and improve the chances of patients’ benefiting from it.

Theory-Oriented Evaluation

One way to develop an ex post theory is to try to examine the available evidence and to speculate, in a relatively free-form way, about what might have produced the results. However, one of the achievements of the family of evaluation methods that we call *theory oriented* (including approaches such as theory-based evaluation, theory-driven evaluation, and realist evaluation) has been to demonstrate the value of using program theory to guide questions and data gathering in evaluation studies. This family of methods insists on explicating the theories or models that underlie programs, elaborating causal chains and mechanisms, and conceptualizing the social processes implicated in the programs’ outcomes. The evaluation scientist Carol Weiss, for example, emphasizes the importance of identifying “program theories of change,” by which she means the rationale and assumptions about mechanisms that link programs’ processes and inputs to outcomes (both intended and unintended). Weiss advises program designers to clearly specify the hypotheses and assumptions that inform programs, especially those concerning how the program is likely to bring about the desired outcomes (Weiss 1995). She—along with others, including Chen and Rossi (1983) and Connell and Kubisch (1998)—encourages program designers and evaluators to specify a program’s theory of change before implementing and evaluating the program, by using both the social science literature and the program stakeholders’ beliefs. Although some of its details differ from those of other approaches in the theory-oriented evaluation family,

realistic (later known as *realist*) *evaluation* (Pawson and Tilley 1997) also seeks to test and refine theories, particularly those that might explain how context, mechanism, and outcome are linked.

Theory-oriented evaluation methods demonstrate that program theories can be a useful starting point for evaluation. But most theory-oriented evaluation approaches are intended for prospectively designed evaluation, and they typically involve collecting data as programs proceed. As we noted earlier, however, QI intervention studies are often undertaken without a concurrent process evaluation tracking the way the program works as it is implemented. This is a serious problem. By the end of the program, the designers' operating theory may look quite different from the theory with which they started. Yet the sparse description of interventions encouraged by academic medical journals may provide very few opportunities for program leaders to discuss their new understandings of how the program worked in practice. These opportunities may be limited even further if one group designs and implements the program and another group evaluates it. Accordingly, we propose that when no fieldwork data are available, the learning and experience of program leaders constitute an important source of data for both describing the program interventions (what really happened, as opposed to what was intended to happen) and refining and improving the program theories.

Of course, the program leaders themselves may not have complete insight into exactly what happened or may not be able to recognize and theorize what occurred and why. In part this is because some of the program designers' (and implementers') decisions may be recognized as decisions only in retrospect. That is, the decisions may have been a series of small steps or disjointed and amorphous accommodations that resulted in lines of action taking shape (Weiss 1988). In part, it is because interventions often are complex and difficult to describe or explain, and since those conducting QI often are not familiar with social science theory, they may not be able to explain the intervention fully. Social scientists, therefore, may be a valuable source of information about ex post program theory, as they can explain aspects of the program that may have been obscure to its developers and implementers, can help hunches mature into theories, and can challenge assumptions.

We summarize our proposal as follows. Updating program theory after the program has been carried out is critically important, and combining the program leaders' experience of implementing the program and the expertise of social scientists may offer a valuable way of achieving

this. An analogy—and that is all it is—might be the Bayesian approach to probability. Bayesian statistics involves the construction of formal statistical models; that is not what we are proposing here. But the principles that inform Bayesian analyses can help structure thinking about how to devise an ex post theory of a program. Bayes's theorem allows conditional probabilities to be calculated by updating initial sources of information, formally expressed as a probability distribution (the “prior” distribution) with new evidence (the “likelihood”) to produce a “posterior” distribution (Roberts et al. 2002; Spiegelhalter et al. 2000). It can thus be regarded as “a formalization of the process of learning from experience” (Spiegelhalter, Abrams, and Myles 2004, 2). The analogue we propose is that prospectively defined program theory and the program leaders' experience can be regarded as “the prior” and can be synthesized with the contribution of social sciences to produce a new ex post theory that can be used and tested in future implementations of the program.

The Michigan program is an example of an initial program theory that is particularly likely to benefit from updating in this way. Even though it did start with an explicit framework for change described in the protocol, the program designers learned much from the program's implementation and used that knowledge dynamically to modify the program to respond to the participants' needs during its implementation. Although the program in action was not identical to the program as designed, limited resources made a process evaluation impossible to run alongside the program.

The Michigan Keystone Project to Reduce Central Venous Catheter Bloodstream Infections (CVC-BSIs) in Intensive Care Units

The purpose of the Michigan Keystone Project (hereafter called the Michigan Program or Michigan) was to reduce central venous catheter bloodstream infections (CVC-BSIs) in more than one hundred participating intensive care units (ICUs) in the state of Michigan. CVCs—also known as *central lines*—are narrow tubes inserted into large veins, with the tip lying close to the heart. CVCs allow vascular access for a variety of clinical purposes, including administering drugs and fluids. But they also may let dangerous bacteria or fungi enter directly into the patient's bloodstream. CVC-BSIs are a major cause of increased morbidity,

mortality, and costs of care in ICU patients (Mermel et al. 2009). The most important contributor to CVC-BSIs is the way that the catheter is inserted and subsequently managed (Gastmeier and Geffers 2006). The original program theory for Michigan (summarized briefly in box 1) was intended to address this both through interventions acting specifically on practices relating to CVC care and through promoting cultural and organizational changes pertaining to safety more broadly.

Box 1

Summary of Original Program Theory for Michigan

The study's conceptual model was based on the Medical Outcomes Study and sought to link care processes to a diverse set of outcomes (see figure A). The specific aims of this project included the following:

1. Implementing and evaluating the impact of the Comprehensive Unit-Based Safety Program (CUSP), which includes the ICU Safety-Reporting System (ICUSRS) in a cohort of Michigan hospitals.

Hypothesis: The CUSP with the ICUSRS will lead to measurable improvements in patients' safety and safety climate.

2. Implementing and evaluating the effect of an intervention to improve communication and staffing in ICUs.

Hypothesis: The use of these targeted interventions will lead to significant improvements in the ICU's quality of care.

3. Evaluating and implementing in a cohort of Michigan ICUs the effect of an intervention to reduce or eliminate catheter-related bloodstream infections in ICUs.

Hypothesis: With this intervention we will eliminate or reduce catheter-related bloodstream infections in Michigan ICUs.

4. Evaluating and implementing in a cohort of Michigan ICUs the effect of an intervention to improve the care of ventilated patients in ICUs.

Hypothesis: With this intervention we will eliminate or reduce ventilator-associated pneumonia, duration of mechanical ventilation, and ICU length of stay.

5. Implementing and evaluating an intervention to reduce ICU mortality.
Hypothesis: We can reduce ICU mortality with this focused intervention.
6. Identifying the characteristics associated with improvements in patients' outcomes.
Hypothesis: Leadership involvement, dedicated staff time, and a physician advocate will be associated with improvement efforts.

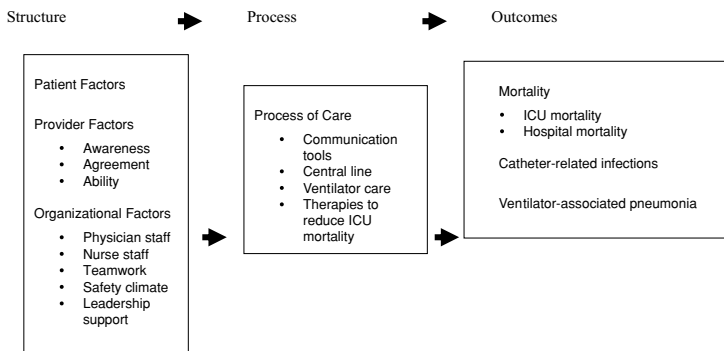


FIGURE A. Conceptual Model for Michigan Study: Medical Outcomes Study Framework

The program reported considerable success in achieving its objectives: the median reported infection rate per 1,000 catheter days dropped from 2.7 at baseline to 0 within three months and stayed at that level until the end of the eighteen-month study. The mean rate of infections decreased from 7.7 BSIs per 1,000 catheter days at baseline to 2.3 at three months and 1.4 at eighteen months (Pronovost et al. 2006). These reductions were sustained over another eighteen months of follow-up (Pronovost et al. 2010). Since the study was based on hospitals' self-reports, we cannot completely exclude "gaming" effects (Bevan and Hood 2006) in the data. But data showing that mortality in the ICUs participating in the program decreased compared with a control group of nonparticipants (Lipitz-Snyderman et al. 2011) offer reassurance that the effects observed were real.

