

**Expanding Research and Evaluation Designs to  
Improve the Science Base for Health Care and  
Public Health Quality Improvement Symposium**

Meeting Summary  
September 13–15, 2005

Wyndham City Center Hotel  
Washington, DC

*Prepared by Barbara DeVinney for the Agency for Healthcare Research and Quality  
under a subcontract to Contract No. 290-01-0001 with Cygnus Corporation, Inc.*

## ***Contents***

Executive Summary _____	3
Summary of the Presentations _____	7
Description of the meeting format _____	7
Expanding Research and Evaluation Designs for Quality Improvement Interventions _____	7
An Integrated Model for Improvement: Implications for Study Design _____	9
Meeting Purposes and Process Revisited _____	15
Session I. Case Studies: QII at the Clinical Microsystem level _____	16
QII to Increase Delivery of Clinical Preventive Services in Office-Based Primary Care ____	16
QII to Increase Timely Delivery of Surfactant to High-Risk Newborns During Hospital Labor and Delivery _____	25
Session II. Research Evaluation Designs for QII at the Health “Plan” Level _____	30
Case: Expanding and Testing VA Collaborative Care Models for Depression _____	30
Session III. QII at the Multiple Clinical Systems Level: Improving Chronic Illness Care Evaluation _____	34
Session IV. QII at the State/Regional Level _____	39
California State Tobacco Prevention and Control, Public Health QII Evaluation Design Issues, and Lessons for Health Care QII in the Policy Environment _____	39
Summary of the Recommendations _____	44
Reports from the Homogeneous Breakout Groups _____	45
Disparities Recommendations _____	45
Research Training Directors’ Recommendations _____	46
Design Experts’ Recommendations _____	47
Peer Reviewers’ and Journal Editors’ Recommendations _____	48
Users’ Recommendations _____	49
Reports from the Heterogeneous Breakout Groups _____	51
Next Steps, Parting Remarks _____	57
The Core Planning Group _____	60
Participant List _____	61
Symposium Agenda _____	70
Acknowledgments _____	76
References _____	77

## *Executive Summary*

Numerous gaps in health care and public health quality recently have been recognized, and studies on the effectiveness of strategies labeled health care and public health quality improvement interventions are a relatively recent development.<sup>1,2,3,4,5,6,7</sup> Quality improvement (QI) in health care has recently been defined as “...identifying those activities that *increase the rate with which practices known to be effective are applied to patient care.*”<sup>8</sup> For this symposium, we defined QI interventions even more broadly, so as to include a range of strategies implemented in “real-world” settings for the purpose of expanding the delivery, reach, and impact of population-level, evidence-based health care and related public health interventions.

The goals for the symposium were:

- To review a range of quality improvement interventions (QIIs) and the critical questions that arise in evaluation of these interventions, including basic questions of internal (does it work in the setting and for the condition in which it was tested) and external (will it work in other settings and conditions) validity.
- To identify the strengths, weaknesses, and tradeoffs of alternative designs and methods for evaluating QIIs.
- To achieve a working consensus about the range of traditional and innovative designs and methods that can be used to answer key QII questions.
- To identify and suggest strategies to facilitate possible changes in funding mechanisms, review processes, research and publication standards, and research training that could help accelerate the development and spread of reliable QII research methods.

The format of the meeting was: 1) an opening night dinner with keynote remarks designed to frame the meeting; 2) presentations on quality improvement projects at a variety of levels of the health care and public health systems followed by commentaries and discussion periods; 3) breakout sessions in which all attendees were encouraged to make recommendations to improve quality improvement intervention research and evaluation, and 4) reports of the recommendations from the facilitators of the breakout sessions. This symposium was designed to be interactive and to elicit the input of the approximately 120 attendees, whose backgrounds spanned quality improvement, health care, public health, research, training, and patient advocacy.

### *Framing*

Introductory remarks focused on the need for quality improvement in health care, desirable characteristics of information from quality improvement studies, and that funding and performing randomized trials will not provide all the answers for QI because of the characteristics of quality improvement interventions:

- The targets of QI interventions (QIIs) are not individual patients.
- QIIs are complex and sometimes change over time.
- The setting is an essential component of the question and QII.

A broad range of evaluation designs were described as potentially applicable to QIIs, including but not limited to, the randomized controlled trial, the group-randomized trial, the case-control study, the interrupted time series, and qualitative methods.

The quality movement, which is directed at the level of systems of care of communities and of the public's health, was characterized as the second phase of several recent important performance initiatives to reduce the "Knowledge-Performance Gap" in health care and public health. The first initiative was the *practice* of evidence-based medicine, which is largely directed at the level of care of individual patients. Drawing on elements of these initiatives, Batalden and Davidoff have developed an integrated model of improvement. The integrating concept that underlies this model is that improvement is fundamentally a learning process. The model is driven by three kinds of learning: a) scientific discovery, which is learning about "what is"; b) experiential discovery, which is "learning about learning" or, in effect, "learning about how learning works"; and c) experiential learning, which is learning "how to" do something. Experiential learning, to put it simply, is learning by doing. The importance of publishing was emphasized, and an announcement was made about forthcoming publication guidelines for QII studies.<sup>9</sup>

#### *Lessons learned from past QIIs*

The symposium included a series of presentations on QIIs and lessons learned through them at four increasingly complex levels of the health care and public health systems. The first level examined was the clinical microsystem level. A clinical microsystem is defined as "a group of clinicians and staff working together with a shared clinical purpose to provide care for a population of patients."<sup>10</sup> Hospitals, hospital units (e.g., intensive care units), primary care practices and other such entities can be considered clinical microsystems. The presentations examined QIIs for prevention in primary care settings and a QII for timely delivery of surfactant in neonatal intensive care units.

The second level examined was the health "plan" level. For this level, the presentation examined a series of QIIs performed within the Veterans Health Administration on collaborative care for depression.

The third example was a QII at multiple levels in the so-called "chain of effect" for systems change and quality improvement.<sup>11</sup> For this level, the presentation reviewed lessons learned from the Improving Chronic Illness Care Evaluation (ICICE). The ICICE was an evaluation of the Institute for Healthcare Improvement's collaborative approach to quality improvement from organizational to clinical microsystem levels. In this intervention, "organization level" interventions included a requirement for the organizations to form interdisciplinary teams, and for senior leadership of the organizations to provide resources and a plan for disseminating the new Chronic Care Model-based care system. Changes actually implemented varied from organization to organization and, within organizations, at the clinic level, and, within clinics, at the individual patient-provider level.

The fourth level examined during the symposium was the state/regional level. For this level, the presentation examined the California state tobacco control and prevention effort begun in 1988.

Recommendations and observations addressed: 1) alternative research designs and methods for QII research 2) how to implement quality improvement (and implications for research), and 3) how to combine quality improvement with research. In addition, specific suggestions were directed toward funders of QI research.

The discussion on research design addressed the value and feasibility of randomized trials, as well as the types of data to be collected to understand more about the “whys” of program success or failure. The cluster randomized trial is a viable design which is applicable to studies of system-level change. However, some practices and health plans have been unwilling to participate in randomized trials. Studies within a single large health plan where the health plan response is one of the units of analysis cannot use randomized designs. In the absence of randomization, several aspects of design can help strengthen inferences about causality and generalizability, including: using an evaluation logic model; using a matched control group with a before and after design; collecting data from multiple sources; planning for and testing potential biases; and being sure that there is a solid evidence base for any clinical intervention as well as for clinical microsystem interventions.

In addition to measuring clinical outcomes, studies also need to measure behavioral and organizational changes that occur (intentionally and otherwise) as a result of QIIs. Currently, resource limitations can hamper the researchers’ ability to collect data on variations in implementation across sites when sites are permitted to develop their own QII strategies. Knowing more about organizations will help the field understand how to tailor interventions to an organization’s stage of readiness for change, and it will enable us to go beyond general statements, such as the importance of “leadership support and buy-in.” New measurement tools may be needed.

Recommendations for combining QIIs with research focused on ways to increase both rigor and relevance: by working at multiple levels of a system, by developing participatory relationships that transcend single projects, by considering different ways of knowing (e.g., qualitative and quantitative data), by fostering shared learning among participants in research, and by pursuing development alongside research in QI. To facilitate more quality improvement engagement, more research is needed on how to entice practice leaders into QI studies. Funding agencies could help create research infrastructures in real-world settings and fund in shorter cycles to take advantage of natural experiments in QI.

The multi-level California tobacco control effort provided lessons for health care as health care QI is beginning to take place at levels beyond clinical microsystems and health plans. In the program, there were state-level policy changes (tobacco taxes) and a wide variety of grants for local programs. Thus, local adaptation was a built-in, essential feature of the program. Evaluations of program success depended on ongoing surveillance data supplemented by intervention-specific measurement and evaluation activities. For health care QIIs with these features, implementers and researchers should:

- Maximize the design within the constraints.
- Maximize learning from variation within the intervention.
- Validate and then use existing surveillance measures.
- Use multiple and differing measures of the critical phenomenon to increase one's confidence that there is a real treatment effect.

### *Challenges to QI Research*

Some of the challenges in QI research include the fact that people want to be able to isolate the effects of individual components within multifactorial interventions, but it is difficult for them to know where to invest resources to address this question. It also is necessary to strike a balance between structure and autonomy. The Institute for Healthcare Improvement's Breakthrough Series, for example, is a general QI process rather than a set of prescribed interventions. Centers want to have a menu of intervention choices that they can adapt to their needs.

### *Overarching Recommendations*

Numerous specific solutions for improving QI science were made during presentations and discussion of these studies or were generated by breakout groups tasked with addressing one of five areas: research designs, peer review and journal editing, health disparities, research training, and users of quality improvement research (stakeholders and evidence synthesizers). The solutions included:

- Develop a taxonomy of QI for health care and public health.
- Develop flexible intervention and evaluation toolkits that can be locally adapted.
- Develop a transdisciplinary theory for health care and public health QI.
- Address disparities in health in all health care and public health QI projects.
- Build cumulative knowledge in QI.
- Organize peer review panels to include transdisciplinary representation.
- Develop funding mechanisms so that quick review and funding is available for the study of acute events and natural experiments in quality improvement and system change.
- Fund a major dataset of QI research so that many types of questions can be asked.
- Include partnership requirements in QII RFAs.
- Work with the academic tenure and promotion system to develop viable career paths for QI researchers and implementers in academic institutions.

## ***Summary of the Presentations***

### ***Description of the meeting format***

The format of the meeting was: 1) an opening night dinner with keynote remarks designed to frame the meeting; 2) presentations on quality improvement projects at a variety of levels of the health care and public health systems followed by commentaries and discussion periods; 3) breakout sessions in which all attendees were encouraged to make recommendations to improve quality improvement intervention research and evaluation, and 4) reports of the recommendations from the facilitators of the breakout sessions. This symposium was designed to be interactive and to elicit the input of the approximately 120 attendees, whose backgrounds spanned quality improvement, health care, public health, research, training, and patient advocacy.

