

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.^{1,2,3,4,5,6,7} 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.*”⁸ 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.⁹

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."¹⁰ 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.¹¹ 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 Balas¹², 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. Wilson¹³ 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:

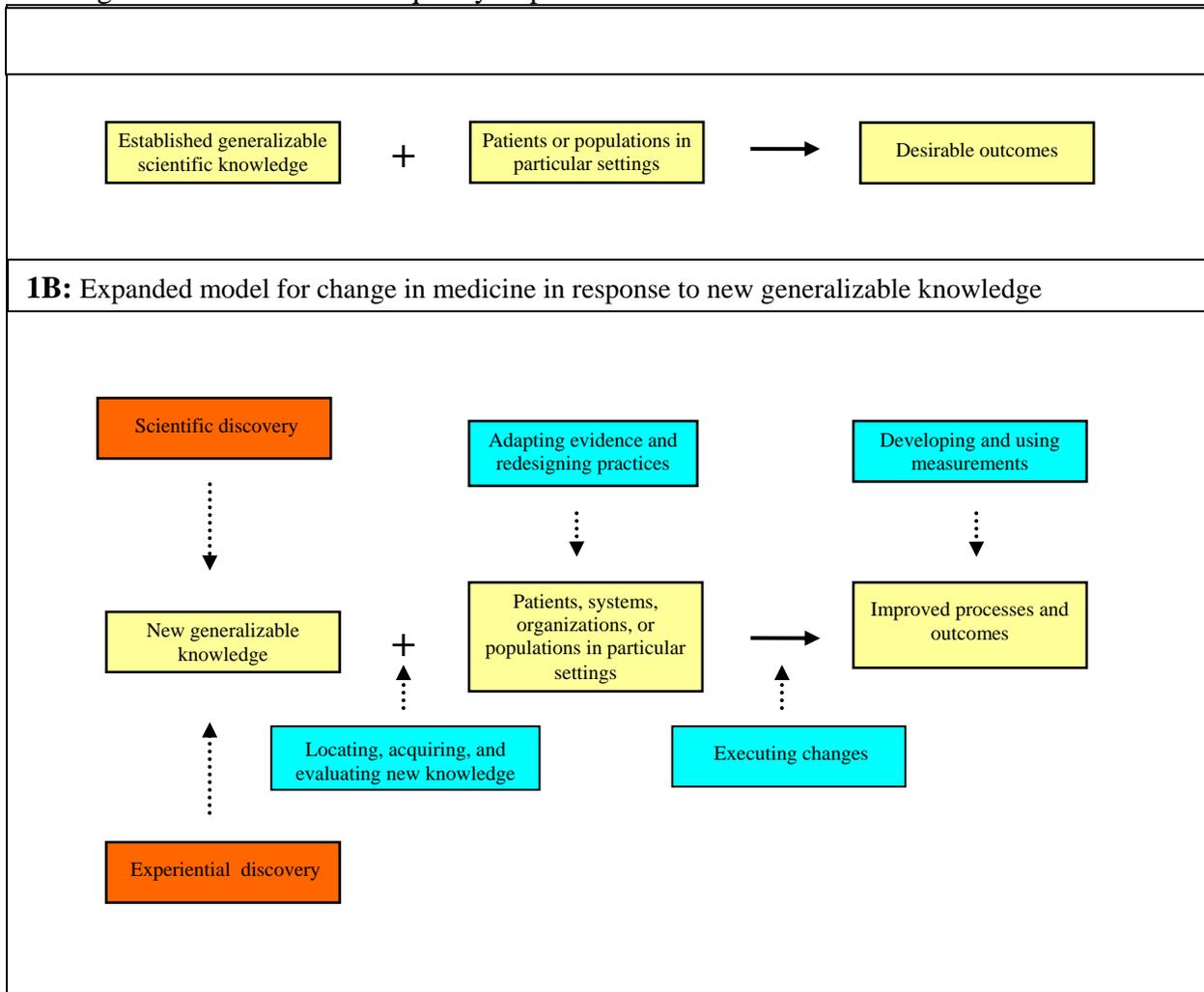
- Makes establishing repeatability difficult.
- Prevents public scrutiny and accountability.
- Reduces the incentives and opportunities to clarify thinking, verify observations, and justify inferences.
- Slows the spread of known improvements.
- Inhibits the discovery of innovations.
- Ethical issue: 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)^{14,15,16} 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.