Our analysis asked why the program was successful. Our procedures for updating the program theory were as follows:

1. The program leaders' (Goeschel and Pronovost) theory of change, based on their original study protocol and their experience from

running the program, was treated as an analogue of the Bayesian prior. This “implementer” team gave their original study protocol to the “evaluator” team of social scientists and participated in interviews and other dialogues with them in order to share learning.

2. New information—an analogue of the Bayesian likelihood—came from the “evaluator” team of social scientists (Dixon-Woods, Bosk, and Aveling), who produced a theoretical interpretation of both the original program theory and the accounts of the program given by the program leaders in their interviews, dialogues, and published reports. Their attempts to theorize did not draw on other reports of programs to reduce infections but, instead, on more general social science theory. This new information can be likened to the Bayesian likelihood.
3. Through a collaborative process of dialogue, challenge, and eventual consensus, the “prior” theory of the implementer team was synthesized with the new evidence from the evaluator team to produce an updated posterior or “ex post” theory of why the Michigan project was so successful. This new program theory helped identify the mechanisms that explained the program outcomes and how those changes were achieved.

Our overall aim was to explain how the program worked on average across the participating units. We did not try to describe the contextual factors that might have modified the effectiveness of the program in different settings.

An Ex Post Theoretical Account

Our synthesis of the original program specification, the program team’s experience, and the social scientists’ expertise identified six reasons that explained why Michigan worked: (1) isomorphic pressures, (2) networked community effects, (3) reframing CVC-BSIs as a social problem, (4) changing practice and culture at the sharp end by using interventions with different effects, (5) using data as a disciplinary force, and (6) skillfully using “hard edges.”

1. *Isomorphic Pressures Explain Why Units Joined the Program.* Funded by the U.S. Agency for Healthcare Research and Quality as a prospective cohort study, the Michigan ICU project tried to assess and improve

the culture of patient safety, standardize the use of evidence-based interventions, and reduce CVC-BSIs and ventilator-acquired pneumonia (VAP). The collaborative model for quality improvement in health care (Mills and Weeks 2004; Øvretveit et al. 2002) was used for the program's design, which eventually involved 103 ICUs across Michigan, all of which agreed voluntarily to commit to the initiative. The project was a collaborative of the Quality and Safety Research Group at Johns Hopkins University, the Michigan Health and Hospitals Association (MHA)-Keystone Center for Patient Safety and Quality, and participating Michigan hospitals. As the original protocol stated, the participating hospitals had to make a number of explicit, though nonbinding, commitments (see box 2) in writing and signed by a hospital senior executive, along with a list of hospital team members and the amount of time they would devote to the project.

Box 2

Unit Requirements for Participants

- Provide a commitment letter from the hospital CEO to the program team.
- Identify a project team leader, typically a nurse manager, who can devote approximately 10 percent time to this effort.
- Form a project team that includes, at a minimum, a physician advocate, a nurse manager/advocate if not the project leader, a data coordinator, and a hospital executive advocate.
- Submit baseline and monthly infection rate data to the program; blinded data are provided to all participants for benchmarking purposes.
- Complete a cultural survey at outset and approximately eighteen months later.
- Participate in weekly immersion calls during the first six weeks.
- Participate in one or two project conference calls a month on content, coaching, and peer learning.
- Participate in statewide face-to-face meetings every six months over the course of the program.
- Implement the program's improvement tools, including CUSP.
- Hold monthly meetings to review data results and apply the improvement tools.

In answer to an invitation published in the Michigan Hospital Association's newsletter, forty ICUs initially signed up for the program. Then, as word got around, other ICUs requested participation, and in the end, 85 percent of Michigan's ICU beds were included. The large number of participants, which increased as time went on, was evidence of a process known in sociology as *institutional isomorphism*. Institutional isomorphism refers to how organizations come to look like one another, for example, by adopting similar processes, practices, and structures (DiMaggio and Powell 1983). This kind of similarity of form is often found in organizations facing similar problems. The Michigan ICUs were a community of organizations engaged in the same kinds of activities in the same area of health care and were subject to similar reputational and regulatory pressures. Accordingly, they were naturally inclined to compare their procedures and structures, and so we might expect that over time the organizations would mimic the more successful ones.

The important practical effect of institutional isomorphism is that organizations may find that it is unacceptable and damaging *not* to participate in a particular program or to adopt particular policies (Sutton et al. 1994). Not participating in the Michigan program thus may not have been an option for most ICUs, especially once the tipping point of participation had been reached. Three different mechanisms may produce institutional isomorphism (Powell and DiMaggio 2001). Organizations that conform because they perceive something to be the "right" way of doing something in terms of values and ethics exhibit *normative isomorphism*. The external environment's imposition of requirements (such as rules and regulations) that are difficult for organizations to escape with impunity is *coercive isomorphism*. Organizations that imitate one another (especially those seen to be successful) with a view to improving their own performance, securing legitimacy, or both are demonstrating *mimetic isomorphism*. Although all three forces were at work in Michigan, normative and mimetic isomorphism probably were the strongest. Coercive forces were not entirely absent, as we suggest later, but they did not explain why ICUs joined the program when they were not required to do so.

2. *Creating a Networked Community.* Once they had joined the project, the ICUs were asked to collect baseline CVC-BSI and VAP data and to administer to their staff a safety attitudes survey, as specified in the original protocol. The protocol also stipulated participation in education and meetings, which took more specific shape as the program

proceeded. Education involved “immersion” coaching, including weekly teleconferences for six to eight weeks with hundreds of callers from the participating hospitals. The coaching gave the participants content information on measurement and the CUSP program (see box 3) and explained the technical interventions. After five months, a residential two-day workshop was held to build rapport within teams and between participating teams and the project team, and to launch the technical interventions. The ICUs then began to submit monthly data to the project team and to implement the technical interventions. The monthly teleconferences and face-to-face workshops at six-month intervals continued for the duration of the program (see box 4), but again, their form evolved over time. These calls delivered the information necessary to carry out the project; provided mentorship on methods for facilitating local change, including identifying and resolving common barriers through theoretical and experience-based strategies; engaged the project team with the participating centers and the centers with one another; and offered encouragement by sharing success stories. The teleconferences were recorded and made available on CDs, and a toolkit of materials to support the implementation of the intervention was distributed.

Box 3

Comprehensive Unit-Based Safety Program

- Educate staff on the science of safety.
- Identify defects in care.
- Engage executives.
- Learn from one defect per month and implement one culture improvement tool.
- Implement teamwork tools.

Box 4***What Participating Hospital Units Received***

- CUSP and CVC-BSI reduction tools and training.
- Support and access to expert faculty through teleconferences each month and access to expertise between calls.
- Opportunities to gather in person at conferences with other members of the collaboration.
- Tools and training for measuring and monitoring CVC-BSIs and safety culture in units.
- Dedicated website with resources, manuals, toolkits, recordings of learning sessions, central line FAQs, notifications of training opportunities, links to other useful websites, and the like.
- Anonymous trend data fed back to units.

The teleconferences, meetings, and other communicative facilities served a number of important functions beyond supplying information. They helped perhaps most in promoting a networked, community-based approach to the problem of CVC-BSIs. Although the original protocol had identified the importance of the collaborative approach, it was the experience of running the program together with the later input of the social scientists that enabled recognition of the importance of the networked community to the success of the Michigan project. The potential of such networks in securing desirable outcomes in health care and elsewhere has drawn increased interest and excitement in recent years (Braithwaite 2008; Buchanan et al. 2007; Dopson et al. 2002; Kickert, Klijn, and Koppenjan 1997; Martin, Currie, and Finn 2009a). Approaches based on professional communities are thought to be especially promising for addressing well-known problems of changing professionals' practices, as they may prefer to take their "directions for performance" from inside rather than outside their professional group (Bate 1994).

Health care professionals are, of course, not alone in this, as rules imposed from outside a community typically have less legitimacy and

are more likely to be violated than those agreed upon within a community (Ostrom 1990). The resistance of physicians in particular to efforts to impose “outside” or “managerial” rules has been well documented (Black and Thompson 1993; Davies, Powell, and Rushmer 2007; Martin, Currie, and Finn 2009b; Waring and Currie 2009). This has been found to be true even when managers who are themselves physicians are acting in administrative roles, because the imposition of rules violates norms of collegiality. Using a professional community as a means of mobilizing action, by contrast, is consistent with long-standing collegial structures and with the corresponding centrality of the community principle in the organization and experience of professional work (Adler, Kwon, and Heckscher 2008). As well as helping avoid professional resistance, networked, community-based participatory approaches, which emphasize cooperation and norms of reciprocity rather than administrative fiat and managerial instructions, are more likely to sustain collaboration and activity over the long term (Lowndes and Skelcher 1998; Bray et al. 2009) and to create programs deemed useful and relevant by clinicians.