### **Tuesday, September 13, 2005**

#### **Welcoming Remarks**

Thomas Chapel, M.A., M.B.A. (Facilitator)

Senior Health Scientist, Office of the Director, Office of Strategy and Innovation, Centers for Disease Control and Prevention

Mr. Chapel welcomed participants to the symposium and thanked them for attending. The symposium was sponsored by the Agency for Healthcare Research and Quality (AHRQ); the Centers for Disease Control and Prevention (CDC); the National Institutes of Health's (NIH's) Office of Disease Prevention, National Cancer Institute (NCI), and National Heart, Lung, and Blood Institute (NHLBI); the Department of Veterans Affairs (VA) Quality Enhancement Research Initiative (QUERI); and the Robert Wood Johnson Foundation (RWJF). Dr. Denise Dougherty, Ph.D., Senior Advisor, Office of Extramural Research, Education, and Priority Populations (OEREP), AHRQ, introduced Dr. Clancy.

#### **Expanding Research and Evaluation Designs for Quality Improvement Interventions**

Carolyn Clancy, M.D.

Director, Agency for Healthcare Research and Quality

Several years ago, the Institute of Medicine (IOM) published the report *Crossing the Quality Chasm* which discusses the gap that persists between the best possible health care available and the care that most patients receive. Dr. Clancy emphasized that progress in improving health care in this country still has a long way to go. The question is: are we making any progress? The Agency for Healthcare Research and Quality's second annual reports focusing on the state of health care quality and disparities in America showed overall improvement and identified areas still in need of improvement, including pervasive disparities related to race, ethnicity, and socioeconomic status. Dr. Clancy highlighted some of the findings in these reports, including the following:

- Health care quality continues to improve at a modest pace across most measures of quality, with the overall rate of improvement being 2.8 percent. This improvement covers a range of areas, including prenatal care, hip fracture, and alcohol dependence.

- Health care quality improvement (QI) is variable, with notable areas of high performance in patient safety, which showed a 10.2% improvement, and in Medicare's QIO (quality improvement organization) measures, which showed a 9.2% improvement.

A survey performed last fall by AHRQ, in conjunction with the Kaiser Family Foundation and the Harvard School of Public Health, showed that the proportion of the American public dissatisfied with health care increased from 44% in 2000 to 55% in 2004. Meanwhile, 40% of people surveyed believe that the quality of health care in this country has gotten worse. Dr. Clancy believes that more people are beginning to recognize that disparities in care do exist in relation to race, ethnicity, income, education, and other factors. She noted the need for collaboration between research groups that address quality and those examining disparities in health care.

AHRQ's mission is to improve the quality, safety, efficiency, and effectiveness of health care for all Americans. To achieve that goal, the Agency must find answers that are valid, timely, convincing, and practical. However, this is a great challenge. There is reason to be humble regarding the rate of turning funded research to the benefit of patient care. According to Balas<sup>12</sup>, it takes 17 years to turn 14% of original research to the benefit of patient care. There are many clinical treatments that have gone by the wayside, such as hormone replacement therapy to prevent cardiovascular disease or sulfuric acid to treat scurvy. Like clinical interventions, quality interventions need to be evaluated. Thus, the methods need to be right. If we want our results to have effect in the short-term, they have to be practical. Dr. Clancy added there are great opportunities to build on what we know and the methods we have, in order to develop the methods we need. Funding and performing randomized trials will not provide all the answers because of the characteristics of quality improvement interventions. Randomized trials are sometimes not appropriate for these studies because:

- The targets of QI interventions (QIIs) are not individual patients.
- QIIs are complex and sometimes change over time.
- The setting is an essential component of the question and QII.

Dr. Clancy noted that we can learn important lessons when things do not go well. According to an AHRQ-funded study at a major urban teaching hospital, the computerized physician order entry systems that are expected to significantly reduce medication errors must be implemented thoughtfully to avoid facilitating certain types of errors. Implementation problems can be minimized through testing and adaptations to meet the needs of individual clinical settings.

Dr. Clancy listed some currently important questions in the field of quality improvement interventions:

- Can a regional health information organization improve interoperability of health information technology systems and improve patient safety and quality of care?
- Can pay for performance improve quality?
- Do changes in hospital culture reduce medical errors?
- What QI strategies work for reducing disparities?

Dr. Clancy noted that to develop the area of QII evaluation designs and methods, we can draw upon the evaluation designs used in medical, clinical, and health services research, such as the randomized controlled trial, group-randomized trial, and case-control study, as well as those used in the behavioral/social science fields, such as the interrupted time series and qualitative methods, to give just a few examples.

Dr. Clancy concluded by suggesting that we need to look at how we use these methods and build on these methods to get to the next phase of quality improvement intervention and evaluation.

### **An Integrated Model for Improvement: Implications for Study Design**

Frank Davidoff, M.D.

Editor Emeritus, *Annals of Internal Medicine*

Executive Editor, Institute for Healthcare Improvement

Dr. Davidoff noted that the goals of this symposium are, first, to find better ways of knowing whether QI in health care works and, second, to find ways to increase the impact of that knowledge. To accomplish those purposes, three questions must be answered:

- What kind of evidence do QI studies produce?
- What are the strengths and limitations of that evidence?
- How can QI and the dissemination of its results help one another?

It is the role of this group to answer these questions. Dr. Davidoff said that his job was to create a context in which these questions can be answered.

Why is it so hard to improve health care?

Dr. Davidoff noted that providing health care that is both science-based and humane (caring) is an extraordinarily complex and demanding undertaking. He presented a “basic model” for the delivery of scientifically-grounded medical care (Figure 1a, below). The model involves the judicious application of a large body of established, generalizable knowledge to the idiosyncratic needs of patients and families, as well as to micro- and macro-systems, to reach the desired outcomes. This must all be done in a demanding and multi-layered local socio-technical environment. Dr. Davidoff noted that the missing element from this model is “improvement.” Although the care provided in that model may be optimal at any moment, it is not going to improve over time unless mechanisms for improving that model are built into the process.

Dr. Davidoff emphasized that there is an enormous gap between medicine’s potential and what it actually delivers, i.e., the “Knowledge-Performance Gap”. However, several recent important performance initiatives have been designed to help close this gap. The first initiative was the *practice* of evidence-based medicine, which is largely directed at the level of care of individual patients, in contradistinction to the *concept* of evidence-based medicine. The second is the quality movement, which is directed at the level of systems of care of communities and of the public health. Drawing on elements of these initiatives, and building on the basic model, Dr. Paul Batalden and Dr. Davidoff have developed an integrated model of improvement (Figure 1b). The integrating concept that underlies this model is that improvement is fundamentally a

learning process. The model is driven by three kinds of learning: a) scientific discovery, which is learning about “what is”; b) experiential discovery, which is “learning about learning” or, in effect, “learning about how learning works”; and c) experiential learning, which is actually learning “how to” do something. Unless newly discovered scientific knowledge, which has entered the “reservoir” of new generalizable knowledge, is translated into practice, it does not have an effect. Dr. Davidoff posed the question, “Where does experiential learning enter in?” We see it in the “performance elements,” in terms of locating, acquiring, and evaluating established knowledge, adapting evidence to local circumstances, redesigning practices, executing changes, measuring outcomes and using those data to modify care accordingly. However, another element is needed. There is learning to be done about what goes on in each of those steps, and we call that learning “experiential discovery.” Scientific discovery and experiential discovery enter the integrated model as inputs to the knowledge reservoir. In other words, knowledge translation is used to redesign care delivery. The changes are then executed and measured, and feedback is used to modify the information for care delivery.

Dr. Davidoff compared the three kinds of learning listed above. Experiential learning is the source of most of our knowledge, not just in medicine, but in general. It is driven by experience and not by doing randomized trials. However, Dr. Davidoff said that despite the ubiquity of experiential learning, it largely has been ignored in academic medicine. Experiential learning necessarily takes place in concrete, local, and real-world contexts, which means that variables are either difficult to control or cannot be controlled. The product of this learning is concrete “know how” and competence. The learning that takes place is also described as “reflexive,” because the product or outcome of that learning is intended to change the thing one is learning. In this way, the effects of experiential learning are intrinsically unstable because of the continual feedback on performance.

Dr. Davidoff contrasted experiential learning with scientific discovery and experiential discovery. The setting for scientific discovery is often artificial in clinical research and very protocol-driven for the purpose of controlling contextual variables. The product of scientific discovery is abstract in that it is conceptual knowledge. The product of scientific discovery is unchanged by its discovery. For example, discovery that an antibiotic is effective for certain diseases does not change its effectiveness. Originality is the hallmark of scientific discovery, because the purpose is to discover what is unknown or not understood.

How does experiential learning work? To describe it simply, it is learning by doing. The generic cycle of experiential learning consists of four elements:

- Experiencing something fully and openly.
- Questioning what has just occurred.
- Conceptualizing: trying to relate the questions the experience has raised to other models or processes and reflecting on how to do it better.
- Going back and trying again.

All four elements are essential. Without questioning and conceptualizing, there is stagnation. Without experience, there is pedantry and inaction.

Why is experiential learning particularly relevant to quality improvement? The characteristics of experiential learning, in that it is “real world,” reflexive, produces know how, is unstable, and applies what is known, are particularly suited to dealing with the problems of everyday health care delivery. These problems are that health care delivery is messy, nonlinear, and happens in complex adaptive systems. The characteristics of experiential learning listed above also present challenges in that experiential learning is not easily amenable to traditional hypothesis-testing research methods. This points to the importance of *experiential discovery*, the science of learning about experiential learning, which is a discipline that is just beginning to coalesce. Experiential discovery can provide the link between the gritty, messy experiential world of local “know how” and the orderly, scholarly world of conceptual knowledge.

Dr. Davidoff presented some thoughts on the nature and role of that link between experiential learning and experiential discovery. The traditional experiential learning cycle (experience, question, conceptualize, retry) describes the informal individual learning process. However, it does not capture the formal work done by organizations to improve their performance. Therefore, the study cycle needs to include an additional planning step and to collapse the questioning and conceptualizing steps into a single study step. This results in the familiar “plan, do, study, act” cycle.