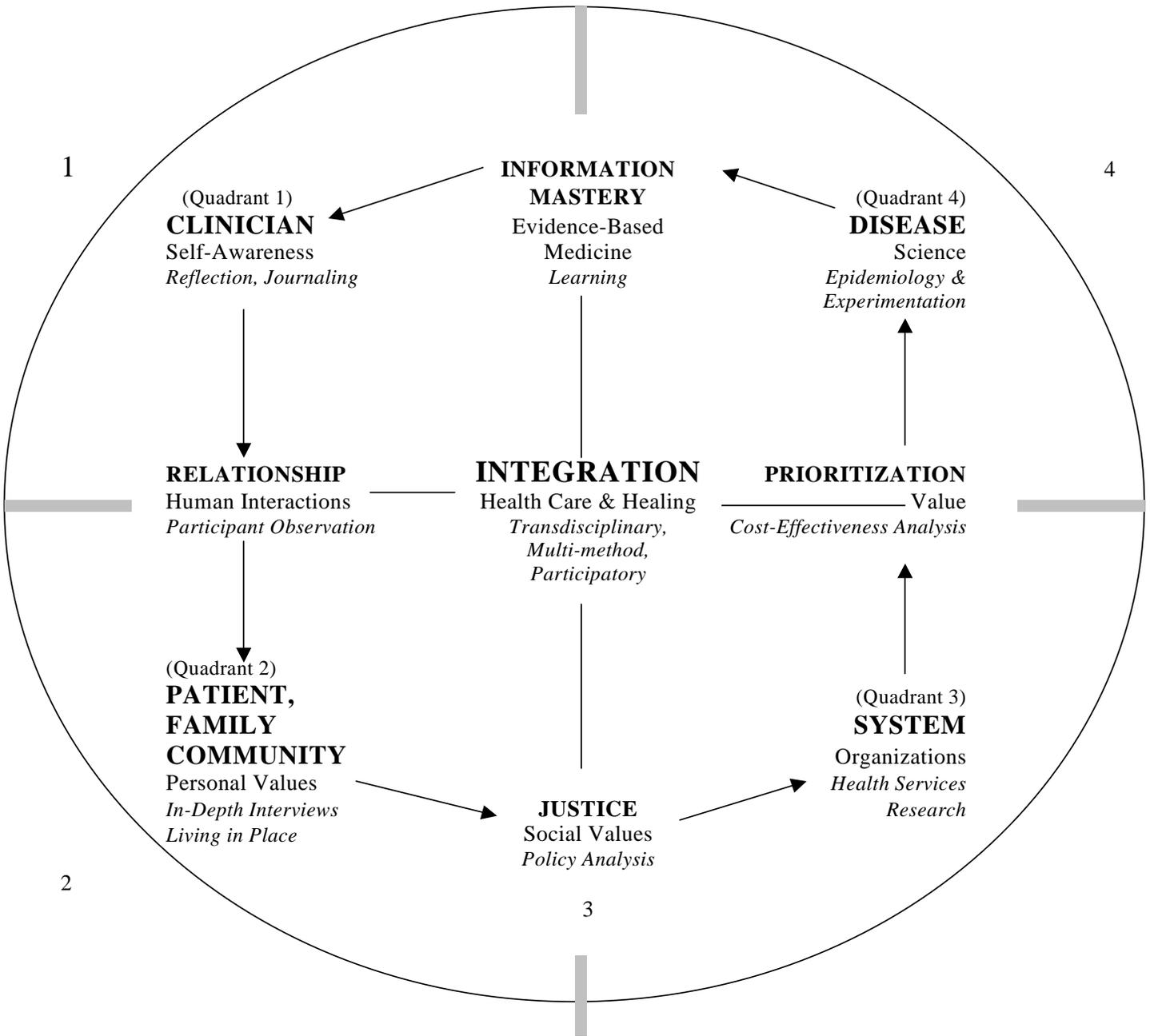
Dr. Stange's understanding of the need for the conference is that we feel some angst because our theory and methods and worldview do not match the problems we are addressing. The fundamental problem is that our ways of thinking, our ways of knowing, and our methods are good at isolating a phenomenon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

	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.¹⁹



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^{20,21} 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.

Greenhalgh²² 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.²³ 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

modest impacts. Even with support for using evidence-based tools (currently available on at least four websites originating from randomized studies), only some practices (those with features fostering success, such as support from mental health and primary care leadership) succeed in adopting collaborative care. Additional work adapting collaborative care, as carried out in research studies, to the needs of routine practice in particular health care systems and settings appears to be necessary for successful dissemination of collaborative care, and for assuring that the disseminated model is implemented in a way that achieves impact on practice performance.

The Translating Initiatives in Depression into Effective Solutions (TIDES) project was a nonrandomized study in which performance reports helped guide PDSA (plan, do, study, act) cycles. The clinical results of the study have been very good. However, the TIDES study found that a “top down” approach to change is only effective in the VA when the “bottom up” side has already been built. The study found there was no handoff of VA-adapted collaborative care, designed through the TIDES QI process by three VA regional networks (Veterans Integrated Service Networks or VISNs), from research to a clinical entity. They concluded that they would have to go further than they had thought to make that handoff happen by interacting with many different government entities with which the VISNs are connected.

The goal of ReTIDES is to expand TIDES to a new VISN and more medical centers, providing the basis for national implementation including a fully developed business case. Regional expansion of TIDES is being evaluated with semistructured stakeholder interviews with an emphasis on performance-measure-based evaluation. This study is being done with a non-equivalent control group design with pre- and post-test measures at multiple time intervals.