The “sense of community” in Michigan was fostered by opportunities for interaction and communication. Project workshops, which were attended by as many as five hundred people, were residential. A cocktail hour and a networking section, neither of which was suggested in the original protocol, became features of these workshops over time. Another later innovation was that at the second and subsequent workshops, participants were given a project token (such as a wristband, a mirror, specially labeled soda, or a project T-shirt listing all the participating hospitals; the letterhead of the project’s notepaper contained the same list). Once the initial “coaching” on content had been completed, teams from the participating ICUs increasingly led most of the agenda in teleconferences and meetings. The teams were encouraged to present success stories and to openly discuss any problems and ways to overcome them. This included, for example, dealing with colleagues who did not want to follow the program’s principles. The project leaders acted as facilitators and moderators in these interactions, focused on making generalizable points and encouraging the teams to learn from one another. Again, none of these features was explicitly described in the protocol but emerged as the program proceeded.

Through these strategies to encourage horizontal relationships across and within hospitals, the members of the Michigan program evolved into a virtual learning community, and they frequently talked with one

another outside the formal structure. Again, even though it was not part of the original plan, these exchanges were important to its success. Earlier work has identified the significance of informal socializing in facilitating the spread of “know-how,” the development of relationships, and the establishment of a “sense of community” (Finn, Currie, and Martin 2010; McMillan and Chavis 1986; Øvretveit et al. 2002). People with a greater sense of community and shared identity are more likely to offer their own time and resources to pursue the collective interest (Gillespie et al. 2008). The sense of community in the Michigan program was critical to the generation of reciprocal ties, shared commitment to group goals, and the sense of having a personal investment (McMillan and Chavis 1986). The program thus had some of the features of a “grassroots” or “bottom-up” movement.

By creating social obligations among the members of the collaborative community, the program offered opportunities for social control. In a voluntary endeavor like Michigan, social relations among members are the key to securing compliance with the rules and agreements of the community, since its members, rather than any external agency, monitor one another (Lazega 2000). Maintaining the “regard” of peers within a community requires conformity to its norms; people have an incentive to cooperate so as not to lose respect and standing (Offer 1997). Frequent interactions, reciprocal communication, and decentralization increase the social pressure to cooperate by reducing the social distance between members of unequal status and authority and strengthening the acceptance of group norms by the group members. All these were important to securing the members’ commitment to the Michigan program, as conforming to its goals and requirements came to appear “more modern, appropriate or rational” (Scott 1991) than nonconformity.

Some caution about the notion of a participatory community is needed, however. Although the concept of community often is associated with notions of harmony, egalitarianism, and consensual relationships, communities can be fragmented, hierarchical, and involve conflict (Cornish and Ghosh 2007; Gujit and Shah 1998). In communities, consensus coexists with conflict. Empirical studies of community-based approaches to change are replete with examples of the perpetuation of power inequalities between (and within) participating groups (Clever 2001; Kothari 2001). In such cases, the voices and interests of certain groups may be marginalized despite rhetorical commitments to participation, partnership, or equality (Aveling 2010; Finn 2008; Wilkinson, Godrey,

and Marchington 1997). This suggests that rather than being left to self-organize (and possibly to reproduce historically established hierarchies and asymmetries), the organization of a quality improvement community is more likely to succeed if it is not entirely “bottom-up.” Strong internal direction and “top-down” leadership that empowers all stakeholders to participate (Cornish 2006) thus are important. Quality improvement communities should combine horizontal or “grassroots” momentum with a vertical integrating structure that can coordinate activity and manage potentially competing interests and motives.

The Michigan project sought to do this in a number of ways, in particular by deliberately trying to include all relevant stakeholder groups. The program targeted three groups: senior leaders in hospitals, team leaders (middle managers), and the staff of ICUs. Hospital chief executives specifically were encouraged to support the program. Local improvement teams in the participating units were designed to not be dominated by any single profession but to have representatives from all stakeholder groups. In addition, program leaders asked team leaders to make sure that staff understood the purpose of the project, what they could do to help, what ideas they had to improve the project, and how they were progressing toward the project’s goals. Project workshops, teleconferences, and other communications were conducted to ensure that all voices would be heard. The program leaders repeatedly emphasized that they did not have all the answers and told the teams that they themselves knew best what would work in their local contexts. Again, the theoretical salience of the community approach, and what was required to optimize it, became evident only as the program theory was being updated in collaboration with the social scientists.

3. *Reframing CVC-BSIs as a Social Problem with a Solution.* One of the most important challenges facing the Michigan program was the need to redefine CVC-BSIs as a social problem that could be solved. By social problem, we mean one involving human action and behavior, not a problem with a simple technical fix. In health care, quality issues often are not perceived as problems by the clinical group targeted by an intervention. In addition, social ills that are seen either as intractable or as an expected outcome of work are tolerated because concerted action is seen as futile and a waste of resources that should be used for less stubborn problems. The first task is persuading people of the existence of a shared problem around which they should organize, that is, a problem that can be fixed. Relevant factors include how obvious the problem is, how

important it is to potential participants, and how amenable it is to resolution (Øvretveit et al. 2002).

In Michigan, achieving the community's consensus on CVC-BSIs as a social problem required (1) disrupting norms and behaviors that treated CVC-BSIs as inevitable and (2) developing a set of standardized interventions that the community would accept and implement. The process of displacing existing norms is, however, often uncomfortable and delicate, requiring multiple negotiations and realignments of identities, understanding of roles, and interests. The Michigan program achieved this by creating something like a professional movement (Bucher 1962) that combined some (though not all) of the characteristics and tactics of a social movement (Bate, Robert, and Bevan 2004) but retained the structure of a formal, federally funded intervention. The idea of a professional movement was not part of the program at the outset but gradually emerged over time and was recognized only later by the social scientists.

Another important task of the Michigan professional movement was to engage in "meaning work" (Benford and Hunt 2003; Benford and Snow 2000). That is, it had to frame infection control related to CVCs as a social problem that could be resolved through strategic action. Efforts to reframe an issue as a social problem frequently lead to resistance, in which the protagonists and antagonists insist on their definitions of the situation and try to discredit their opponents' claims or stand fast and refuse to budge (Benford and Hunt 2003; Hilgartner and Bosk 1988). Denying the existence of a problem or denying injury is a key strategy of movement antagonists, which allows them to claim that their opponents are ill informed, irrational, or insincere (Benford and Hunt 2003).

In Michigan, the different groups' perceptions of the extent to which CVC-BSIs were an authentic and soluble problem varied: the infection control professionals saw the CVC-BSIs as a problem, whereas the clinicians working in ICUs were initially more likely to downplay infections' significance and susceptibility to amelioration. The Michigan program used two strategies, storytelling and "hard data," to overcome resistance and transform into a social problem the perception of CVC-BSIs as a normal occurrence. Both strategies were identified in the project protocol but were not at that time linked to the concept of a social problem; this was an understanding that emerged later when updating the program theory.

Social problems, by definition, cause avoidable harm to someone, and the task of constructing a victim group (usually one that is "innocent")

is therefore a key to creating a successful social movement (Dunn 2005). At the program workshops, the storytelling included the tragedy of eighteen-month-old Josie King, who died following a series of hospital errors that began with a catheter infection. By dramatizing and personalizing the story of a single individual and using visual images of preventable disaster, her story triggered the same generosity of response as found when individualized victims are used to solicit charitable donations (Ariely and Norton 2009). By presenting the harm done to patients as clearly avoidable, storytelling succeeded in undermining health care institutions' claims to be safe places and reinforced the perception that something had to be done. "Hard data" demonstrating the variable rates of infection across different ICUs further discredited the "safe places" claims and heightened the drama of the stories but, at the same time, showed that improvement was possible. Together, the stories and the data were discomfiting, challenging the view that health care institutions in general and ICUs in particular were already doing the best they could and could not do better. Identifying the innocent victims of infections helped construct an uncomfortable identity for ICU staff as the accidental perpetrators. At the same time, however, the possibility of redemption through participation in the program completed the narrative arc.