This formalizing of experiential discovery brings it a step closer to the way scientific discovery works because it lends an element of formality and planning to it. However, they still differ in important ways. E.O. Wilson<sup>13</sup> has stated that in science a discovery does not exist until it has been reviewed and safely is in print. In other words, publication is an integral part of the scientific process. Dr. Davidoff mentioned the importance of published findings, as proof and lack of disproof lie at the heart of the logic of science. Therefore, only full and open publication provides the type of access and transparency needed to exercise that logic. Thus, the action cycle in scientific discovery and experiential discovery would be “plan, do, study, publish.” What is missing from the experiential learning action cycle, since experiential learning is performance- and action-oriented, is the publishing step. What are the implications of the difference between these two types of cycles for quality improvement? It suggests the need for new action cycles for quality improvement. A new generic cycle might be “plan, do, study, act/disseminate.” For experiential learning, it might be “plan, do, study, act/teach/coach.” Finally, for experiential discovery, to learn what works better in a more formal way, it might be “plan, do, study, act/publish/discuss. The failure to publish reports of new knowledge from experiential learning could result in the following consequences:

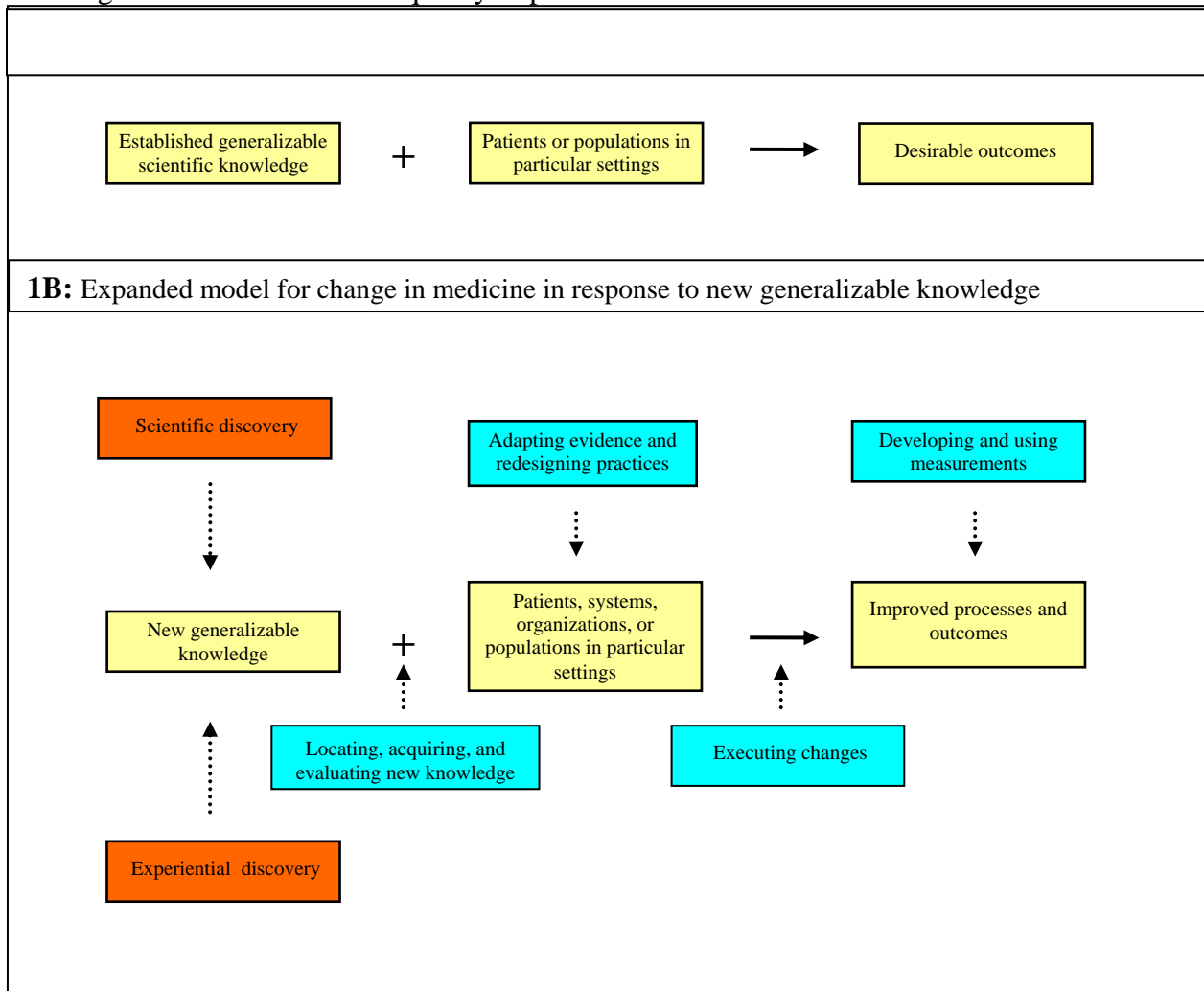
- Makes establishing repeatability difficult.
- Prevents public scrutiny and accountability.
- Reduces the incentives and opportunities to clarify thinking, verify observations, and justify inferences.
- Slows the spread of known improvements.
- Inhibits the discovery of innovations.
- Ethical issue: fails to give information or a product back to the public.
- Limits the influence of publications on QI.

We usually think of scientific discovery shaping publication, with editors and journals being passive recipients of discoveries, but, as Dr. Davidoff explained, publication also shapes scientific discovery. He described some examples, in addition to peer review, of the publication process's reciprocal influence on discovery. There are new requirements for authors to have control of data and the decision to publish. A new proposal is for randomized clinical trials to be registered in formal, nonprofit registries. Another example of the way publication "pushes back" is the development of publication guidelines, such as the general guidelines developed by the International Committee of Medical Journal Editors. Their "Uniform requirements for manuscripts submitted to biomedical journals" is a framework that sets the tone that there can be a standardized way of representing one's work in print. More recently, specific guidelines have been developed for studies done using a particular study design or content areas. The first was the CONSORT (Consolidated Standards of Reporting Trials) guidelines for reporting randomized trials. Some additional functions have flowed from these guidelines documents. Explication of CONSORT guidelines has an educational function, not only on writing up research, but also on planning and conducting the research. When articles are reported in a consistent way, it potentially may make aggregating studies into higher levels of analysis easier. The next question is: is there a way, in parallel, that publication can shape quality improvement? Recently, there was a set of guidelines developed by Richard Thompson and Fiona Moss for quality improvement reports published in *Quality and Safety in Health Care*, and these guidelines were adopted by *BMJ*. These are particularly relevant for case-report types of articles. For more formal and complex types of studies, Drs. Batalden and Davidoff developed a new version of publication guidelines for quality improvement reports. (These were included in the symposium binder and have since been published in *Quality and Safety in Health Care* (2005, Oct 14(5):319-325).) These guidelines follow the IMRaD format, with the learning cycles that take place during the study being reported in the results section. Dr. Davidoff proposed that these guidelines have limitations in that they are only appropriate for certain kinds of quality improvement work. The strengths of the guidelines are that they potentially may have educational value as well as a positive influence on the planning, funding decisions, and editorial processes for QI studies.

To summarize, science-based health care is extraordinarily complex and trying to improve it is like trying to change the tire on a car while the car is running in the Indianapolis 500, but the process of improvement needs to be built in to health care. Because improvement is so difficult, unless improvement becomes an integral part of the health care process, it will be very challenging to make it work and to sustain it. Drs. Davidoff and Batalden have proposed an integrated model including three kinds of learning: scientific discovery, experiential discovery, and experiential learning, with the last having a key role. Experiential learning is uniquely suited to quality improvement; however, it is difficult to study using traditional methods. Experiential discovery, or learning about learning and publishing the results, is one important way to link experiential learning into the body of scientific discovery. Finally, dissemination of results is an essential element of scientific discovery, thus it must be part of improvement, whether through teaching and coaching or publication and discussion. Publication guidelines can help shape quality improvement via their reciprocal interaction with the study methods used, the completeness of reports, funding decisions, and the potential for aggregating results.

**Figure 1**

An integrated model of medical quality improvement



**Discussion**

Lawrence Green, Dr.P.H., Adjunct Professor, School of Medicine and Comprehensive Cancer Center, University of California at San Francisco, noted how well Dr. Davidoff had set the stage for the rest of the symposium. In particular, Dr. Davidoff described the complexity of the environment in which quality improvement interventions are conducted. Dr. Green posed the question, “how ‘out of control’ can we let controlled trials become?” We cannot impose so much control that trials are not representative of the environments we hope to understand. Dr. Green noted his mantra is “if we want more evidence-based practice, we need more practice-based evidence.” This means that we need to find a place for practice-based quality improvement studies that will inform evidence-based practice. Dr. Green would add the practice of engaging practitioners in participatory research, putting scientists in practice settings, and engaging patients in the research enterprise to the set of tools that Dr. Davidoff recommended.

In the discussion following Drs. Clancy, Davidoff, and Green's remarks, participants made a number of cogent comments:

- Several areas of expertise that could be brought to bear on this discussion include the study of knowledge management and tacit knowledge as well as social network research. The potential for partnerships with science agencies is great and should be actively pursued.
- QI and performance improvement are going to occur in the health care system whether or not they are studied. The key component is not the researchers or scientists, but the organizations and the people they serve.
- Health services research (HSR) brings in nothing of the phenomenon being studied in that the principal theory used is statistical theory. We send subjects down two arms of a study and record a few measures at the end. This research model misses out on the role of experience and its effect on performance. It will be important for us to deal with a different way to learn throughout this symposium's sessions.
- Experience must be linked to reflection and feeding back on performance.
- Experience can lead to discovery or learning, but learning from one experience could be termed superstition. We need to be careful about experiential learning that draws an inaccurate conclusion and thus does not generate knowledge.
- There are many ways that experiential learning could be amenable to empirical study. One could use the critical incident technique used in nursing and salesmanship, or the technique of eliciting theories in practice from highly expert people in their fields. Classification of phenomena is, of course, part of science.
- We need to figure out what are the appropriate tradeoffs to the stakeholders. We need to think about research in the context of the people who are using that research. We need many different ways to proceed with research. Dr. Davidoff responded that we can short-circuit the dilemma that arises from the long time span between a study's findings and their publication. He has been emphasizing peer-reviewed publication, but there always have been alternatives. When all publication was print, it was called the "gray literature." And there are venues for "works in progress" such as meeting abstracts. Nowadays, it is a totally different universe with electronic publishing on the Internet and people putting out working papers, white papers, and other kinds of reports, but a better term for this is *dissemination* rather than publication.
- A participant asked how dissemination relates directly to organizations. Organizations are concerned about spread, or how to spread a change concept within the organization, even just to get an idea from one department to another. How do you integrate that into Dr. Davidoff's model so ideas get back to the organization and are not just known among researchers? Dr. Davidoff responded that the publication process can enter into the

process, even if it is internal publication, but “spreading” means getting the word out and it involves social processes among other things. That process is so much more complicated than doing individual projects. Dr. Davidoff acknowledged that he does not know that literature too well, but that there is a literature on spread within the VA system which describes various dimensions of the process in which there are social, administrative, and management issues that do not exist in smaller, community-based improvement projects. Ultimately, the improvement process involves making plans, doing trials, learning from experience, and modifying the efforts. Some organizations are just bigger and slower to move because of their scale.

- We could learn from the treatment of pediatric cancer. Pediatric oncologists enroll every single patient in either a randomized trial or an observational trial. In that way, every patient’s treatment is recorded in a massive database from which there is constant knowledge being generated from every case. If we used this same concept in QI projects, we would have a powerful tool for sharing experiential learning across institutions.
- The publication of negative QI projects is very important, as we have found from clinical trials where negative findings are hardly ever published. A significant portion of experiential learning comes from failure. An important question for QI is what is the gold standard to know whether a technique is yielding the truth?
- In managed care quality improvement, most projects have no publication pressure, and spread may be just within an institution. This is where most of the unproven and unpublished knowledge exists. The question is: how do we change this? The quality improvement directors who could say whether something worked or not are not in attendance at this symposium. We need to put dissemination pressure on the system to bring out the existing knowledge so that it can be peer-reviewed, disseminated, and used by the rest of the health care system. Dr. Davidoff responded that this point gets to the issues of incentives and rewards and intellectual gratification.