Threats to the current design of ReTIDES include not being able to achieve the scope of an intervention that would affect the performance measures as well as all the usual threats to nonrandomized designs. For the study’s purpose, a nonrandomized design was better than randomization. The previous randomized trials present a strong evidence base, and there would be a low level of gain from one more randomized trial compared to the gain from learning about and fostering system implementation. Randomization would further restrict their ability to study the naturalistic process of implementation. For example, there are only a limited number of comparable primary care practices in any regional network, accounting for size, rural versus urban location, academic affiliation, and existing depression care organization. Randomizing within strata of comparable practices would significantly limit management choice regarding how to roll out improved depression care at a regional level. Furthermore, randomization reinforces the idea that the implementation is research, not quality improvement, allowing participants to think of the implementation as temporary and not to think about all of the routine administrative policies, procedures, and resource allocation decisions needed to sustain collaborative care on an ongoing basis.

The ReTIDES approach has been adopted into the national VA primary care strategic plan and the hope is that this approach will be adopted into routine care throughout the VA over the next three to five years.

Improving Quality: A Steeplechase of Sorts? The Necessity of Science for Practice and Science for the Public

Junius Gonzales, M.D.

Acting Director, Division of Services and Intervention Research, National Institute of Mental Health

According to a December 6, 2003, *BMJ* report about the disconnect between research and practice and between policy and practice, there is a need to reconnect funders, users, researchers, and those who might benefit from research. Users and researchers agreed that no one was taking responsibility for disseminating findings and supporting application of findings. Dr. Gonzales commented on randomization versus ecologically relevant results. Dr. Rubenstein matches pairs of sites based on organizational and patient data, and Dr. Gonzales raised the issue of whether there might be contextual, dynamic, or process issues or events that would mean that sites actually were not a great match.

A rigorous examination of process data, as well as strategies for theorizing based on process data, would help make the business case as well as inform the designs of future QI studies. It is important to address balancing accuracy, generality, and simplicity. Clinicians want to know how the results from a particular RCT apply to the patient sitting in front of them who, for example, is outside the age group examined in the study and has co-existing conditions.

A number of challenges and opportunities exist for any QI study. Basic questions revolve around generalizability to other settings and populations. How collaborative care works has not been deconstructed; thus, we do not know about the potency of interactions between different elements of collaborative care. Also, no one knows what mechanisms, such as continuity of care or patient trust, make the model work. Business managers want something better, faster, and cheaper, thus they are focused on wanting to know what the “active ingredient” is that makes collaborative care work.

What we need in the area of QI research is to identify and understand modifiable mechanisms of change, and not just to describe patterns of change. We need new ways to address implementation science, such as taxonomies of near/far, serial, and expert transfer, that increase the potential for local adaptation rather than being limited to a “one size fits all” approach. On the funding side, it would be valuable to invest in the research infrastructure for “real world” settings. Sometimes the needs are very basic, such as computers for data collection. The National Institute of Mental Health (NIMH) has some funding mechanisms outside the usual grant schedule which allow researchers to move quickly to seize research opportunities on “natural experiments” as they happen.

Finally, Dr. Gonzales noted that the big issue for the improvement in care seen in Dr. Rubenstein’s research is sustainability. In other words, will the changes stick?

Discussion

A participant asked about the issues of contamination and floor and ceiling effects. In addition, he noted that when control groups show improvement, some people describe this as a

contamination effect. He considers that pejorative and makes it likely that one will miss the opportunity to find out *why* the controls also improved. Perhaps one reason it was difficult to see improvement is that most people were doing well already, thus there would be a ceiling effect on improvement. To avoid that, one would have to look at the degree of improvement as a function of the baseline rate. Dr. Rubenstein responded that this is related to the issue of taking advantage of one's results, and is relevant to randomized trial findings as well. What one person would consider contamination, another would consider spread. At the patient level, in Dr. Rubenstein's studies, there is control for baseline sickness.

A participant asked why "deconstruction" of collaborative care is needed when it seems clear that the deconstructed components are not working. Dr. Gonzales replied that not all aspects of collaborative care were proven effective or ineffective. The package is transferable, but we do not always know why it works for certain groups but not others. Also, there are different models of collaborative care. The reason that elements of the collaborative care package need to be deconstructed is that practices lack funding for the whole package in a collaborative care model. Dr. Gonzales feels that the active ingredients still are not known. Dr. Rubenstein added that they are deconstructing the model in terms of taking apart its operational components in order to make the business case. For example, patients must be assessed, thus the research group is examining what happens when a patient is assessed. Her group also is attempting to test innovations in this model performed at individual sites, while trying to ensure that the model is not watered down. They want to keep the core model very tight while still being able to examine innovations.

Another participant commented that the placebo effect, as well as the Hawthorne effect, are poorly understood in cluster randomized trials. Blinding is impossible in many of these trials. Innovative designs, such as randomization in which the control group is unaware they are in a trial, or Zelen randomization, should be considered.