One of Michigan's key achievements was its success in creating a "cultural frame": a representation "of collective problems and solutions that help other actors link their own interests and identities to a collective purpose" (Stone Sweet, Fligstein, and Sandholtz 2001, 8–9). The program was able to position the elimination of CVC-BSIs as an obtainable collective prize, motivated by a collective conscience. Nonetheless, finding a consensus on how to win this prize proved to be a challenge as well. Within professional networks, different stakeholders have different knowledge and expertise, and plans for actions are often highly contested. This was addressed in several ways.

The project team's "vertical core" undertook the scientific review of evidence for the interventions used in the program. The project leaders secured the legitimacy of and consensus on its key interventions (e.g., hand hygiene, use of chlorhexidine to prepare patients' skin before CVC insertion) by presenting them as based on clear and credible scientific evidence. This meant that the program could focus on an identifiable, well-bounded, and solvable problem. This single social fact helped diminish arguments about the usefulness of the interventions.

But the credibility and legitimacy of the evidence and the proposed action had to be established through social processes (Latour 1987). The importance of the source in constructing the authority, credibility, and persuasiveness of interventions is now well recognized in the social science literature. Clinicians' behaviors are influenced not only by abstract knowledge and formal bodies of evidence but also by trusted peers (Dopson et al. 2003). Leaders with the necessary authority to "breathe legitimacy" (Hwang and Powell 2005) into what is being advocated are therefore critical. This reflects a more general finding that people are more likely to change their behaviors when they see liked and trusted peers doing the same (Dube and Wilson 1996). The authority of evidence does not stand on its own but requires support from the moral authority of those seeking to deploy it. Critical to the Michigan program's ability to create a professional community focused on reducing CVC-BSIs and to secure legitimacy within the community was its use of leaders who were ICU "insiders" with whom members of the community could identify. It also helped that one of the program leaders (Goeschel) had dual legitimacy as a critical care clinician and a respected health system executive in the state's hospital association. Although the study protocol had listed the project team's strengths, they were made part of the theory only after discussions with the social scientists.

Consensus and legitimacy within the community were further enhanced through collective processes of critical reflection and discussion (Campbell and MacPhail 2002; Guareschi and Jovchelovitch 2004). Repeatedly bringing together large numbers of people from many different organizations was important to creating a sense of community and also to building a consensus on how to reduce CVC-BSIs. Through engagement and discussion, members of the program were able to negotiate, and renegotiate, and it was "this process of constructing meaning which provides organizational members with identity and cohesiveness" (Bate and Robert 2002, 654). Face-to-face and group communication by telephone allowed the sincerity, legitimacy, comprehensibility, and accuracy of the program leaders to be tested, and the stakeholders with opposing views to be questioned, which is recognized as important to consensus building (Innes and Booher 1999). In this way, the program deflated the classic "counterframing" strategy (Benford and Hunt 2003) used by antagonists, in which they argue that the solutions to a problem should be different from those recommended by the program's leaders.

Notably, the original study protocol indicated use of Plan-Do-Study-Act cycles but they were not promoted during the program. Instead, the program leaders spelled out which elements of the program must be used and where local variations were both possible and desirable, but they did not state what these variations should be. Thus the program's key elements were presented as "essential ingredients" that should be made part of local recipes. The program did not, for example, prescribe or impose a uniform format for the checklist for CVC insertion and management, but instead encouraged ICUs to devise their own, as long as they retained the main principles. Empirical studies support this approach of standardization with local variations (Bate, Robert, and McLeod 2002; Bosk 2003; Dopson et al. 2002; Øvretveit et al. 2002). This approach is also consistent with social science theory indicating that some level of local autonomy is much more likely to increase people's willingness to accept standardization and tight control (Poteete 2010). This quality of dynamic standardization probably enhanced the sense of ownership and commitment felt by members of the clinical community.

4. *Changing Practice and Culture at the Sharp End through Interventions That Functioned in Different Ways.* Several features of the Michigan program were simultaneously directed toward the achievement of a defined goal and also served social functions. For example, in accordance with the project protocol, the ICUs were asked to create a dedicated trolley/cart for CVC insertion that would contain in one place all the items required for a successful line insertion. This was "instrumental" in that having everything available to the operator reduced the risk of infection because it averted delays and encouraged aseptic techniques. But providing and stocking the trolley/cart also had an important expressive function that we recognized when we updated the program theory: it signaled that logistical, operational, administrative, and financial resources had been mobilized to serve the needs of infection control. Similarly, such aspects of the CUSP as setting daily goals, partnering the unit with a senior executive, and training on safety science all marked an organizational commitment to patients' safety (Dixon-Woods 2010) and heightened both the awareness and the organizational priority of the program.

As specified in the project protocol, the Michigan program asked that participating units devise and use a checklist of the five practices related to CVC insertion and management (see box 5) that were known to reduce the risk of infection. It further asked that every CVC

insertion (a procedure carried out by a doctor) be witnessed by a nurse who would record the doctor's compliance with the checklist requirements, and stop the insertion if the checklist was not followed. At its simplest, the checklist was a prompt to repair any lapses in the five steps. But confining an understanding of how the checklist works to this simple, though important, role risks a serious misunderstanding. Our updating of the program theory enabled us to recognize that the checklist's social functions extended to a greater contribution to the process in which the actions prescribed by the checklist were accepted and "taken for granted" as the right thing to do: the checklist functioned as a device for social control. It made visible the discrepancies between actual practice and "ideal practice," directing attention to those areas of practice that required attention, and changed the role obligations of the nurse. Furthermore, the use of a locally adapted checklist supported by a strong evidence base provided a "safe space" for ICU workers to develop and practice team behaviors and accept collective responsibility for their outcomes. Because the evidence base supporting the intervention was very strong, the nurses felt safe in speaking up, thereby reducing the risk that they would be criticized.

Box 5

CVC-BSI Elimination Intervention

- Educate staff on evidence-based practices to reduce CVC-BSIs, and implement a checklist to ensure compliance with the following:
 - Observing appropriate hand hygiene.*
 - Using chlorhexidine for skin preparation.*
 - Using full-barrier precautions during central venous catheter insertion.*
 - Choosing subclavian vein placement as the preferred site.*
 - Removing unnecessary CVCs.*
- Empower nurses to ensure that doctors comply with the checklist.
- Provide feedback on infection rates to hospitals and at the unit level.
- Implement monthly team meetings to assess progress.

The way the checklist functioned socially addressed known problems in infection control. Preventing infections requires sustained, collective, and fastidious attention by all members of teams caring for patients but is inherently susceptible to the problem of diffusion of responsibility, perhaps most commonly among those whose job is not explicitly identified as infection control. There are two reasons for this. First, individuals may feel that no matter how much they try to prevent infection, their efforts will be undermined by others who fail to make the same commitment; just one careless handling of a catheter can result in infection. Second, hospitals are complex organizations characterized by “the problem of many hands” (Bovens 1999; Thompson 1980), in which it often is difficult to determine who is responsible for what. In this way, hospitals can be seen as sites of “risk-sharing” and “guilt-sharing” devices (Hughes 1958) that prevent any one person’s contribution to poor outcomes from being singled out. Complying with the checklist, which was overseen and signed off by a nurse, increased the visibility of individuals’ contributions to the process, which has repeatedly been demonstrated to reduce or eliminate the problem of social loafing—the problem that people (often unconsciously) tend to make less effort when they work collectively rather than individually (Latané, Williams, and Harkins 1979).

In addition to increasing the visibility and accountability of the staff’s performance, the checklist helped institutionalize good practices by making the insertion of a CVC into a routine with some of the characteristics of a ritual. This ritual function, again, became evident only after the project was completed and the program theory was being updated. Using the checklist to structure an event involving particular characters (the doctor and nurse) and ceremonial activities and scripts (the signing off, the option for the nurse to intervene) gave the insertion of the catheter a stylized form, helping identify it as a bounded episode distinct from other activities happening at the same time. The creation of a ritual is important because remembering is a fundamentally social activity (Halbwachs and Coser 1992); rituals are a form of collective remembering that helps reinforce memories of important events, including their moral significance (Connerton 1989). Within organizations, rituals are “mechanisms through which organizational members influence how other members are to think and feel—what they want, what they fear, what they should regard as proper and possible, and ultimately, perhaps, who they are” (Kunda 2006, 93). Organizational

members may begin by cooperating with a ritual for the sake of expedience, but they eventually may become so habituated to it that they no longer can conceive of alternative ways of acting (Powell and Colyvas 2008).