### **Wednesday, September 14, 2005**

#### **Meeting Purposes and Process Revisited**

Thomas Chapel, M.A., M.B.A.

Senior Health Scientist, Office of the Director, Office of Strategy and Innovation, Centers for Disease Control and Prevention

Denise Dougherty, Ph.D.

Senior Advisor, Office of Extramural Research, Education, and Priority Populations,  
Agency for Healthcare Research and Quality

Dr. Dougherty identified the goals for the meeting:

- To review a range of quality improvement interventions (QIIs) and the critical questions that arise in evaluation of these interventions, including basic questions of internal (does it work in the setting and for the condition in which it was tested) and external (will it work in other settings and conditions) validity.

- To identify the strengths, weaknesses, and tradeoffs of alternative designs and methods for evaluating QIIs.
- To achieve a working consensus about the range of traditional and innovative designs and methods that can be used to answer key QII questions.
- To identify and suggest strategies to facilitate possible changes in funding mechanisms, review processes, research and publication standards, and research training that could help accelerate the development and spread of reliable QII research methods.

Dr. Dougherty asked the group to think about what it will take to develop more designs, to develop more information about designs, and to make a broader range of designs acceptable to the scientific community. Further, she asked participants to identify changes to broaden the field of evaluation design in terms of review processes, funding mechanisms, publication standards, research training, implementation, and designs and methods themselves.

It is difficult to find a standard definition for QI, especially when considering both health care and public health, and a range of QI strategies exists. The criteria for the definition chosen by the core planning group are that the QI strategies are implemented in “real-world” settings; are used to expand the delivery, reach, and impact of evidence-based interventions at the population level; and include health care and public health interventions. Some examples are policy, organization, system changes, practice design, and strategic linkages to community programs and policies. The breakout sessions scheduled throughout the two days of the symposium were designed to enable a variety of professionals with different backgrounds and specialties to discuss the challenges facing QI designs and ways to improve the science base.

The overarching goal is moving forward to improve the science base for quality improvement interventions and evaluation in the spirit of a quote from Richard Grol:

*The challenge for the years to come is to design strategies for quality improvement that ... step from anecdotal evidence for those strategies to systematic evaluation in order to distinguish between faith and fact in the field of improving care.*

## **SESSION I. CASE STUDIES: QII AT THE CLINICAL MICROSYSTEM LEVEL** **QII to Increase Delivery of Clinical Preventive Services in Office-Based Primary Care**

Lawrence Fine, M.D., Dr.P.H., Chair/Moderator

Leader, Scientific Research Group on Clinical Prevention and Translation, National Heart, Lung, and Blood Institute, National Institutes of Health

### ***QII in the Practice Partner Research Network: Group Randomized Trials and Other Designs***

Steven Ornstein, M.D.

Professor of Family Medicine and Director, Practice Partner Research Network, Medical University of South Carolina

Dr. Ornstein presented two Practice Partner Research Network (PPRNet) studies. The Translating Research Into Practice (TRIP II) study, which was a group-randomized trial, and the Accelerating TRIP in a Practice-Based Research Network (PBRN) project (A-TRIP), which is a

time series study currently underway. Dr. Ornstein briefly presented the results and then discussed strengths and weaknesses of each project and lessons learned that might be applicable to other QI studies.

PPRNet is a practice-based learning and research organization designed to improve health care in its member practices first and then throughout the United States. PPRNet comprises interested users of the Physician Micro Systems, Inc., Practice Partner Patient Record, an electronic medical record (EMR); consultants and collaborators; and research offices at the Medical University of South Carolina, Charleston, South Carolina. The organization includes 101 practices and 502 clinicians in 37 States.

Data are collected and entered into EMRs at sites that use the PPRNet EMR system. Data are extracted and sent to the vendor, who in turn submits the completed data to Dr. Ornstein's group. His group then develops practice reports. The motto of the PPRNet is "to blur the distinction between quality improvement and research" meaning that when they work with practices, they are a quality improvement organization, in that most practices are not overly interested in research, and, to funders, they are a research organization.

*Translating Research Into Practice (TRIP) II: Primary and Secondary Prevention of Cardiovascular Disease and Stroke in Small Primary Care Practices.* This group-randomized trial was funded by AHRQ and the results were published in *Annals of Internal Medicine* (2004;141:523-532). The study ran for two years in 20 non-academic family practice and internal medicine practices and included 87,291 patients. The 10 control group practices received quarterly practice-level performance reports on 21 indicators of care. These indicators were based on guidelines from the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure VI; the National Cholesterol Education Program Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults; the American College of Cardiology/American Heart Association; and the American Diabetes Association. Each practice had EMRs and they organized approaches to improvement on their own.

Practices randomized to the intervention group received quarterly practice performance reports and six to seven practice site visits to obtain descriptive data and to facilitate QI via the use of participatory planning, EMR tools, complexity science approaches, and best practices. They also participated in two network meetings in Charleston to share best practices approaches. Quantitatively, practice-level analyses showed improvement in both intervention and control group practices in the percentage of clinical targets reached, although the intervention group had greater improvement than the controls in 18 of 21 indicators. The patient-level analyses showed that there was improvement in intervention and control subjects, but improvement in the intervention group was greater than that in the control group for only 2 of the 21 measures. The qualitative analyses showed that the most successful practices made quality a priority, involved the entire staff in the effort, redesigned the office, made efforts in patient activation, and used their EMR.

Several challenges had to be faced in this study, particularly in the area of QI "buy-in." Some providers in larger practices did not believe in the study for various reasons, such as not

believing the information in the practice performance reports, or having competing demands, or having a perceived lack of self-efficacy. The response of the investigators was to focus on the more amenable members of a practice and have them model change for others. This had varying levels of success.

In addition, they found their initial concept for the practice visits was misguided. They thought that if they taught providers the practice guidelines and how to use their EMRs correctly, that that would be a panacea. However, most physicians knew the guidelines and wanted to use EMR in their own idiosyncratic ways. The research team's response was to change the focus of site visits to encouraging QI approaches at the microsystem level, meaning that they helped practices in the way that they wanted to be helped.

In planning the study, Dr. Ornstein's group had not recognized the importance of non-provider staff to the organizations studied. As they implemented the study, the team's response was to encourage non-provider staff to play a greater role in participatory planning and implementation, as well as to participate in a second network meeting.

The lessons and conclusions from the TRIP II study were:

- 1) Clinicians will volunteer to participate in these activities, particularly those who have EMR systems that facilitate QI reporting and interventions.
- 2) Clinicians will participate in interventions that they deem beneficial to their patients.
- 3) The EMR is not a panacea, and a more robust quality improvement intervention model is needed.
- 4) Simply giving practices the information (academic detailing) and the tool (EMR) is insufficient.
- 5) "One size does not fit all": a) intervention approaches and emphases have to be customized at the microsystem level; and b) study sections accustomed to specific protocols that require rigid adherence may need to appreciate that customization is needed.

*Accelerating TRIP in a Practice-Based Research Network (PBRN) project (A-TRIP).* A-TRIP is an effort to enhance PPRNet practice performance reports to include approximately 80 indicators in 8 discreet clinical areas. This study is an AHRQ-funded expansion of the TRIP-II study. This is a demonstration project with descriptive and time-series evaluation components. The project involves more than 100 non-academic family practice and internal medicine practices and there will be approximately 500,000 patients included. The intervention methods are practice performance reports, practice site visits every six months in practices that want them, and participation in network meetings in practices that want to attend. They are examining a broad spectrum of clinical indicators, such as 13 related to diabetes mellitus and 21 related to heart disease and stroke. In their practice performance reports, they use comparison with national benchmarks as well as statistical process control methodology to let practices know when they make an improvement. The study is ongoing and the results are preliminary, but there has been greater than expected participation in practice site visits and less than expected participation at network meetings. Preliminary data on practice-level improvement are encouraging.

A-TRIP challenges and resolutions include:

- 1) Participants wanted patient-level reports in addition to practice-level reports to better identify those in need of specific interventions.
- 2) 10% practice attrition annually will challenge data analysts; recruitment of new practices has been ongoing; “duration of exposure” will be included as a variable in the analyses.
- 3) Lack of any control group may compromise the validity of the findings. The hope is that looking at a broad range of indicators and for a large effect size will be considered sufficient evidence of an effect.

A-TRIP lessons to date include:

- 1) Physicians will participate in such a project a) when they receive a tangible benefit, such as free practice reports or continuing medical education (CME) credit; b) when they believe the project is in the best interest of their patients, and c) when they can titrate their levels of involvement to suit their needs or level of interest.
- 2) Non-provider staff are key to project success. One challenge is that their training is variable and they need proper supervision, focused training, and inclusion as a respected team member in QI planning activities.

Dr. Ornstein’s recommendations for future work were that:

- 1) Practice leaders, either physicians or office managers, need to be developed so as to better incorporate these individuals in QIIs. Research is needed on the best approaches for this.
- 2) In addition to clinical outcome measures, studies need to look for the behavioral/organizational changes that take place as a result of different intervention strategies so approaches can be tailored efficiently to the needs of specific practices.

### ***A Multimethod Tailored QII for Sustainable Practice Change***

Mary Ruhe, R.N., B.S.

Project Coordinator for the Enhancing Practice Outcome through Communities and Healthcare Systems (EPOCHS) Project, Department of Family Medicine, Research Division, Case Western Reserve University

& Kurt Stange, M.D., Ph.D.

Professor, Family Medicine, Epidemiology and Biostatistics, Oncology, and Sociology, Case Western Reserve University

Ms. Ruhe focused on the Study to Enhance Prevention by Understanding Practice (STEP-UP)<sup>14,15,16</sup> series of interventions, part of a line of inquiry in collaborating practice-based research networks (PBRNs). This group’s collaboration is based on the premise that understanding practice is important to successful intervention efforts. Two conditions, pursuing a line of inquiry and having collaboration among PBRNs, offer a fertile ground for conducting QIIs. The line of inquiry started with observational studies, moved to intervention studies, and returned to observational studies. Their studies have been funded by AHRQ, NCI, NHLBI, and the Robert Wood Johnson Foundation.

The theoretical framework for these studies is the competing demands/competing opportunities theory, which states that many worthwhile services compete for time on the agenda of primary care patient visits, and that when primary care clinicians are not performing an activity under

scrutiny, they may be doing something else that is more compelling. Creating change in the primary care setting may be enhanced or inhibited by the charge to primary care practices of offering integrated, prioritized care within an ongoing personal relationship with patients. The basic premise of this research is that understanding practice is a process involving learning before, during, and after an intervention.