A participant asked about the business case for QI initiatives. What is needed to examine the business case is not a shift in the study design, but an interest in a range of variables, such as economic variables, that are not typically included in cluster randomized trials. Dr. Rubenstein responded that one can get much of the information needed for the business case from classical study designs in terms of cost effectiveness and how patients flow through the system. However, one has to look past randomized trials to truly capture the system costs, which is what the business managers are interested in. One has to let the system operate freely in order to capture things such as how cost shifts from primary care to mental health care. Dr. Rubenstein noted that it is not easy for researchers and business managers to interact given their different vocabularies. A participant noted the problem with the business case is that it depends on context (e.g., patient location, revenue, etc.) and that the level of care will vary with these circumstances.

Another participant commented that when considering the business case, one must examine who is paying. For purchasers, outcome measures should be included that focus on staff "presenteeism" and absenteeism. In areas such as depression care, there is evidence of a return of dollars from improved labor outcomes.

SESSION III. QII AT THE MULTIPLE CLINICAL SYSTEMS LEVEL: IMPROVING CHRONIC ILLNESS CARE EVALUATION

Tracy Orleans, Ph.D. (Chair/Moderator)

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

The Improving Chronic Illness Care Evaluation

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

Director, Center for Health Studies, MacColl Institute for Healthcare Innovation

Dr. Wagner noted that when RWJF funded the Improving Chronic Illness Care (ICIC) program, there had been advances in chronic illness care that the IOM's *Crossing the Quality Chasm* report showed were not being implemented in practice. In addition, a reasonable body of evidence showed what worked for translating research findings into practice, but these findings also were not being used to change care for most patients. In addition, there was evidence that the continuous quality improvement (CQI) methods implemented in the 1980s and 1990s and traditional continuing medical education were not effective. The Chronic Care Model (CCM) came out of the ICIC team's reading of the literature and their attempts to implement the findings in the Group Health system. His group next wanted to examine whether busy practices could change their practice systems in accord with a multi-component change model, such as the CCM, since multi-component approaches are more likely to be successful than a single component approach. The questions they wanted to address were: can busy practices change their practice systems in accord with a multi-component change model, and, if so, what impact will it have on the quality of care and the outcomes of their patients?

The most successful model for implementing the CCM appeared to be the Breakthrough Series. At the heart of these collaboratives is an approach to QI developed by Lloyd Provost and his colleagues called the "model for improvement," which has three basic elements:

- Set a clear aim.
- Have a measurement system in place that charts whether or not progress is being made.
- Implement a set of grounded changes and test them using rapid cycle PDSA methods to determine whether or not the changes accomplish what they were hypothesized to do.

The Breakthrough Series is a yearlong process that brings together teams from organizations wanting to make change with faculty in a series of in-person meetings, or learning sessions, with much electronic communication in between the meetings. During learning sessions, teams plan sets of changes to be tested in the action periods that follow. One of the questions they hoped to tackle with the Chronic Care Model was whether or not it is truly a generic model, meaning that if it works for diabetes, it should also work for other conditions. Dr. Wagner said over the course of several years, Improving Chronic Illness Care conducted collaboratives in a variety of chronic conditions including diabetes, depression, asthma, and congestive heart failure. In the earliest collaboratives, participating organizations reported mild to moderate improvements in the quality of care. However, they wanted to conduct a more rigorous research-oriented evaluation. Through a competitive process, RANDHealth was selected to evaluate early

collaboratives. The program, the Improving Chronic Illness Care Evaluation (ICICE), was next described by Dr. Keeler.

Design Decisions and Lessons Learned in the Improving Chronic Illness Care Evaluation (ICICE)

Emmett Keeler, Ph.D.

Senior Mathematician, RAND Corporation

The Improving Chronic Illness Care Evaluation looked to answer these main questions:

- Did participating in the collaboratives induce positive change?
- Did implementing the CCM improve processes of care?
- Did implementing the CCM improve patient health?
- What did participation and implementation cost?
- What factors were associated with success?

On average, organizations in the study made more than 30 systemic changes over the year. These changes enabled the researchers to determine if moving toward the CCM is good for patients. The results showed that the process of care, patient self-management, and some outcomes improved more for intervention patients than control site patients. The areas that were emphasized in the learning sessions were the ones that changed the most. The website <http://www.rand.org/health/ICICE> includes findings from multiple papers and other information about the study.

Dr. Keeler discussed design considerations. In any intervention, it is important for the “signal” of the treatment effect to stand out by reducing the statistical “noise.” One way to achieve a high signal to noise ratio is to have a large number of subjects, which will eliminate or decrease problems caused by random error. Another approach is to have tight, homogenous criteria for enrollment and tight protocols for treatment. Tight designs reduce noise, but limit the generalizability of the results. The RCT paradigm eliminates biases further through randomization and blinding.

Is randomization possible in all studies? Organizations want success (meaning that their organization improves), not necessarily science. Moreover, the organizations with which the ICICE team worked stated that their patients did not like to “feel like guinea pigs.” His group was not able to convince participating organizations to accept randomization of sites for the trials, and patient-level randomization was not possible because this project addresses systemic changes. Components of strong study designs that are alternatives to the RCT include:

- Before and after with a matched control group.
- Multiple sources of data and an evaluation logic model.
- Planning for and testing potential biases.