The checklist and the nurse's monitoring expressed a profound restructuring of organizational and professional roles, relationships, and identities (at least in relation to CVC insertion). In discussing this, it is helpful to resort to a little caricature. It is useful, for example, to understand doctors as being bred to deal with crisis. Their role is often one of heroic salvage, engaged in retrieving situations that are unusual or aberrant, and tinged with the possibility of catastrophe. They must be prepared to seize the initiative, to respond to unexpected events, and to come up with creative yet skillful solutions. Through their long socialization into the profession, they have learned that their peers do not always admire demonstrations of procedural obedience (Bosk 2003). Nondoctors' direct supervision of doctors' work outside an explicitly defined training context is unusual, as physicians are assumed to be self-monitoring and self-correcting. In fact, an attempt to supervise or correct a doctor during a procedure may be seen as breach of etiquette. Therefore, not only routinizing the CVC insertion procedure but also having an individual who is a member of a different, typically lower-status, professional tribe (nursing) supervise this routine might well be seen as a radical departure from normal practice.

Resistance to the program's effort to make this happen included complaints from doctors who said that they might be made to look stupid in front of colleagues and from nurses who said they did not want their "heads bitten off" by doctors if they intervened to stop a procedure. To help neutralize this resistance to renegotiating professional roles, the program leaders encouraged role modeling (senior doctors asking nurses to ensure that they challenged them when appropriate) and gave the participants new "vocabularies of motive," in C. Wright Mills's (1940) sense of accepted justifications for present, future, or past actions. The Michigan program supplied a new way of justifying action in regard to CVC-BSIs, one that emphasized the benefits for the patient. This "motive talk" did more than overcome the challenges associated with introducing the checklist; it also promoted "value congruity" (Sitkin and Roth 1993) in the professional community, at both the institutional level and on the wards. None of these features was specified in the

protocol but emerged over time as the program leaders learned from the participants.

5. *The Program Harnessed Data on Infection Rates as a Disciplinary Force.* Measuring phenomena and events and translating them into data enables evaluation, comparison, and intervention (Latour 1987) and so is indispensable to QI efforts. A key feature of the Michigan program was its systematic collection of data on CVC-BSIs and feedback to the units of these findings, expressed as the number of infections per one thousand CVC days. The vertical core of the project team centralized the data collection. This was a departure from many quality improvement efforts, in which the data remain local to individual organizations. Each hospital received anonymous reports of their infection rates compared with those of the entire cohort. In order not to violate the teams' trust, the infection rates were not made public. Local teams were encouraged to share among themselves the data showing the gap between the program's goal of eliminating CVC-BSIs and their own current performance. Many teams reported sharing the data throughout their unit, often posting performance reports on unit bulletin boards or in staff lounges or conference rooms. Some teams also routinely reported their data at medical and nursing staff meetings, project team meetings with their executive partner, and management and board meetings. Posters reporting weeks without an infection were displayed in ICUs, and moving a counter to show infection-free weeks became a ritual on many units. Although these activities were not written into the protocol, they emerged spontaneously as the program took place.

Measurement always has consequences: Heisenberg's uncertainty principle states that every system of measurement acts on the system being measured. The data collection process in Michigan had a performative dimension. By having ICUs disclose their data (albeit only within the clinical community), the program acted as a spur to action to control infections, a rather concrete demonstration of Louis Brandeis's claim that "sunlight is the best disinfectant." The data were used initially to rupture perceptions or assumptions that there was no "problem" to be addressed. At a psychological level, this disruption in taken-for-granted assumptions led to a need for meaning making and a reorganization of identities and norms, and motivated participants to action (Zittoun 2008).

Even though the project protocol specified collecting data on infections according to the Centers for Disease Control and Prevention's

criteria, how this would improve quality was not fully understood at the outset. Our updating of the theory suggests that the data on infection rates helped establish shared norms within a geographically disparate clinical community. Teams were eager to perform in accordance with group norms and particularly to reduce their rates in line with the overall downward trend. As Rose and Miller (1992) pointed out, measurement enables “action at a distance” and the coordination of those in different locations by enjoining those within these locales to work out where they are, calibrate themselves in relation to where they should be, and devise ways of getting from one state to the other. This is likely to have been encouraged by the tight coupling between infection rates and behavioral/cultural change: the teams received positive feedback from seeing rates of infection decline in response to their efforts. Regular feedback on performance can be used to motivate sustained efforts, as it can provide a sense of progress or help keep participants “on task,” particularly between collaborative meet-ups and when it is provided by peers (Bate 2000). Sharing blind data on other units’ performance in Michigan motivated the units to match the performance of those reporting the lowest rates of infection, and indeed some teams identified themselves to others so that they could discuss the comparative data. As more teams’ infection rates dropped, social pressure, reputational incentives, and hope for further improvements rose.

6. *The Program Made Skillful Use of “Hard Edges.”* Thus far we have suggested that Michigan worked primarily through consensus and with the voluntary consent of the program’s participants, deriving much of its force from social norms and the prospect of social sanction or loss of “face.” The program had no formal (legal) authority, although it did have some coercive features, which operated in a number of ways at a number of levels.

At the ward level, rituals such as those associated with the checklist made explicit the demands on the way people presented and conducted themselves. As Ericson (1995) argued, showing conformity to formal mechanisms of accountability has become a primary requirement for sustaining legitimacy in modern organizations. By leaving a visible trace of what happened, the checklist enhanced procedural accountability. It not only provided an outline of good conduct (Timmermans 2003), but it also functioned as a formal means of recording and inspecting that conduct through opening up to external scrutiny what the practitioners had done (compared with what they were supposed to have done). By

being “latently supervisory” (Freidson 1988), the checklist may have increased the staff’s commitment to their performance so as to avoid any later censure. This gave the checklist a hard edge, since it provided a means of determining and assessing procedural propriety. “Going by the book” was likely the safest action for practitioners, since this provided the most secure defense in the event of a challenge (legal or otherwise) about the quality of care provided and would provide strong evidence that their actions conformed with expectations of optimal care.

A second hard edge to the program was its use of activist tactics to ensure cooperation. Although it was not written into the project protocol, those ICUs that failed to return their data were subject to a number of sanctions by the program leaders. Their hospitals’ CEOs were contacted by the program leaders and asked for the data, and if the data were not forthcoming, the ICUs were asked to withdraw from the program. No hospital withdrew. In another innovation, the program team also encouraged nurses to call them if they were unable to prevent a doctor proceeding with a CVC insertion that did not meet the requirements, thereby enhancing the social functions of the checklist. Although no nurse called the program leaders to intervene, the nurses did reveal (in teleconferences and other forums) that they sometimes used the threat of calling to great effect, and thereby succeeded in discouraging the physicians’ colleagues who were intent on continuing their poor practice. Tactics for not reaching the point of having to call the program leaders were frequently discussed during the conference calls. Some physicians did call the program leaders to complain that the program was fomenting revolution among nurses, and these calls were used as an opportunity to explain that the program was trying to ensure that patients got the best care. All the callers appeared to accept this.

Third, the data on infection rates themselves were another hard edge. By making visible what otherwise might have remained opaque and by ranking units’ performance, the rates may have been regarded as coercive, or as a forceful prompt to action.

Conclusions

Not understanding how programs work when they are complex social interventions is likely to result in nontransferability, a limited ability to improve the program and its outcomes, and disappointment. There

can be little doubt that conducting ethnographic fieldwork while the program is running is an ideal first step in generating a high-quality description and theoretical understanding of a QI program. At present, this kind of concurrent evaluation of QI efforts is rare. Nonetheless, the need for good program theory that benefits from the experiences of running the program does not disappear, so other ways of capturing learning and improving the account of the program must be found. We believe that implementers can be connected with evaluators after the program has been completed. Updating the original program theory using the program leaders' experience and social scientists' contributions can produce an *ex post* theory. We used this approach to further understand the social factors contributing to the success of the Michigan project.

Our analysis has uncovered valuable lessons for others seeking to implement both the Michigan program and QI more generally, as well as areas for evaluation in future fieldwork studying similar programs. First, if institutional isomorphism explains why the ICUs joined the program, QI models that create a few organizations as "beacons" or "showcases" may have difficulty persuading others to be "followers." Therefore, it may be more helpful to engage a large number of organizations at once and then to allow mimetic or normative forces to encourage others to succumb.

Second, the community-based model developed in Michigan was likely one of the keys to its success. It allowed lessons to be shared and community bonds and obligations to be formed that generated highly effective (and inexpensive) normative effects and encouraged peer monitoring. By developing horizontal links between the participating units, the program was able to mobilize social forces beyond what would have been possible had the model been solely vertical (each unit engaging with the program team individually). Operating as a professional movement, the community was able to generate the energy and momentum of a grassroots movement while at the same time the program team was able to provide enough vertical structure, leadership, and resource to contain activities, ensure focus and direction, and secure the inclusion and cooperation of all relevant stakeholders. It enabled local people to assume leadership roles and achieve change. Third, the combination of strategies used at the sharp end of the ICUs had both expressive and instrumental functions that probably helped both practically and symbolically to institutionalize good practice and change culture.