The STEP-UP study was a group-randomized trial in 77 practices which involved individualized interventions based on the multimethod assessment process (MAP). Control groups get a (refined) delayed intervention with pre/post evaluation. MAP involves observation of practice operations and patient visits, key informant interviews, and practice genograms, a tool for understanding practices which depicts the structure and relationships within primary care practices. Understanding practices includes understanding key stakeholders and their motivations, promoting current approaches to preventive service delivery, and understanding levers for change and the practice's capacity to change. The intervention was tailored to each practice and feedback on rates of preventive services delivery were given to each practice.

The rate of improvement was variable across practices, ranging from 31 to 43% improvement in preventive services delivery rates. Improvements were sustained during a 24-month follow-up period. The most substantial variability was in health habit counseling. Thus, the study demonstrated that a tailored QII can have a variable but sustained effectiveness, even in a changing health care environment. However, greater individualization of intervention approaches, based on a greater understanding of practice variation, is needed.

STEP-2 was a refined QII among the STEP-UP control practices that did not show an increase in preventive services delivery. Their stasis was attributed to the practices having been given more choice in where to focus their attention.

The lessons learned from the STEP-UP studies led to the group's newest study, EPOCHS, which stands for Enhancing Practice Outcomes through Communities and Healthcare Systems. This study is a randomized controlled trial (RCT) of 30 primary care practices in three systems and includes engagement of resources from the practices, the health care system, and community organizations.

The bottom line throughout the entire line of inquiry was the need to understand the practices as complex adaptive systems in which complex behavior emerges from relationships among agents, simple rules and recurrent patterns exist, initial conditions are important, and co-evolution of the organization is non-linear. Grounded in this framework, a reflective quality improvement process offers insights into the practice change process. For example, understanding practices' vision and mission is useful in guiding change. Tension and discomfort are essential and normal during change.

Ms. Ruhe indicated that some of the lessons learned were that:

- 1) Change is difficult to predict, but a practice being at the "edge of chaos," meaning stressed but not set in its ways, facilitates practice change.
- 2) It is important to tailor facilitation over time to optimize emergent opportunities and malleable moments of readiness to change.

- 3) Motivated change agents are important. Once motivation exists, instrumental needs can be addressed.

In summary, efforts to improve practice should be preceded by efforts to understand practice. Primary care practices are complex adaptive systems. The implications of a complexity science perspective are that relationships are critical, learning is more important than knowing, and problems cannot be solved by muscle, but instead require creativity and improvisation.

***Comments: Rigor and Relevance***

Kurt Stange, M.D., Ph.D.

Professor, Family Medicine, Epidemiology and Biostatistics, Oncology, and Sociology, Case Western Reserve University

In Dr. Stange's talk, he commented on: 1) Dr. Ornstein and Ms. Ruhe's presentations; 2) how we think about the problem of rigor versus relevance; 3) how we approach working on the issues of quality improvement; and 4) the problem of how we synthesize knowledge.

Although the papers by Ornstein and Ruhe represent the work of two independent research groups, they have a lot of commonalities in their approach, likely because there are common stimuli that have led them to similarities in how they approach their interventions and their evaluations. Both presentations worked with real-world primary care practices that are characteristic of the industry. The business and reimbursement model for primary care practice does not match QI intervention goals. To keep afloat financially, most primary care visits need to be kept to less than 10 minutes and most practices have had to reduce their number of qualified staff to maintain the bottom line. There is a mismatch between the staff available and the complexity of the competing demands for optimizing care.

Both of these studies worked in practice-based research networks. Whereas most quality improvement work aims to reduce variation, the work presented here was done in a manner that is designed to promote desirable variability as well as reduce variation in the delivery of evidence-based services. The desirable variability reflects local adaptation.

The research was done in a peer review and funding environment that emphasizes the RCT and single disease foci. In contrast, the intervention approaches that both presentations described were multifaceted, emphasizing multiple processes and tools, and addressing a diverse set of outcomes.

How should we optimize this package? In light of the competing demands theory described by Ms. Ruhe, it is important to look broadly so that improvement in one area is not occurring at the expense of a deficit in another. We need to think about the commonalities of what the systems were trying to optimize when trying to improve multiple outcomes. Both presentations addressed the practice level, but with a direction to begin to include elements of the health care system level and community level. Both studies individualized shared best practices via outside facilitation and consultation, and also with mechanisms to facilitate shared learning within a practice and across practices using complexity science principles. Both presentations described evaluations with mixed methods designs which included concurrent qualitative evaluation to

understand the process of change and to understand individuals' learning processes. This information fed back into the intervention.

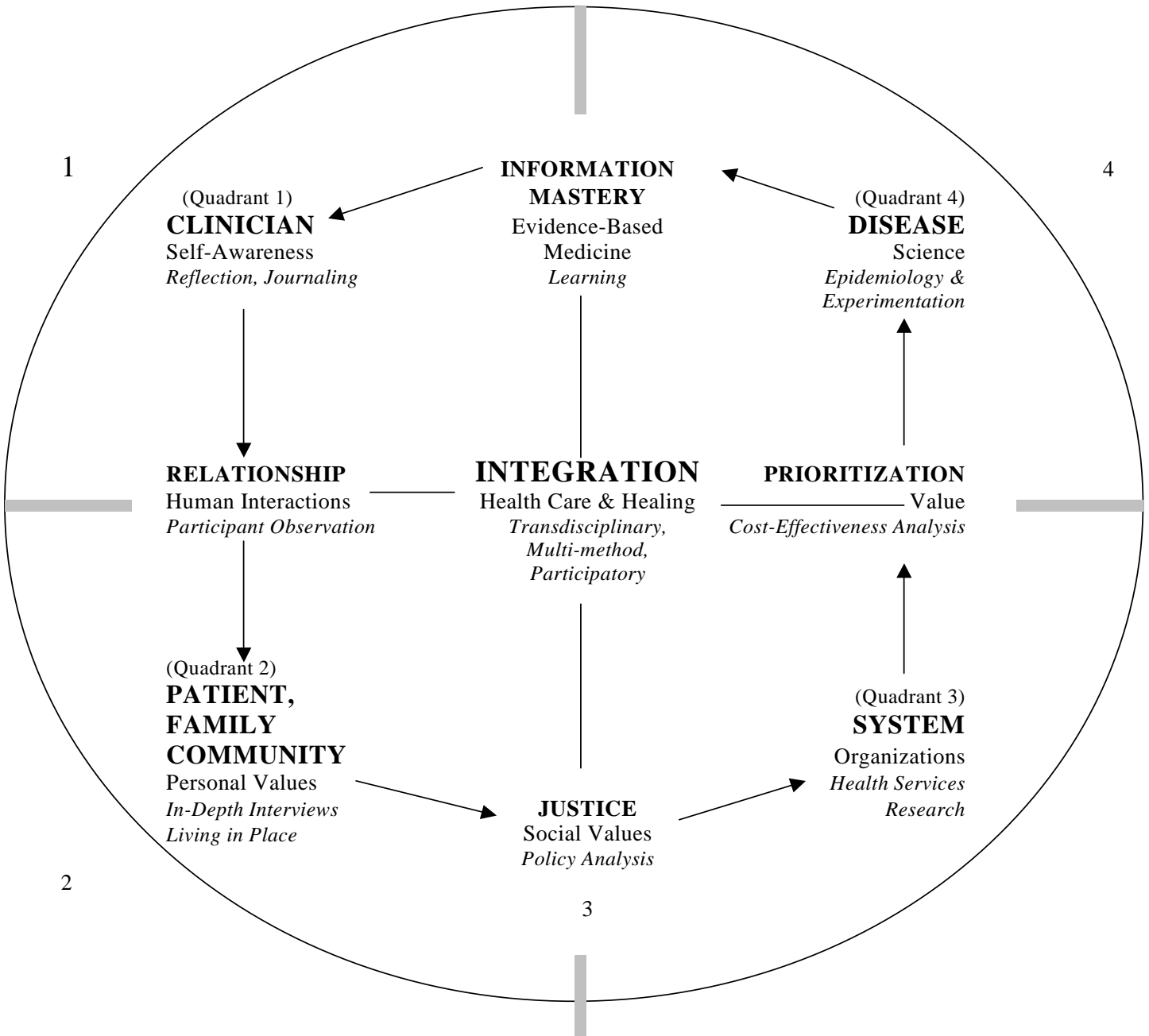
Dr. Stange's understanding of the need for the conference is that we feel some angst because our theory and methods and worldview do not match the problems we are addressing. The fundamental problem is that our ways of thinking, our ways of knowing, and our methods are good at isolating a phenomenon from its context. Yet, we believe that context matters. There are four different ways of "knowing" about health and health care<sup>17,18</sup>:

	Inner Reality	Outer Reality
Individual	"I"	"It"
Collective	"We"	"It"

When applied to the health care system, individual outer reality would be the study of disease and treatment and collective reality is about systems such as health services research.

When we do research we tend to focus on one way of knowing at a time, but we need to remember that other ways of knowing are always there. There also is a need to synthesize these ways of knowing (Figure).

Figure. Multiple ways of knowing.<sup>19</sup>



For each item, bold capitalized words on the first line signify "FOCUS OF KNOWLEDGE," normal text on the second line signifies "Task of Understanding," and italicized words on the third line signify "Mode of Inquiry."

One way of honoring these different ways of knowing is to readily acknowledge when we plan interventions and when we interpret the findings that there are different perspectives.

It is important to consider other ways of knowing even when we are working within one way of knowing. One way is to use transdisciplinary approaches in whole systems that involve collaboration. Thomas<sup>20,21</sup> described QI needing both “top down” and “bottom up” leadership. “Top down” leadership involves working with systems and addressing power structures. “Bottom up” leadership means involving the perspective of people on the front lines. To do whole system collaboration in QIIs, Dr. Stange suggested using three different forms of collaboration:

- Multidisciplinary—Multiple disciplines contribute their individual piece to solving the problem. Multiple experts can do this through an edited book or separate presentations.
- Interdisciplinary—Interdisciplinary research can focus on a conversation between and among disciplines, with both working together to solve a common problem.
- Transdisciplinary—Transdisciplinary research is a sustained conversation across and beyond disciplinary boundaries that creates a new, shared language.

It is critical to think about where transdisciplinary teams should be developed. Bringing together research and development would help overcome problems currently faced in translating research into practice. Three models mentioned for research and development and QI are integrated health care systems where one can look at the population of enrollees all at once such as the Health Maintenance Organization Research Network, the NIH Research Center Model, and PBRNs. PBRNs are affiliated practices that are primarily devoted to patient care, and that engage front-line wisdom to develop questions, gather data, and interpret and implement the findings. These networks have more generalizable patient populations than the typical settings for research. Community-based participatory research (CBPR) is parallel to PBRNs and has three characteristics: collaboration, mutual education, and acting on the results that are relevant to the community.