The first study design decision concerns the constitution of the control group. Dr. Keeler suggested using another section of the organization, such as a clinic or a group of doctors, to serve as an internal control group. A before and after with a matched control group design can

be used to control for secular trends and unrelated changes in the organization. Also, patients with chronic diseases that are fairly stable, such as diabetes, can be used as their own controls. One of the benefits of an external control site is the lack of contamination from intervention, although these sites are less likely to cooperate and are more expensive to include. The ICICE used internal control sites. To identify patients, Dr. Keeler's group used sampling frames from site registries of patients with the diseases of interest. There were some inaccuracies in the registry, and patients who said they did not have the disease in question or did not obtain care at the sites were excluded.

Five different sources of information were used:

- Patient telephone surveys—The patient is the best source of information on what the care provider does in terms of patient education and effective communication and whether information was retained. With chronic diseases, the patient is the one who has to do the work of self-care. The phone calls were used to cross-check if improvements noted in medical charts also were reported by the patient. The telephone survey could be used to determine if improvements in charts are documented or real. The cost of the survey was \$100 per call, with approximately 4,000 patients called. Phone surveys were done at the end of the collaborative.
- Medical record abstraction—Information from charts can provide true before and after intervention data.
- Clinical and administrative staff surveys.
- Monthly progress reports—The collaborative asked the team at each intervention site to complete a brief report each month for the leaders and for the Institute for Healthcare Improvement (IHI). The surveys were helpful in determining what the sites did, but the lack of standardization across sites reduced the value of their statistics.
- Final calls with leaders—The team conducted follow-up phone calls with leaders one year later to determine 1) major successes, 2) major barriers in implementing the CCM efforts and how they were addressed, and 3) continuation and spread of CCM efforts.

The team examined how “close” control sites were to intervention sites in terms of geography and overlap of staff. “Closeness” was a mild predictor of control sites showing improvement. The team also asked during exit interviews whether control sites were engaging in other QI activities.

Several other important points were made regarding contamination, bias from the use of volunteers, concerns about pressure from funders, and potential ways to decrease evaluation costs, such as 1) to lower multi-site IRB and consent costs perhaps via prior patient consent for quality improvement and QI research activities, and 2) to reduce data collection costs via the use of electronic medical records or clever use of existing data, like claims.

Comments on QIIs at the Multiple Clinical Systems Level

Marshall Chin, M.D., M.P.H.

Associate Professor of Medicine, University of Chicago

Dr. Chin concurred with most of Dr. Keeler's points and used his time to build upon them. He noted that people sometimes propose a time series design as an alternative to a before-after design, but often it is difficult to gather the necessary data points. In addition, the patient-level cohort design (tracking the same patients longitudinally) can be difficult to implement because of concerns related to the Health Insurance Portability and Accountability Act (HIPAA) of 1996. When institutional review boards require informed consent from individual patients for such studies, the resulting sample of patients may be biased. He compared studies using the "population of focus" (meaning a subset of the organization) mentioned by Dr. Keeler with those using a random sample of the whole clinic population. The results are frequently different depending on which method is used, although the direction of bias can vary. For short-term evaluations, it is reasonable to use the population of focus model because in a short period of time, one does not expect the intervention to spread to the whole clinic. However, over the long term, the more important goal is improving the overall quality of care of all patients; thus, the analysis of the whole clinic population is more appropriate in that case.

Dr. Chin endorsed Dr. Keeler's approach of using multiple sources of data. However, in Dr. Chin's studies of health centers, monthly reports have been of limited use because participants do not always supply detailed data.

Dr. Chin reviewed the key research questions and methods challenges in health care QI. He noted that there are some nice reviews that profile critical components of QI collaboratives and the traits of successful collaborative teams. Dr. Chin recommended articles by Wilson²⁴ and Ovretveit²⁵.

Some of the challenges in QI research include isolating the effects of individual components within multifactorial interventions, since policymakers and managers want to know where to invest limited resources. Another challenge is how to go beyond general statements, such as the importance of "leadership support and buy-in," via more detailed analyses and new measurement tools. Third, there is a need to be able to tailor interventions to an organization's stage of change. Fourth, one must strike a balance between structure and autonomy. The Breakthrough Series is a general QI process rather than a set of prescribed interventions. Centers also have found it useful to have a menu of intervention choices that they can adapt to their needs. Fifth, an important area for future research is to identify what incentives and assistance can enhance and sustain quality improvement efforts further.

It is very hard to get the appropriate data and there are few sophisticated models for analyzing the business case for quality improvement, which may be why there is so little on this in the literature. Yet, for QI to be viable, we need to be able to make the business case. The societal perspective is also important and difficult to study. Sustainability is a difficult issue with limited data, partly because funding cycles are short and studies frequently are limited to one to two years. What are the unintended consequences of QI? This topic is rarely examined, but should be.