A further benefit of our analysis is that it suggests points of vulnerability in the program. Finding the right team of project leaders to act as what we call the *vertical core* is critical. The program's participants must trust and respect the team. In turn, the team must be able to combine the right quality of scientific expertise with the ability to engage emotionally with participants and to use the correct combination of hard and soft tactics to ensure discipline but still give the participants responsibility and local ownership. The local teams are the ones who deliver results. The vertical core must focus on enabling teams to make changes, figuring out why some things are hard for staff, and making them easier to do. The disciplinary effects of data collection depend on the data being able to command legitimacy among participants and being used to encourage learning, rather than being used solely to punish or reward. Those collecting, processing, or feeding back data to the participating teams should have the necessary skills and resources for data management (Øvretveit et al. 2002), or the program will risk losing its credibility and persuasiveness. None of this is easy.

More issues arise at the sharp end. The restructuring of relationships implied by the checklist is a significant challenge to those seeking to establish new rituals. Having the nurse supervise the doctor's work may produce tensions like those found whenever broader patterns of authority relations are rearranged (Katz 1977). If nurses are unable or unwilling to observe and intervene, perhaps because this form of actor-monitor relationship (Stelling and Bucher 1972) may appear to violate normal professional relationships or to lead to personal challenges, problems will result. Relationships between doctors and nurses are subject to strongly reinforced etiquette rules, which include showing regard for others through deference, even when the recipient of that regard may not deserve it (Goffman and Best 2005). Notwithstanding the new vocabularies of motive provided by the program, some participants still may have problems with the emotional challenge of negotiating tension and conflict.

On average, the Michigan program succeeded, but different units demonstrated different levels of success. Contextual influences are therefore likely to be important sources of vulnerability for the program (Davidoff 2009). A major influence on organizational legitimacy is the extent to which the problem being addressed is perceived as being genuinely a problem. Establishing that the problem being targeted is not already under control may therefore be an important first step in future

iterations of the program. Organizational problems of legitimacy may also arise when ICUs struggle to find resources to divert to the program (e.g., staff time to collect data, complete questionnaires, and participate in teleconferences) away from what appear to be other (at least as important) clinical and other demands (Dixon-Woods et al. 2009). For instance, hospitals that enter the program for mimetic reasons (appearing to conform to expectations) may never fully engage with what the program requires and thus fall back on displays of compliance. In addition, while the reputational incentives within the community may have acted as a spur to improvement, it may also have been an incentive to “gaming” (Bevan and Hood 2006), or underreporting infections. A broader lesson for QI programs may be to choose interventions in areas with a reasonable consensus on the scientific evidence and supporting interventions to improve performance and the validity of the measures to be used. Without this, the prospect of endless and unproductive squabbling and the loss of legitimacy becomes real. Finally, QI program leaders should be aware that working in areas in which the outcome measures are less sensitive to changes in behavior, and thus less likely to provide positive feedback to teams, may be more difficult.

Our new theory is the result of updating the original program theory in light of the program leaders’ experiences and the social scientists’ theorization. Even though our example is based on a successful program, updating the program theory when a program has failed in this way is likely to be equally important, allowing the program leaders to point out what they felt went wrong, where their program design needs to be changed, and so on. Having program designers and leaders update their theory in light of their experience accomplishes many of the things that Carol Weiss and others see as helpful when done prospectively, including persuading the designers to explain and agree on what they have been doing, making programmatic assumptions explicit, and helping identify poorly defined, implausible, or contentious understandings of the program (Birckmayer and Weiss 2000). Updating the theory as we have done is a way for program leaders to say “now that we have done it, this is how we now think our program worked.” Improving the program theory in this way is likely to be extremely useful for others seeking to implement the program, providing shortcuts to learning and helping avert cargo-cult imitations. It also formalizes hypotheses that can be tested in future evaluations of the program, ultimately improving the science of QI. Developing an *ex post* theory as we have done is likely

valuable even when a process evaluation has been conducted, enabling continuous feedback, creating discussion circles, and promoting ongoing learning and improvement.

Our approach belongs firmly in the family of theory-oriented evaluation methods and thus shares much with other approaches in this tradition. It differs from some of these approaches in its emphasis on *updating* the program theory and its collaboration between program leaders and social scientists (implementers and evaluators), which is why we find the Bayesian analogy helpful. This analogy—though we do not want to stretch it too far—helps direct attention to the updating of prior beliefs in accordance with new data. Some aspects of prior beliefs (especially in the project's original protocol) may diminish or disappear, and new beliefs may be added. The realist evaluation approach may also be used retrospectively, but by contrast it is preoccupied with identifying context-mechanism-outcome configurations and testing these against various forms of evidence (which we have not attempted to do). Realist evaluation also differs from our approach in its rejection of correlational logic and its claim that “programs are theories incarnate” (rather than our view, which is that programs are what actually happen).

A theory that relies solely on program leaders' experience and social scientists' contribution, such as that which we have presented here, does, of course, have many limitations. We caution that it should not be regarded as a substitute for well-conducted fieldwork; rather, it is one way of making the best of an unsatisfactory situation in which no concurrent process evaluation data are available. The process of updating the theory like this, however, may fail, especially if the program leaders are not willing to be challenged and insist on recovering the sunk costs of their initial theory. The process we describe does not include testing the theory against the data (mainly because there are few suitable data available). For purposes of the Michigan program, our analysis should therefore be treated as well-informed hypotheses that can be explored empirically in future research, rather than as definitive conclusions. Future research may be able to distinguish the components of the program theory that have the most explanatory power from those that are less crucial to securing the program's outcomes. For example, empirical studies may be able to establish whether the measurement of and feedback on infection rates can by themselves significantly reduce CVC-BSIs.

Implementing a program like Michigan again without understanding what needs to be done and how and why it needs to be done risks the

program's failure, disenchantment among the target audience, and loss of the potential to address other quality issues in health care. As ICUs around the world seek to "Match Michigan," this ex post program theory may help inform not only what they need to do but also where they need to direct their energy and attention.

References

- Adler, P.S., S.W. Kwon, and C. Heckscher. 2008. Professional Work: The Emergence of Collaborative Community. *Organization Science* 19(2):359–76.
- Ariely, D., and M.I. Norton. 2009. Conceptual Consumption. *Annual Review of Psychology* 60:475–99.
- Aveling, E.L. 2010. The Impact of Aid Chains: Relations of Dependence or Supportive Partnerships for Community-Led Interventions? *AIDS Care* 22(1):1588–97.
- Bate, S.P. 1994. *Strategies for Cultural Change*. Oxford: Butterworth-Heinemann.
- Bate, S.P. 2000. Changing the Culture of a Hospital: From Hierarchy to Networked Community. *Public Administration* 78(3):485–512.
- Bate, S.P., and G. Robert. 2002. Knowledge Management and Communities of Practice in the Private Sector: Lessons for Modernising the National Health Service in England and Wales. *Public Administration* 80(4):643–63.
- Bate, S.P., G. Robert, and H. Bevan. 2004. The Next Phase of Healthcare Improvement: What Can We Learn from Social Movements? *Quality & Safety in Health Care* 13(1):62–66.
- Bate, S.P., G. Robert, and H. McLeod. 2002. Report on the "Breakthrough" Collaborative Approach to Quality and Service Improvement within Four Regions of the NHS. Birmingham: HSMC, University of Birmingham.
- Benford, R.D., and S.A. Hunt. 2003. Interactional Dynamics in Public Problems Marketplaces: Movements and the Counterframing and Reframing of Public Problems. In *Challenges and Choices: Constructionist Perspectives on Social Problems*, ed. J.A. Holstein and G. Miller, 153–86. New York: Aldine de Gruyter.
- Benford, R.D., and D.A. Snow. 2000. Framing Processes and Social Movements: An Overview and Assessment. *Annual Review of Sociology* 26(1):611–39.
- Benning, A., M. Ghaleb, A. Suokas, M. Dixon-Woods, J. Dawson, N. Barber, B.D. Franklin, A. Girling, K. Hemming, M. Carmalt, G.