Fostering a multimethod approach that integrates quantitative and qualitative methods is the way to move the field forward by allowing us to understand the meaning of a study and what its generalizable lessons are. The strength *and* weakness of quantitative methods is that they isolate a phenomenon from its context. Qualitative methods are good for helping one understand context.

A complexity science perspective makes us think about the relationships among agents in a system. Simple rules can be used to describe the components of complex behavior. Different parts of a system co-evolve, meaning that if we are studying practices, we need to look at the communities in which they exist because the systems will co-evolve. We need to look at where a system started before we intervene.

Greenhalgh<sup>22</sup> states that the next generation of intervention research needs to be theory-driven rather than by thoughts about how to disseminate a particular package. It also is important to look at ecological context while pursuing a multidisciplinary and participatory approach.

In conclusion, Dr. Stange asked, “How do we get started?” He suggested these approaches:

- We need to work at multiple levels of a system.
- We need to consider different ways of knowing.
- We need to pursue development alongside research in QI.
- We need to foster shared learning among participants in research.
- We need to develop participatory relationships that transcend single projects.

Dr. Stange recommended that quality improvement be pursued not as single projects but as lines of inquiry. Thus, we need to integrate quantitative and qualitative research either sequentially or simultaneously. With incremental approaches, we are just part of the problem, enabling the current dysfunctional system. Instead, we should address how we can transform the system.

## **Discussion**

Discussants noted the emphasis was on learning how we are going to learn rather than on studying the strength of the interventions. QII research to date suggests that the design of the intervention is where things are lacking. Do the constraints of the study design result in weaker interventions? In response to a comment about a lack of science on the intervention side compared to a lot of theory, but a robust use of science on the evaluation side, Dr. Stange agreed that methods could be restrictive and problematic, but robust interventions build in iterative review cycles. He countered that a great deal of science is used on the intervention side.

Another participant asked whether there would be a model self-sustaining microsystem, and how will people be directed to get there? Dr. Ornstein noted that there are practices that need to hear information one time only and then can implement it. There are other practices that after repeated exposure to an intervention fail to make changes.

## **QII to Increase Timely Delivery of Surfactant to High-Risk Newborns During Hospital Labor and Delivery**

David Atkins, M.D., M.P.H. (Chair/Moderator)

Chief Medical Officer, Center for Outcomes and Evidence, Agency for Healthcare Research and Quality

### ***QII Case Study: Surfactant Use in Preterm Infants***

Gautham Suresh, M.D., D.M., M.S.

Assistant Professor, Division of Pediatrics, Medical University of South Carolina  
& Laura Leviton, Ph.D.

Senior Program Officer, Department of Research and Evaluation, Robert Wood Johnson Foundation

Drs. Suresh and Leviton presented the findings of a cluster-randomized trial on the timing of surfactant use in preterm infants for which the principal investigator was Dr. Jeffrey Horbar.<sup>23</sup> This study was funded by AHRQ and the results were published in *BMJ*.

Prematurity is a common problem with infants, and respiratory distress syndrome (RDS) is one of its most common morbidities, primarily from the lack of endogenous surfactant. The development of exogenous surfactant therapy to treat RDS was a significant advancement in the field, leading to decreases in mortality and morbidity. Randomized trials have shown that surfactant therapy is best used on a prophylactic basis within the first two hours after birth, before the infant becomes too sick. According to studies by the Vermont Oxford Network, only 19% of premature infants receive the first dose of surfactant less than 15 minutes after birth, and 27% received surfactant more than two hours after birth.

The trial was conducted in the Vermont Oxford Network which is a network of 500 neonatal intensive care units in North America. The mission of the network is to improve the quality and safety of care for newborn infants and their families through a coordinated program of research on improvement. The network has an ongoing process of data collection and no additional data collection was needed for the sake of the trial. The network provides its units with quarterly and annual reports on outcome. A multidisciplinary team of neonatologists, outcomes researchers, statisticians, practice improvement experts, evaluation experts, and behavioral scientists worked together to prepare for the trial. Before the trial, focus groups met to refine and customize the intervention's design. These groups used the PRECEDE framework and included neonatal practitioners not affiliated with the network. The focus groups' input helped refine the intervention. Out of 355 eligible hospitals, 114 were enrolled and split into 57 intervention and 57 control sites. The network provided good baseline data at the site and individual patient levels. The chief component of the intervention was a 2-day workshop consisting of multidisciplinary teams (physicians, nurses, and respiratory therapists from each hospital) who were taught the principles of evidence-based medicine, the effectiveness of surfactant, and the importance of early administration, as well as the principles of quality improvement. The groups were not told what to do. Instead, they received data about their hospitals' performance in comparison with the body of evidence and data on other hospitals in the network. Teams were told, "you decide what to do" in terms of interventions. The teams were asked to set aims and refine their aims after conferring with others at their hospital. Ongoing support was provided through a listserv and periodic conference calls.

Dr. Leviton noted that there is a wealth of information about the ecology of neonatal intensive care units and their logistics, but the focus here is on outcomes. Dr. Leviton noted that a cluster randomized design, which is known in other fields as a "place-based randomized experiment," was used in which neonatal intensive care units (NICUs) were randomized to treatment or control. The units for measurement and analysis are nested within other units. They chose the infant and NICU levels for direct study because the NICU staff essentially treats the infants as a team; therefore, the individual practitioners were less relevant for direct study. It was important to directly study the infants as evidence of the practitioner decisions about individual infants. Referral systems were indirectly studied because many infants were born in referring hospitals and transported to tertiary care centers. The researchers expected differences between in-born and out-born infants, thus they prospectively planned analyses of all infants combined and separate analyses for in-born and out-born infants.

The proportion of infants 23–29 weeks old receiving surfactant in the delivery room was significantly higher in the intervention group (55%) than the control group (18%). Dr. Leviton

contrasted the effect size found here, a 37% difference, with those cited in a review of multifaceted QI interventions, which were in the range of a 2-9% difference. Thus, the effect size seen in this study was much greater than typically is seen in multifaceted QI interventions. Treatment subjects were less likely than controls to have surfactant administered more than two hours after birth. However, no differences in infant mortality or pneumothorax were found between treatment and control infants. There could be many reasons for this, but the most plausible is a lack of statistical power. As often happens in prevention trials, the world had changed such that when this study began, corticosteroid therapy for women in premature labor, which matures the organs of the fetus, had become prevalent and is a possible explanation for some of the findings.

“The road not taken” in methods choices precluded other things being done. They chose to “go broad” for a causal test of an organization-level intervention rather than to “go deep” to understand the intervention within organizations. Basically, this was a resource-allocation decision. When deciding whether to go broad or go deep, one must consider whether one is looking for causal information. If you do not have a mature intervention, it does not make sense to perform a causal study. The consequence of “the road not taken” is that the research team did not have enough data to examine the mechanisms by which these improvements occurred.

The team has some attitudinal data for the participants showing a high level of endorsement for the need for a practice change. They would like to better characterize and classify the rapid-cycle-improvement cycles that the NICUs underwent and know why the evidence was persuasive to the participants in the workshop and how they made their commitment to change. They would also like to investigate whether features of the NICU environment are associated with the degree of change in the institutions.

The take-away message from this study is not that “they had everything going for them (meaning a large sample size and a receptive audience) and that is why they could do a cluster randomized trial.” Symposium participants should take away that one must make deliberate choices in setting up studies. We want to use the most rigorous methods we can, if it is possible and if a causal question is being asked. One would not want to do a randomized trial prematurely. Two key points from this presentation are:

- Research networks help and can keep costs down.
- Strong cases can be made for group randomized designs, and they can provide information about the organizations of interest.

## **Discussion**

A participant asked 1) why the presentations did not mention the economic implications of their interventions, and 2) how their process evaluation was conceptualized. The idea of mediators and moderators is important when considering causal pathways for how interventions achieve their effects. Concerning economic implications, Dr. Suresh responded that the grant was for \$1.3 million for a period of 3 years. The dilemma for effectiveness was the absence of improvement of clinical outcomes, while the main improvements observed were in the process measures where the time from birth to administration of surfactant was decreased. Dr. Leviton added that Canadian colleagues had to judiciously use surfactant because it is very expensive.

Concerning mediators and moderators, the process variables that they collected included observation of workshop participants and their deliberations, their stated aims, content analysis of recordings of conference calls, and review of listserv discussions of problems. All of this material can be characterized, as can hospital characteristics. A secondary analysis is underway to see if this information has any explanatory power.

After a basic cost analysis, Dr. Ornstein's group found the costs were lower in the intervention group than in the controls. This difference is thought to be based on efficiencies and new processes resulting from the intervention. However, this cost analysis was primitive, but highlighted some interesting things. Dr. Stange responded that inductive analyses of the qualitative data led to some interesting findings about mediators and moderators. For example, some of the practices that seemed like failures had made substantial improvements, but not in the areas for which outcome measures were recorded.

A participant remarked that it is unclear whether we know what questions need to be answered to be useful in the field. Do the participants know enough to show that some QIIs work without knowing what is in the "black box?" He also asked whether we know enough about the substance of the changes and what changes in process have been put in place. Or, do we need to learn what are the processes that facilitate change? These are different kinds of questions. Do we need to be conducting research at the level of asking questions such as how do we motivate leadership or how do we work with non-physician staff? Dr. Leviton replied that the dilemma described is between resources allocated to addressing a causal question versus resources allocated to understanding what it is that we are producing when implementing an intervention. We have limited resources for research so we must decide how to allocate resources to an appropriate causal test and methods decisions related to understanding. She regrets that they cannot do more to understand the molar construct, or package of intervention that was delivered, but would make the same resource allocation choice again.

A participant asked if information is available about the spread of the study's findings and if enough is known about what worked to persuade others to examine the results at their respective institutions. Dr. Suresh's group has not studied the spread of the intervention to determine who has applied the findings.

A participant noted that these projects have looked at small microsystems and individual practices that are relatively independent. We need to improve our models of QI involving larger organizations and address how we deal with them.

A participant offered that even when individuals report on the quantitative piece, other things occur that are not being reported and not being discussed because of a lack of training or terminology for some items. The problem for trying to generalize this work is that we are not learning as much as we should from other people because some parts of the process do not get described in publications.