Dr. Chin concluded by noting that RWJF has a new initiative titled “Finding Answers: Disparities Research for Change” to apply QI and other innovations to disparities issues. The goal is to identify what interventions work to reduce disparities in real world practice.

Discussion

Dr. Wagner was asked to address the value of the external evaluation of his program. Dr. Wagner responded that the RAND evaluation has been very helpful. New papers are becoming available that will aid in dissemination of the CCM. Health care leaders are becoming increasingly sophisticated and understand the limitations of their own QI data.

A participant indicated it will be difficult to “unpack” some multi-factor interventions, especially when certain studies have found that some interventions perform better when they are “packed.” Concepts and methods are available for examining bundles of activities or configuration work in organizational science, which is not often referenced in health services research. We may find that the “key ingredient” only works when it is in interaction with four or five other ingredients. Another point is that there may be multiple routes to the same destination, or equifinality. It would be useful to have studies comparing different approaches to improvement to test the equifinality hypothesis.

In response, Dr. Keeler cited a recently published a meta-analysis of the components of QI efficacy studies. Each of the four parts of the CCM that was testable was effective, but none was essential. Statistical analysis showed that the remaining components are effective even when one component is removed.

In response to a question about why there was so little use of provider-level statistics, Dr. Keeler replied those statistics lacked standardization. His study found that not every practice had an existing registry. It would be possible to put more pressure on participants to standardize in order to get usable data, but Dr. Keeler’s group decided just to use their own measures. Dr. Wagner noted that these were early collaboratives and their ability to guide teams has gotten better since then. IHI has a philosophy of not letting measurement slow down improvement, adding that if an organization has a measure it is accustomed to using, the researchers just use that measure in order to facilitate the improvement process. However, over time, there are more and more standardized measures in use.

One participant noted the years of work on the part of Dr. Wagner and IHI to determine what model was going to be tested to aid in the rapid cycle of change. Dr. Wagner’s ability to publish his early work and refine the model over time was important in warranting a major investment in this research.

A participant asked what will be the QI example paralleling the clinical world’s example of hormone replacement therapy (HRT), noting that HRT had been found to be safe in non-RCTs and its cardio-protective nature had biological plausibility. For HRT, a randomized trial was thought to be nearly unethical since it seemed self-evident that HRT was a good thing. When an RCT was conducted, it was found that HRT put women at increased risk of cardiovascular

disease in the first 12 months of treatment. In other words, what is the QI disaster that is waiting to happen? Dr. Orleans replied that the U.S. Preventive Services Task Force conducted a systematic review of the literature on this topic. Similar standards are needed for accumulating, critiquing, and synthesizing the evidence as it becomes available, and that is a role for journal editors, funders, and everyone involved in this new field. We need to build the ability for systematic review in the field of QI. Others noted that we have a positive publication bias and that many studies are short-term and describe the intervention generally. There needs to be more attention to disseminating the specific details of an intervention.

Dr. Keeler was asked to comment about the variation across sites and analyses of the factors contributing to success. Dr. Keeler replied that there are about 20 sites and hundreds of variables on site characteristics, subjects, etc. Statistically, it is not appropriate to examine so many variables for so few sites. However, it is valuable to take a common sense approach to talking to people about what worked. Dr. Wagner noted that Steve Shortell et al. have an article in *Medical Care*²⁶ on organizations and quality improvement teams that is an important step in the direction of understanding process and mediators. Dr. Orleans noted that one important outcome of Shortell's research was the new taxonomy that was developed.

SESSION IV. 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

Peter Briss, M.D., M.P.H. (Chair/Moderator)

Chief, Community Guide Branch, Centers for Disease Control and Prevention

Shawna Mercer, M.Sc., Ph.D. (Chair/Moderator)

Health Scientist and Senior Advisor, Office of the Chief Science Officer, Centers for Disease Control and Prevention

Dr. Briss introduced this session by mentioning that the California tobacco control model was a complex multi-component strategy with evolving interventions and different approaches to measuring outcomes. It presents an important, real world problem mixing rigor and relevance.

Program Evaluation in Public Health: California's Effort to Reduce Tobacco Use, 1989–2005

David Hopkins, M.D., M.P.H.¹

Staff Scientist, Community Guide Branch, Centers for Disease Control and Prevention

Dr. Hopkins noted that this presentation was on program evaluation in public health, and that at the end of the presentation, he would discuss bridging the gap between how public health thinks about these issues versus how health care thinks about these issues. Dr. Hopkins asked participants to think about how California's experience mirrors participants' experience with QI or differs from it, and about what can we learn from this population-based public health approach.

¹ Dr. Terry Pechacek of CDC contributed to the preparation of this presentation and had planned to present jointly with Dr. Hopkins but was unable to attend the symposium.

Dr. Hopkins began by presenting background information on California at the state's tobacco control program's inception in 1988. With a population of more than 28 million people, California had 4.8 million adult smokers (22.8 percent), which was the second lowest average in the country. Led by a health coalition, voters approved Proposition 99 in November 1988, which increased the excise tax on cigarettes by 25 cents per pack and earmarked funding for a statewide program, which became the largest tobacco prevention and control program in the world.