- Rudge, T. Naicker, U. Nwulu, S. Choudhury, and R. Lilford. 2011. Large Scale Organisational Intervention to Improve Patient Safety in Four UK Hospitals: Mixed Method Evaluation. *BMJ* 342:d195.
- Bevan, G., and C. Hood. 2006. What's Measured Is What Matters: Targets and Gaming in the English Public Health Care System. *Public Administration* 84(3):517–38.
- Birkmayer, J.D., and C. Weiss. 2000. Theory-Based Evaluation in Practice: What Do We Learn? *Evaluation Review* 24(4):407–31.
- Black, N., and E. Thompson. 1993. Obstacles to Medical Audit: British Doctors Speak Out. *Social Science & Medicine* 36(7):849–56.
- Bosk, C.L. 2003. *Forgive and Remember: Managing Medical Failure*. Chicago: University of Chicago Press.
- Bosk, C.L., M. Dixon-Woods, C.A. Goeschel, and P.J. Pronovost. 2009. The Art of Medicine: Reality Check for Checklists. *The Lancet* 374(9688):444–45.
- Bovens, M. 1999. The Quest for Responsibility—Accountability and Citizenship in Complex Organizations. *Organizational Studies* 20(4):706–7.
- Braithwaite, J. 2008. The Regulatory State? In *The Oxford Handbook of Political Institutions*, ed. R. Rhodes, S. Binder, and B.A. Rockman, 407–30. Oxford: Oxford University Press.
- Bray, P., D.M. Cummings, M. Wolf, M.W. Massing, and J. Reaves. 2009. After the Collaborative Is Over: What Sustains Quality Improvement Initiatives in Primary Care Practices? *Joint Commission on Quality and Safety* 35:502–8.
- Buchanan, D., R. Addicott, L. Fitzgerald, E. Ferlie, and J.I. Baeza. 2007. Nobody in Charge: Distributed Change Agency in Healthcare. *Human Relations* 60(7):1065–90.
- Bucher, R. 1962. Pathology: A Study of Social Movements within a Profession. *Social Problems* 10(1):40–51.
- Campbell, C., and C. MacPhail. 2002. Peer Education, Gender and the Development of Critical Consciousness: Participatory HIV Prevention by South African Youth. *Social Science & Medicine* 55(2): 331–45.
- Chen, H.T., and P.H. Rossi. 1983. Evaluating with Sense: The Theory-Driven Approach. *Evaluation Review* 7(3):283–302.
- Cleaver, F. 2001. Institutions, Agency, and the Limitations of Participatory Approaches to Development. In *Participation: The New Tyranny?* ed. B. Cooke and U. Kothari, 36–55. London: Zed Books.
- Connell, J.P., and A.C. Kubisch. 1998. Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives: Progress, Prospects and Problems. In *New Approaches to Evaluating*

- Community Initiatives*, ed. J.P. Connell, A.C. Kubisch, L.B. Schorr, and C.H. Weiss, 15–44. Washington, DC: Aspen Institute.
- Connerton, P. 1989. *How Societies Remember*. Cambridge: Cambridge University Press.
- Cornish, F. 2006. Empowerment to Participate: A Case Study of Participation by Indian Sex Workers in HIV Prevention. *Journal of Community and Applied Social Psychology* 16(4):301–15.
- Cornish, F., and R. Ghosh. 2007. Necessary Contradictions of “Community-Led” Health Promotion: A Case Study of HIV Prevention in an Indian Red Light District. *Social Science & Medicine* 64(2):496–507.
- Davidoff, F. 2009. Heterogeneity Is Not Always Noise. *JAMA* 302(23):2580–86.
- Davies, H., A. Powell, and R. Rushmer. 2007. Why Don’t Clinicians Engage with Quality Improvement? *Journal of Health Services Research & Policy* 12(3):129–30.
- DiMaggio, P.J., and W.W. Powell. 1983. The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review* 48(2):147–60.
- Dixon-Woods, M. 2010. Why Is Patient Safety So Hard? A Selective Review of Ethnographic Studies. *Journal of Health Services Research & Policy* 15(suppl. 1):11–16.
- Dixon-Woods, M., A. Suokas, E. Pitchforth, and C. Tarrant. 2009. An Ethnographic Study of Classifying and Accounting for Risk at the Sharp End of Medical Wards. *Social Science & Medicine* 69(3):362–69.
- Donaldson, S.I. 2007. *Program Theory–Driven Evaluation Science: Strategies and Applications*. New York: Erlbaum.
- Dopson, S., L. Fitzgerald, E. Ferlie, J. Gabbay, and L. Locock. 2002. No Magic Targets! Changing Clinical Practice to Become More Evidence-Based. *Health Care Management Review* 27(3):35–47.
- Dopson, S., L. Locock, E. Ferlie, and L. Fitzgerald. 2003. Evidence-Based Medicine and the Implementation Gap. *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine* 7(3):311–30.
- Dube, N., and D. Wilson. 1996. Peer Education Programmes among HIV-Vulnerable Communities in Southern Africa. In *HIV/AIDS in the South African Mining Industry*, ed. B. Williams and C. Campbell, 107–110. Johannesburg: Epidemiology Research Unit
- Dunn, J.L. 2005. “Victims” and “Survivors”: Emerging Vocabularies of Motive for “Battered Women Who Stay.” *Sociological Inquiry* 75(1):1–30.

- Ericson, R.V. 1995. The News Media and Accountability in Criminal Justice. In *Accountability for Criminal Justice: Selected Essays*, ed. P.C. Stenning, 135–61. Toronto: University of Toronto Press.
- Feynman, R.P. 1999. Cargo Cult Science: Some Remarks on Science, Pseudoscience, and Learning How to Not Fool Yourself. In *The Pleasure of Finding Things Out*, by R.P. Feynman and J. Robbins, 205–16. Cambridge, MA: Perseus Books.
- Finn, R. 2008. The Language of Teamwork: Reproducing Professional Divisions in the Operating Theatre. *Human Relations* 61(1):103–130.
- Finn, R., G. Currie, and G. Martin. 2010. Team Work in Context: Institutional Mediation in the Public-Service Professional Bureaucracy. *Organization Studies* 31(8):1069–97.
- Freidson, E. 1988. *Profession of Medicine: A Study of the Sociology of Applied Knowledge*. Chicago: University of Chicago Press.
- Gastmeier, P., and C. Geffers. 2006. Prevention of Catheter-Related Bloodstream Infections: Analysis of Studies Published between 2002 and 2005. *Journal of Hospital Infection* 64(4):326–35.
- Gillespie, A., F. Cornish, E.L. Aveling, and T. Zittoun. 2008. Conflict-ing Community Commitments: A Dialogical Analysis of a British Woman's World War II Diaries. *Journal of Community Psychology* 36(1):35–52.
- Goffman, E., and J. Best. 2005. *Interaction Ritual: Essays in Face-to-Face Behavior*. Piscataway, NJ: Transaction Publishers.
- Guareschi, P.A., and S. Jovchelovitch. 2004. Participation, Health and the Development of Community Resources in Southern Brazil. *Journal of Health Psychology* 9(2):311–22.
- Gujit, I., and M.K. Shah. 1998. *The Myth of Community: Gender Issues in Participatory Development*. London: Intermediate Technology Publications.
- Halbwachs, M., and L.A. Coser. 1992. *On Collective Memory*. Chicago: University of Chicago Press.
- Hilgartner, S., and C.L. Bosk. 1988. The Rise and Fall of Social Problems: A Public Arenas Model. *American Journal of Sociology* 94(1):53.
- Hughes, E.C. 1958. *Men and Their Work*. Glencoe, IL: Free Press.
- Hwang, H., and W.W. Powell. 2005. Institutions and Entrepreneurship. In *Handbook of Entrepreneurship Research*, ed. H. Hwang and W.W. Powell, 179–210. Dordrecht: Kluwer.
- Innes, J.E., and D.E. Booher. 1999. Consensus Building as Role Playing and Bricolage toward a Theory of Collaborative Planning. *Journal of the American Planning Association* 65(1):9–26.