A participant asked if the conference organizers are only interested in quality improvement efforts that are conducted by outside investigators, or are they, and participants, interested in developing a model for an organization that wants to make a change on its own. Another way to

ask the question is: are participants interested in studying change efforts when they (the researchers) do not control the intervention? One would want to create a model for how organizations go about making that change and how they set priorities. This would mean conducting natural experiments, which are abundant, but typically go unstudied. Dr. Atkins responded that the group is interested in finding innovative ways to make those types of changes. Francis Chesley, Jr., M.D., Director, OEREP, AHRQ, said the Agency is always interested in taking advantage of and funding a natural experiment. However, the current research funding cycle does not allow them to catch up with that. In other words, the 7-9 month cycle of peer review makes it difficult to get the timing right to examine natural experiments. Peter Briss, M.D., M.P.H., Chief, Community Guide Branch, CDC, said the planning committee is interested in real-life experiences, and the California tobacco presentation on September 15 will present a longer term look at a series of natural experiments, with less emphasis on the single study approach. Joseph Francis, Jr., M.D., M.P.H., Associate Director, Office of Research and Development, VA, indicated that the VA has an intramural research program and is interested in the aforementioned type of research. The VA has a variety of mechanisms of rapid response funding to take advantage of naturalistic events and other current “hot topics.” However, even with rapid-cycle funding, there are other barriers, such as delays for institutional review board (IRB) approval or the hiring of staff, that often preclude fully exploiting the “natural experiments” that happen within health care systems. Neil Thakur, Ph.D., of the VA, stated that a question in which we are interested is: does doing this type of research better mean we will shorten the time to having interventions ready to broadly implement; or will it not save any time, but rather make the resulting intervention more effective? Maybe taking a long time in the research process is a good thing if it leads to a more effective intervention. David Abrams, Ph.D., Director, Office of Behavioral and Social Sciences Research, NIH, encouraged participants to think about how to combine new technologies with statistical methods presented at the symposium to combat the challenges they currently face. New technologies, such as web-based data collection and the use of Palm Pilots to collect data in real time, will provide new opportunities for the collection of experiential data. He also advocated combining process and outcome measures in new and rigorous ways. Lori Melichar, Ph.D., of RWJF, offered a potential solution. She stated that we should determine what level of evidence we really need when funding a research project. Maybe a communications firm or research firm could query key stakeholders, saying, for example, “We have a project to teach nurses QI interventions. What will it take for you to adopt this program in your hospital? Would it be statistical significance? Do you need to see case studies? What kind of evidence do you need?” This could allow researchers to propose a project of appropriate scale and scope using the methods that will best lead to a project that can have the impact it intends.

A participant encouraged everyone to take the ideas presented so far to their breakout groups for discussion to develop recommendations about levels of evidence and other issues raised.

**SESSION II. RESEARCH EVALUATION DESIGNS FOR QII AT THE HEALTH “PLAN” LEVEL**  
**Case: Expanding and Testing VA Collaborative Care Models for Depression**

Joseph Francis Jr., M.D., M.P.H. (Chair/Moderator)

Associate Director, Office of Research and Development, Department of Veterans Affairs

*Translating Initiatives in Depression into Effective Solutions: Regional Expansion Project*

Lisa V. Rubenstein, M.D., M.S.P.H.

Professor of Medicine, VA Greater Los Angeles Healthcare System and UCLA

Senior Scientist, RAND

Dr. Rubenstein intended to help participants understand this field, which is in development, and how it may or may not apply to areas outside of depression. She proposed that there is an evolution of designs that goes along with the evolution of goals in going from efficacy to effectiveness to quality improvement to routine care. Dr. Rubenstein profiled a series of studies culminating in the Regional expansion of the Translating Initiatives in Depression into Effective Solutions (ReTIDES) study. This is an ongoing project that aims to help practices implement evidence-based treatments for depression care.

Depression efficacy research based on classic randomized trials shows that both antidepressants and short-term manualized psychotherapy are effective. Descriptive studies of the quality of depression care nationally identified cases of low quality of care, disparities, and variations in routine practice. Initial randomized provider behavior change interventions to improve depression care quality focused on knowledge-related barriers through clinician education, screening and feedback, and computer reminders. These single-component interventions, however, did not improve quality of care for this condition. Finally, a series of provider behavior change interventions using multi-component models identified the effectiveness of collaborative care for depression. Collaborative care is a multicomponent model, similar to the Chronic Care Model, that fills the gap between primary care and mental health specialty care using a care manager, who sometimes is a nurse. Collaborative care also activates and educates patients for self-care. Initially, these studies randomized patients and were designed similarly to classic trials. Studies randomized at the patient level showed effectiveness across all major demographic groups, including minorities, adolescents, and the elderly. In addition, to test the effectiveness of implementing collaborative care organizational changes in practices, as would be required to disseminate the effective models, investigators randomized practices (cluster randomized designs) and evaluated the effects of collaborative care on consenting patients in experimental practices. These studies showed effectiveness and cost-effectiveness in large and small, rural and urban, managed care, and other types of practices. Finally, a series of randomized studies evaluated how practices could use QI methods to self-adopt improved depression care and improve practice-wide performance. In these designs, outcomes are measured across representative patients with depression attending the experimental or control practices, independently of their participation in improved depression care—similarly to the way HEDIS (Health Plan Employer Data and Information Set) or other performance measures are carried out.

These studies showed that without evidence-based tools, practices do not impact depression care performance. Incorporating tools from previous studies is associated with perceptible, but















































































## *Symposium Agenda*

### **September 13, 2006 (evening)**

5:30 p.m. Registration  
Lower Level Lobby

5:30–6:30 Networking Reception at Wyndham City Center Hotel  
City Center Ballroom

6:30–6:50 Welcoming Remarks  
City Center Ballroom  
*Facilitator:* Thomas J. Chapel, M.A., M.B.A.  
*Welcoming Remarks:* Carolyn Clancy, M.D.

6:50–7:00 Planning Group and Co-Sponsor Self-Introductions  
City Center Ballroom

7:00–7:20 Dinner  
City Center Ballroom

7:20–8:45 Setting the Stage  
City Center Ballroom  
*Moderator:* Lawrence W. Green, Dr.P.H.  
*Remarks:* Frank Davidoff, M.D. (7:20–8:00)  
*Brief Response:* Lawrence W. Green, Dr.P.H. (8:00–8:05)  
*Discussion:* (8:05–8:45)

8:45 Brief Statement of Meeting Purpose/Housekeeping Details  
City Center Ballroom  
Denise Dougherty, Ph.D.

9:00 Adjourn for the Evening

### **September 14, 2006 (8:00 a.m.–5:15 p.m.)**

7:30 a.m. Registration and Continental Breakfast  
Lower Level Lobby

8:00–8:30 Meeting Purposes and Process Revisited  
New Hampshire Ballroom  
Thomas Chapel, M.A., M.B.A.

8:30 –12:15 p.m. SESSION I. CASE STUDIES: QII AT THE CLINICAL  
MICROSYSTEM LEVEL  
New Hampshire Ballroom

8:30–9:30 QII to Increase Delivery of Clinical Preventive Services

## in Office-Based Primary Care

*Chair/Moderator:* Lawrence J. Fine, M.D., Dr.P.H.

*Presenters:* Steven Ornstein, M.D. (25 minutes)

Mary C. Ruhe, R.N. (10 minutes)

*Commenter:* Kurt C. Stange, M.D., Ph.D. (25 minutes)

## 9:30–10:30 QII to Increase Timely Delivery of Surfactant to High-Risk Newborns During Hospital L & D

*Chair/Moderator:* David Atkins, M.D., M.P.H.

*Presenters:* Gautham Suresh, M.D., D.M., M.S.,  
and Laura C. Leviton, Ph.D. (20 minutes)

*General Discussion:* To be led by Drs. Leviton and Suresh  
(35 minutes)

## 10:30–10:45 Break

Lower Level Lobby

## 10:45–11:45 5 BREAKOUT SESSIONS Led by a Facilitator with Brief Kickoff Remarks by Two Commenters

Goals: Identify solutions to the challenge of improving the science base for QII evaluation at the clinical microsystems level from the breakout group's perspective

Peer Reviewers (Journal Editors and Study Section members)

New Hampshire Three

*Facilitator* Edward Wagner, M.D., M.P.H.

*Kickoff commenters:* Alan S. Go, M.D. (invited), and  
Sankey Williams, M.D. (invited)

*Recorder:* Emily DeVoto, Ph.D., M.S.P.H.

*Recorder on easel:* C. Tracy Orleans, Ph.D.

## Design Experts

New Hampshire Two

*Facilitator:* Laura C. Leviton, Ph.D.

*Kick-off commenters:* Jeremy Grimshaw, M.B., Ch.B., Ph.D., and  
Benjamin Crabtree, Ph.D.

*Recorder:* Lori Melichar, Ph.D.

## Users of QII Research

New Hampshire One

*Facilitator:* David Atkins, M.D., M.P.H.

*Kick-off commenters:* Marian F. Earls, M.D., M.T.S., and  
Paul Wallace, M.D.

*Recorder:* Shawna Mercer, Ph.D., M.Sc.

## Research Training Directors

Potomac Room

*Facilitator:* Joseph Francis Jr., M.D., M.P.H.

*Kick-off commenters:* Richard Hirth, Ph.D., and

Joan Reede, M.D., M.P.H., M.S.

*Recorder:* Neil Thakur, Ph.D.

Disparity Reductions within Quality

Dupont Room

*Facilitator:* Marshall Chin, M.D., M.P.H.

*Kick-off commenter:* Robin Weinick, Ph.D.

*Recorder:* Denise Dougherty, Ph.D.

11:45–12:00 noon Break

Lower Level Lobby

12:00–12:45 p.m. Lunch Buffet/Networking Break/Facilitators and Recorders

Prepare Summary of Breakout Groups' Solutions

New Hampshire Ballroom

12:45–1:30 Plenary: Reports Back by Breakout Groups' Facilitators

(10 minutes each) and General Discussion (15 minutes)

New Hampshire Ballroom

*Moderator:* Arnold Milstein, M.D., M.P.H.

1:30–2:30 SESSION II. RESEARCH EVALUATION DESIGNS FOR QII

AT THE HEALTH "PLAN" LEVEL

New Hampshire Ballroom

Case: Expanding and Testing VA Collaborative Care Models for  
Depression" (ReTIDES QII)

*Chair/Moderator:* Joseph Francis Jr., M.D., M.P.H.

*Presenter:* Lisa V. Rubenstein, M.D., M.S.P.H.

(20 minutes)

*Commenter:* Junius Gonzales, M.D. (10 minutes)

*General Discussion:* (30 minutes)

2:30–2:45 Break

Lower Level Lobby

2:45–3:45 SESSION III. QII AT MULTIPLE CLINICAL SYSTEMS LEVEL:

ICICE (Improving Chronic Illness Care Evaluation)

New Hampshire Ballroom

*Chair/Moderator:* C. Tracy Orleans, Ph.D.

*Presenters:* Edward Wagner, M.D., M.P.H., and

Emmett Keeler, Ph.D. (30 minutes)

*Commenter:* Marshall Chin, M.D., M.P.H. (10 minutes)

*General Discussion:* (20 minutes)

3:45–4:00 Break

Lower Level Lobby

4:00–5:15 5 BREAKOUT SESSIONS led by facilitators (see above)

Goals: Identify solutions for QII evaluation at the health and

multiple clinical systems levels from the breakout group's perspective

Peer Reviewers (Journal Editors and Study Section members)

New Hampshire Three

*Facilitator:* Molla Sloane Donaldson, Dr.P.H., M.S.

*Kick-off commenters:* Martin Eccles, M.D., M.B., and Gayle Weaver, Ph.D.

*Recorder:* Lawrence W. Green, Dr.P.H.

Design Experts

New Hampshire Two

*Facilitator:* Francis D. Chesley Jr., M.D.