California then had a great deal of money for programs and an urgent need to decide what to do and how to evaluate its efforts. In 1998, there was limited experience with effective population-based tobacco control interventions. There were some studies at the community level on cardiovascular disease risk reduction, and some experience with multi-component programs. Meanwhile, there was an appreciation of the effects of policy change on behavior as econometric studies on the effects of price on tobacco use were being published.

California had program options and had to decide which interventions to use and how to implement them. California opted not to take a top-down approach in which all programs would be implemented statewide. Instead, they had a couple of statewide program components and left the rest to local decision-makers. They opted to pursue a "full court press" from the outset, instead of developing programs built on the results of smaller-scale demonstration projects. Meanwhile, NCI was advocating a comprehensive approach that was based on local control of the issue and emphasis on, and funding of, community coalitions to examine issues. The goal of NCI's approach was to have multiple channels, such as media campaigns and school-based programs, and multiple targets, such as increased cessation and reduced initiation of tobacco use, for the interventions. California provided a field test for the approach that NCI advocated. In 1990, California adopted tobacco prevention and control initiatives, which included a paid mass media campaign, funding for school-based programs, and funding for intervention and treatment research. Although there were challenges in the evaluations of this program, the state built evaluations into the mandate. Surveillance systems were put into place and more surveys were added, enabling the state to receive answers quickly. Components of the program were evaluated through contracts (independent evaluators) and a research program was funded within the University of California. This had the benefit that both good news and bad news got published. Funding for local intervention and research projects came with the stipulation that 10% of the budget be spent on evaluation, and support for local groups was provided that directed them to experts who could consult on or conduct evaluations.

An oversight committee was appointed to conduct an annual review of the surveillance and research results and to provide advice and recommendations based on the findings.

Outcomes of the California program included a decrease of 32.5% in smoking prevalence among California adults between 1988 and 2004. In addition, consumption decreased 55.6% in California, compared to a 32% decrease in the rest of the U.S., between 1988 and 2003.

The California strategy plays a large role in the tobacco control literature, having produced dozens of publications influencing tobacco prevention and control efforts. The evaluations have documented the overall impact of a comprehensive tobacco control effort as well as the

independent contribution of some components, such as the helpline and smoke-free policies. The evaluations contributed to program survival. At the same time, local program impact is unclear because most evaluations have not been published and comparisons have been difficult. The effectiveness of some interventions, such as school-based programs, remains unclear. There have been some adjustments made to enhance evaluation, such as adopting more uniform surveillance tools.

California's program has become the model for state-level comprehensive tobacco control and the California experience contributed to the contents of the CDC's *Best Practices for Comprehensive Tobacco Control Programs*.²⁷

Discussion

Questions revolved around issues of sustainability and ceiling effects.

Dr. Hopkins noted that once the tobacco control and prevention program was de-funded in Minnesota, there was a relatively quick increase in susceptibility to smoking among teenagers. He noted that many political issues make funding these programs difficult. Several states have de-funded their programs, suggesting that the weight of the evidence was not sufficient to counter the political processes that undermine such programs.

A participant asked if Dr. Hopkins has observed a plateau in the decrease in smoking. If so, are there plans to change the approach to deal with the "bedrock" of smokers who will not quit? A national survey on systems did show a leveling off in the decline of the numbers of smokers.

Comments on the California Tobacco Program

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

Director, Center for Health Studies, MacColl Institute for Healthcare Innovation

Dr. Wagner noted that the California tobacco control and prevention experience has some analogies or lessons for the improvement of medical care. The program, while it used a multilevel approach that goes beyond most published health care QIIs, has many parallels with health care QIIs. These include an attempt to change a system and a culture (tobacco acquisition, initiation, and use). The intervention occurred at the levels of society/policy, organizations, schools, community institutions, and individuals, the last occurring through quitlines and other direct-to-consumer services. Local adaptation was a built-in, essential feature of the program; thus, internal variation was encouraged and program standardization was not encouraged. Evaluations depended on ongoing surveillance data supplemented by intervention-specific measurement and evaluation activities. This is a nice model for complex quality improvement interventions.

The same questions we would ask of a medical care quality improvement intervention were asked in the California effort. General evaluation questions included: Do QIIs work? If so, what is the contribution of individual components? If not, were there promising components whose effect perhaps was overwhelmed by ineffective components?

For this complex intervention, the gold standard of evaluation of a rigorous test of a standardized, simple, unchanging treatment over a relatively short time span would have been totally irrelevant. The ability to randomize was not the issue. Researchers randomized where they could, but some situations, such as the state-level change in excise tax, precluded randomization. The evaluators “played the hand they were dealt” and developed the strongest quasi-experiment possible within the context of this natural experiment. Evidence and theory, such as the COMMIT trial experience and studies of economic effects, suggested that comprehensive system change was indicated, and this approach indeed was taken in California. The standardization of an intervention across sites and over time precludes learning. It was clear before the program began that excise tax increases and quitlines were two effective methods in tobacco control. There was not such a strong evidence base for what local communities could do, so innovation at that level was encouraged.