- Katz, J. 1977. Cover-up and Collective Integrity: On the Natural Antagonisms of Authority Internal and External to Organizations. *Social Problems* 25(1):3–17.
- Kickert, W.J.M., E.H. Klijn, and J.F.N. Koppenjan. 1997. *Managing Complex Networks: Strategies for the Public Sector*. London: Sage.
- Kothari, U. 2001. Power, Knowledge and Social Control in Participatory Development. In *Participation: The New Tyranny?* ed. B. Cooke and U. Kothari, 139–52. London: Zed Books.
- Kunda, G. 2006. *Engineering Culture: Control and Commitment in a High-Tech Corporation*. Philadelphia: Temple University Press.
- Latané, B., K. Williams, and S. Harkins. 1979. Many Hands Make Light the Work: The Causes and Consequences of Social Loafing. *Journal of Personality and Social Psychology* 37(6):822–32.
- Latour, B. 1987. *Science in Action*. Milton Keynes: Open University Press.
- Lazega, E. 2000. Rule Enforcement among Peers: A Lateral Control Regime. *Organization Studies* 21(1):193–214.
- Lipitz-Snyderman, A., D. Steinwachs, D.M. Needham, E. Colantuoni, L.L. Morlock, and P.J. Pronovost. 2011. Impact of a Statewide Intensive Care Unit Quality Improvement Initiative on Hospital Mortality and Length of Stay: Retrospective Comparative Analysis. *BMJ* 342:d219.
- Lomas, J. 2005. Using Research to Inform Healthcare Managers' and Policy Makers' Questions: From Summative to Interpretive Synthesis. *Healthcare Policy* 1(1):55–71.
- Lowndes, V., and C. Skelcher. 1998. The Dynamics of Multi-Organizational Partnerships: An Analysis of Changing Modes of Governance. *Public Administration* 76(2):313–33.
- Martin, G., G. Currie, and R. Finn. 2009a. Leadership, Service Reform and Public-Service Networks: The Case of Cancer-Genetics Pilots in the English NHS. *Journal of Public Administration Research and Theory* 19(4):769–94.
- Martin, G., G. Currie, and R. Finn. 2009b. Reconfiguring or Reproducing Intra-Professional Boundaries? Specialist Expertise, Generalist Knowledge and the “Modernization” of the Medical Workforce. *Social Science & Medicine* 68(7):1191–98.
- McMillan, D.W., and D.M. Chavis. 1986. Sense of Community: A Definition and a Theory. *Journal of Community Psychology* 14(1):6–23.
- Mermel, L.A., M. Allon, E. Bouza, D.E. Craven, P. Flynn, N.P. O'Grady, I.I. Raad, B.J. Rijnders, R.J. Sherertz, and D.K. Warren. 2009. Clinical Practice Guidelines for the Diagnosis and Management of Intravascular Catheter-Related Infection: 2009 Update by the

- Infectious Diseases Society of America. *Clinical Infectious Diseases* 49(1):1–45.
- Mills, C.W. 1940. Situated Actions and Vocabularies of Motive. *American Sociological Review* 5(6):904–13.
- Mills, P.D., and W.B. Weeks. 2004. Characteristics of Successful Quality Improvement Teams: Lessons from Five Collaborative Projects in the VHA. *Joint Commission Journal on Quality and Patient Safety* 30(3):152–62.
- Offer, A. 1997. Between the Gift and the Market: The Economy of Regard. *Economic History Review* 50(3):450–76.
- Ostrom, E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- Øvretveit, J., P. Bate, P. Cleary, S. Cretin, D. Gustafson, K. McInnes, H. McLeod, T. Molfenter, G. Robert, S. Shortell, and T. Wilson. 2002. Quality Collaboratives: Lessons from Research. *Quality and Safety in Health Care* 11(4):345–51.
- Pawson, R., and N. Tilley. 1997. *Realistic Evaluation*. London: Sage.
- Poteete, A.R. 2010. *Working Together: Collective Action, the Commons, and Multiple Methods in Practice*. Princeton, NJ: Princeton University Press.
- Powell, W.W., and J. Colyvas. 2008. Microfoundations of Institutional Theory. In *Handbook of Organizational Institutionalism*, ed. R. Greenwood, C. Oliver, K. Sahlin-Andersson, and R. Suddaby, 276–98. London: Sage.
- Powell, W.W., and P.J. DiMaggio. 2001. *The New Institutionalism in Organizational Analysis*. Chicago: University of Chicago Press.
- Pronovost, P.J., C.A. Goeschel, E. Colantuoni, S. Watson, L.H. Lubomski, S.M. Berenholtz, D.A. Thompson, D.J. Sinopoli, S. Cosgrove, J.B. Sexton, J.A. Marsteller, R.C. Hyzy, R. Welsh, P. Posa, K. Schumacher, and D. Needham. 2010. Sustaining Reductions in Catheter Related Bloodstream Infections in Michigan Intensive Care Units: Observational Study. *BMJ* 340:c309.
- Pronovost, P., D. Needham, S. Berenholtz, D. Sinopoli, H. Chu, S. Cosgrove, B. Sexton, R. Hyzy, R. Welsh, G. Roth, J. Bander, J. Kepros, and C. Goeschel. 2006. An Intervention to Decrease Catheter-Related Bloodstream Infections in the ICU. *New England Journal of Medicine* 355(26):2725.
- Roberts, K.A., M. Dixon-Woods, R. Fitzpatrick, K.R. Abrams, and D.R. Jones. 2002. Factors Affecting Uptake of Childhood Immunisation: A Bayesian Synthesis of Qualitative and Quantitative Evidence. *The Lancet* 360(9345):1596–99.
- Rose, N., and P. Miller. 1992. Political Power beyond the State: Problematics of Government. *British Journal of Sociology* 43(2):173–205.

- Scott, W.R. 1991. Unpacking Institutional Arguments. In *The New Institutionalism in Organizational Analysis*, ed. W.W. Powell and P.J. DiMaggio, 164–82. Chicago: University of Chicago Press.
- Shojania, K.G., and J.M. Grimshaw. 2005. Evidence-Based Quality Improvement: The State of the Science. *Health Affairs* 24(1):138–50.
- Sitkin, S.B., and N.L. Roth. 1993. Explaining the Limited Effectiveness of Legalistic “Remedies” for Trust/Distrust. *Organization Science* 4(3):367–92.
- Spiegelhalter, D.J., K.R. Abrams, and J.P. Myles. 2004. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Hoboken, NJ: Wiley.
- Spiegelhalter, D.J., J.P. Myles, D.R. Jones, and K.R. Abrams. 2000. Bayesian Methods in Health Technology Assessment: A Review. *Health Technology Assessment* 4(38):1–129.
- Stame, N. 2004. Theory-Based Evaluation and Types of Complexity. *Evaluation* 10(1):58.
- Stelling, J., and R. Bucher. 1972. Autonomy and Monitoring on Hospital Wards. *Sociological Quarterly* 13(4):431–46.
- Stone Sweet, A., N. Fligstein, and W. Sandholtz. 2001. The Institutionalization of European Space. In *The Institutionalization of Europe*, ed. A. Stone Sweet, N. Fligstein, and W. Sandholtz, 1–28. Oxford: Oxford University Press.
- Sutton, J.R., F. Dobbin, J.W. Meyer, and W.R. Scott. 1994. The Legalization of the Workplace. *American Journal of Sociology* 99(4):944–71.
- Thompson, D.F. 1980. Moral Responsibility of Public Officials: The Problem of Many Hands. *American Political Science Review* 74(4):905–16.
- Timmermans, S. 2003. *The Gold Standard: The Challenge of Evidence-Based Medicine and Standardization in Health Care*. Philadelphia: Temple University Press.
- Waring, J., and G. Currie. 2009. Managing Expert Knowledge: Organizational Challenges and Managerial Futures for the UK Medical Profession. *Organization Studies* 30(7):755–78.
- Weiss, C. 1988. Evaluation for Decisions: Is Anybody There? Does Anybody Care? *Evaluation Practice* 9(1):5–19.
- Weiss, C. 1995. Nothing as Practical as a Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families. In *New Approaches to Evaluating Community Initiatives: Concepts, Methods and Contexts*, ed. J. Connell, A. Kuchisch, L.B. Schorr, and C. Weiss, 65–92. New York: Aspen Institute.

- Wilkinson, A., G. Godrey, and M. Marchington. 1997. Bouquets, Brickbats and Blinkers: Total Quality Management and Employee Involvement in Practice. *Organization Studies* 18(5):799–819.
- Zittoun, T. 2008. Learning through Transitions: The Role of Institutions. *European Journal of Psychology of Education* 23(2):165–81.

Acknowledgments: This paper is based on work commissioned by the Health Foundation (Registered Charity Number: 286967). Charles L. Bosk's work on patient safety is supported by a Health Investigator Award from the Robert Wood Johnson Foundation. Peter J. Pronovost and Christine A. Goeschel received support for this work from the U.S. Agency for Healthcare Research and Quality. We are very grateful to Frank Davidoff, Trish Greenhalgh, Myles Leslie, Graham Martin, and Julian Bion for comments on earlier drafts of this article. We thank Lisa Hallam for help in preparing the manuscript.