*Kick-off commenters:* Kelly J. Devers, Ph.D., and Kaveh G. Shojania, M.D.

*Recorder:* David M. Introcaso, Ph.D.

Users of QII Research

New Hampshire One

*Facilitator:* Peter Briss, M.D., M.P.H.

*Kick-off commenters:* Anne-Marie J. Audet, M.D., M.Sc., and Evelyn Whitlock, M.D., M.P.H.

*Recorder:* Shawna Mercer, Ph.D., M.Sc.

Research Training Directors

Potomac Room

*Facilitator:* Thomas J. Chapel, M.A., M.B.A.

*Kick-off commenters:* Ann Barry Flood, Ph.D., and Mark Splaine, M.D., M.S. (invited)

*Recorder:* Joseph Francis Jr., M.D., M.P.H.

Disparity Reductions within Quality

Dupont Room

*Facilitator:* Denise Dougherty, Ph.D.

*Kick-off commenters:* Constance Fung, M.D., M.S.H.S., and Marsha Gold, Sc.D.

*Recorder:* Debra Joy Perez, Ph.D., M.A., M.P.A.

5:15–6:00 Breakout Group Facilitators and Recorders Convene;

Others Adjourn

New Hampshire Three

5:15 –?::?? Dinner on Your Own (Except Core Planning Group Members)

6:00–7:00 Core Planning Group Members Convene to Debrief

Potomac Room

**September 15, 2006 (8:15 a.m.–1:00 p.m.)**

7:45 a.m. Registration and Continental Breakfast  
Lower Level Lobby

**8:15–9:15 SESSION V. QII AT THE STATE/REGIONAL LEVEL:**

California State Tobacco Prevention and Control, Public  
Health QII—Evaluation Design Issues and Lessons for Health  
Care QII in the Policy Environment

New Hampshire Ballroom

*Chair/Moderator:* Peter Briss, M.D., M.P.H., and  
Shawna Mercer, Ph.D., M.Sc.

*Presenters:* David Hopkins, M.D., M.P.H. and  
Terry Pechacek, Ph.D., M.A. (25 minutes)

*Commenter:* Edward Wagner, M.D., M.P.H. (10 minutes)

*General Discussion* (25 minutes)

9:15–10:00 Plenary: Review Recommendations from September 14  
Breakout Sessions and Discuss

New Hampshire Ballroom

*Moderator:* Tanja Popovic, M.D., Ph.D.

10:00–10:15 Break

Lower Level Lobby

**10:15–11:15 SESSION VI. DEVELOPING INTERSECTING  
RECOMMENDATIONS BREAKOUT GROUPS**

Charge: Expanding the range of acceptable evaluation research  
designs for QII will take action by the many interdependent  
components of the QII evaluation universe over the short- and longterms.  
How would you take the solutions identified to date by the  
homogeneous breakout groups and combine them into solutions that  
would be actionable by the key intersecting stakeholders, acting in a  
complementary fashion? Are there some solutions that could be  
implemented in the short-term (e.g., over the next 1-3 years)? What  
are some solutions that might take a longer time to implement?

Heterogeneous I Last names A-D

New Hampshire One

*Facilitators:* Lisa A. Simpson, M.B., B.Ch., M.P.H.,  
and Denise Dougherty, Ph.D.

*Recorder:* Emily DeVoto, Ph.D., M.S.P.H.

Heterogeneous II Last names E-I

New Hampshire Two

*Facilitators:* Susanne Salem-Schatz, Sc.D., and  
Shawna Mercer, Ph.D., M.Sc.

*Recorder:* David Haggstrom, M.D., M.C.R.

Heterogeneous III Last names J-M  
New Hampshire One  
*Facilitator:* Lawrence J. Fine, M.D., Dr.P.H.  
*Recorder:* Barbara Wells, Ph.D.

Heterogeneous IV Last names N-R  
Dupont Room  
*Facilitators:* Lawrence W. Green, Dr.P.H., and  
Patricia Dolan Mullen, Dr.P.H., M.P.H.  
*Recorder:* David Atkins, M.D., M.P.H.

Heterogeneous V Last names S-Z  
Potomac Room  
*Facilitator:* Joseph Francis Jr., M.D., M.P.H.  
*Recorder:* Neil Thakur, Ph.D.

11:15–11:30 BREAK; GET BOX LUNCH; FACILITATORS AND  
RECORDERS CONVENE  
Lower Level Lobby

11:30–12:30 p.m. WORKING LUNCH: PLENARY AND DISCUSSION:  
REPORTS BACK AND DISCUSSION  
New Hampshire Ballroom  
*Facilitator:* Thomas J. Chapel, M.A., M.B.A.

12:30–1:00 Next Steps, Parting Remarks (Core Planning Group)  
New Hampshire Ballroom

1:00 Adjourn  
New Hampshire Ballroom

2:00–4:00 Core Planning Group Convenes to Debrief  
Potomac Room

## ***Acknowledgments***

The core planning group would like to thank the following individuals for their help in facilitating and recording the breakout sessions:

*Facilitators:* David Atkins, Peter Briss, Thomas Chapel, Francis Chesley, Marshall Chin, Molla Donaldson, Denise Dougherty, Lawrence Fine, Joseph Francis, Lawrence Green, Laura Leviton, Shawna Mercer, Patricia Dolan Mullen, Susanne Salem-Schatz, Lisa Simpson, & Edward Wagner.

*Recorders:* David Atkins, Emily DeVoto, Denise Dougherty, Joseph Francis, Lawrence Green, David Haggstrom, David Introcaso, Lori Melichar, Shawna Mercer, Debra Perez, Neil Thakur, & Barbara Wells.

---

## References

- <sup>1</sup> Institute of Medicine. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: National Academy Press; 2003.
- <sup>2</sup> Institute of Medicine. *To Err is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 1999.
- <sup>3</sup> Institute of Medicine. *Priority Areas for National Action: Transforming Health Care Quality*. Washington, DC: National Academy Press; 2003.
- <sup>4</sup> Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press; 2001.
- <sup>5</sup> Institute of Medicine Board on Health Care Services. *The First Annual Crossing the Quality Chasm Summit: A Focus on Communities*. Washington, DC: National Academies Press; 2004.
- <sup>6</sup> Agency for Healthcare Research and Quality. National Healthcare Disparities Report. [http://www.qualitytools.ahrq.gov/disparitiesreport/download\\_report.aspx](http://www.qualitytools.ahrq.gov/disparitiesreport/download_report.aspx). Accessed June 10, 2004.
- <sup>7</sup> Agency for Healthcare Research and Quality. National Healthcare Quality Report. [http://www.qualitytools.ahrq.gov/qualityreport/download\\_report.aspx](http://www.qualitytools.ahrq.gov/qualityreport/download_report.aspx). Accessed June 10, 2004.
- <sup>8</sup> Agency for Healthcare Research and Quality. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies--Volume I: Series Overview and Methodology. Rockville, MD: AHRQ; July 2004. AHRQ Pub No. 04-0051-1.
- <sup>9</sup> Davidoff F, Batalden P. Toward stronger evidence on quality improvement. Draft publication guidelines: the beginning of a consensus project. *Qual Saf Health Care*. 2005;14:319-325.
- <sup>10</sup> Mohr JJ, Batalden P, Barach P. Integrating patient safety into the clinical microsystem. *Qual Saf Health Care*. 2004;13(Suppl II):ii34-ii38.
- <sup>11</sup> Cretin S, Shortell S, Keeler E. An evaluation of collaborative interventions to improve chronic illness care: Framework and study design. *Evaluation Review*. Feb 2004;28(1):28-51.
- <sup>12</sup> Balas AE, Boren S. Managing Clinical Knowledge for Health Care Improvement. *Yearbook of Medical Informatics 2000*. Stuttgart, Germany: Schattauer; 2000.
- <sup>13</sup> Wilson EO. *Consilience: the unity of knowledge*. New York: Kopf; 1998:59.
- <sup>14</sup> Goodwin MA, Zyzanski SJ, Zronek S, et al. A clinical trial of tailored office systems for preventive service delivery: The Study to Enhance Prevention by Understanding Practice (STEP-UP). *Am J Prev Med*. 2001;21(1):20-28.
- <sup>15</sup> Stange KC, Goodwin MA, Zyzanski SJ, Dietrich AJ. Sustainability of a practice-individualized preventive service delivery intervention. *Am J Prev Med*. 2003;25:296-300.
- <sup>16</sup> Ruhe MC, Weyer SM, Zronek S, Wilkinson A, Wilkinson PS, Stange KC. Facilitating practice change: Lessons from the STEP-UP clinical trial. *Prev Med*. 2005;40(6):729-734.
- <sup>17</sup> Wilbur, K. *Sex, Ecology, Spirituality*. 1995/2000, Boston: Shambhala Publications, Inc.
- <sup>18</sup> Wilbur, K. *A Brief History of Everything*. 1996, Boston: Shambhala Publications, Inc.
- <sup>19</sup> Stange KC, Miller WL, McWhinney IR. Developing the knowledge base of family practice. *Fam Med*. 2001;33(4):286-297.

- 
- <sup>20</sup> Thomas P, Griffiths F, Kai J, O'Dwyer A. Networks for research in primary health care. *BMJ*. 2001;322:588-590.
- <sup>21</sup> Thomas P. *Integrative Primary Health Care: Leading, Managing, Facilitation*. London: Radcliff Press, 2006.
- <sup>22</sup> Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q*. 2004;82:581-629.
- <sup>23</sup> Horbar, JA, Carpenter, JH, Buzas, J, Soll, RF, Suresh, G, Bracken, MB, Leviton, LC, Plsek, PE, and Sinclair, JC. 2004. Collaborative quality improvement to promote evidence based surfactant for preterm infants: a cluster randomized trial. *British Medical Journal*. 2004;329:1004-1010.
- <sup>24</sup> Wilson T, Berwick DM, Cleary PD. What do collaborative improvement projects do? Experience from seven countries. *Jt Comm J Qual Safety*. 2003;29:85-93.
- <sup>25</sup> Ovretveit J, Bate P, Cleary P, et al. Quality Collaboratives: Lessons From Research. *Quality And Safety In Health Care*. 2002;11:345-351.
- <sup>26</sup> Shortell SM, Marsteller JA, Lin M, Pearson ML, Wu SY, Mendel P, Cretin S, Rosen M. The role of perceived team effectiveness in improving chronic illness care. *Med Care*. 2004;42:1040-8.
- <sup>27</sup> Centers for Disease Control and Prevention. *Best Practices for Comprehensive Tobacco Control Programs—August 1999*. Atlanta GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, August 1999. Reprinted, with corrections.
- <sup>28</sup> Novotny TE, Seigel MB. California's Tobacco Control Saga. *Health Affairs*. 1996;15:58-72.
- <sup>29</sup> Kupersmith J, Sung N, Genel M, et al. Creating a new structure for research on health care effectiveness. *Journal of Investigative Medicine*. 2005;53:67-72.