Comprehensive system change often is necessary. Interventions such as the molar randomized trial of a single pill or a single machine are irrelevant when one needs to change multiple interacting components. In the California tobacco case, the conceptual model was to make tobacco less attractive to buy, more difficult to use, and easier to give up. In the future, health care interventions also will have to be multilevel because practices do not exist in a vacuum. They interact with other practices, insurance providers, and larger organizations and they are affected by policy changes. Future interventions will need to consider these interactions. Sequential testing and factorial designs may not be feasible or possible in all cases to test all of the multiple components; thus, we need to increase our efforts with multi-component interventions and increase our ability to learn from them. Dr. Wagner asked rhetorically whether multi-component, multilevel, adaptive, system change interventions are evaluable. For this question, we can learn valuable lessons from evaluation sciences in other disciplines, such as education, social services, and welfare, and in the use of regression discontinuity designs, for example. At a conference that Dr. Wagner attended many years ago, Mark Lipsey emphasized small“t” theories of treatment, which are not general theories of how the world works. Instead, they are attempts to describe how complex interventions work by positing how the various elements of the intervention work, what the mediators and moderators are, and what outcome variables will be affected in a given subpopulation at a certain time. To develop a small“t” theory, first one would lay all that out and then attempt to develop an appropriate measurement strategy. For example, one might theorize that leadership influences the effectiveness of quality improvement teams in an organization and then the effectiveness of the quality improvement team affects the depth and the relevance of the changes which next lead to improvement in care.

The lessons that Dr. Wagner drew from the California example are:

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

Discussion

Discussion focused on expansion of Dr. Wagner's comments and enhancement strategies for doing quick turnaround research on natural experiments, the interface of clinical care with public health messages, and political pressures.

Dr. Green affirmed a point made by Dr. Wagner, which is that a comprehensive program has merit not just because the various elements support each other synergistically but also because some elements work for some people and other elements work for other people. Dr. Green attended meetings with the Oversight Committee (California's Evaluation Advisory Committee) which was being asked which elements of the program worked and which ones did not. Certain groups were looking to cut funding for parts of the program that could not be defended, but the Committee argued that the program should not be dismantled. Contracts that were given to private companies to evaluate the intermediate effects for specific interventions indicated that some things were not working in some places, but did work in other areas. In a multi-level, multi-component intervention, individuals are surrounded by variations on the program message through different channels and different means of communication; some people are more responsive to certain channels than others, and different people respond to different channels. All of this is what makes comprehensiveness valuable. This is what made it worthwhile that the planners remained steadfast regarding holding the elements of the program together, even when some elements were faltering in some places. Dr. Green encouraged those in medical care settings to consider the principle of avoiding disaggregation of complex programs and to look for ways to evaluate them comprehensively.

Commenting that many policy changes are occurring which are not preplanned and have not incorporated evaluation, suggestions were sought concerning funding mechanisms and strategies for capturing information about the effects of the policy changes. Currently, we are not doing a good job of capturing real world changes. To what extent does one benefit from having baseline data? What does one do when one does not have baseline data? Typically, a policy change will have already occurred by the time one secures the funding needed to study the change. Dr. Briss responded that it is a conventional practice in public health to measure important outcomes over time using surveillance, thereby providing for baseline measurement. When changes occur, scientists can then take advantage of the natural experiment. This approach could be taken in health care systems as well. Dr. Wagner noted that a survey of communities interested in RHIOs (regional health information organizations) was conducted. Participants were asked what function they saw RHIOs serving. Only half reported that they were going to use them for community-based performance measures, thereby missing an opportunity to get community-based information exchanges to produce community-wide performance measures. Resulting data could provide a platform for evaluating policy change.

Pointing out that there are many public health messages in California, another participant asked what happens when teenagers visit their primary care providers and are not asked if they smoke. She is involved in an intervention program with primary care providers and feels this program interacts nicely with the different levels of influence adolescents receive from public health messages.

A participant noted that both the intervention and the target were evolving over time in California. The panel was asked if they had any way of examining industry pushback in response to their evaluations. Dr. Hopkins referred participants interested in this topic to T.E. Novotny and M.B. Seigel's article titled "California's Tobacco Control Saga" published in *Health Affairs*.²⁸ Dr. Hopkins described a period in the 1990s when tobacco industries contributed more to California political campaigns than they did to Congressional campaigns, made an effort to have the word "addiction" removed from program materials, and made other efforts to interfere with the program. Dr. Green noted that the industry pushback was felt on the evaluation side. The tobacco control effort in Massachusetts was able to build on California's effort and began with a higher tax and a faster take-off for the program. The CDC extracted from these efforts a set of "best practices." There were concerns about using that term for findings that were not based on randomized studies; however, other states gave greater weight to these two states' experiences than they did to decades of findings from randomized controlled trials. This may be an important perspective for quality improvement efforts, in that other practitioners may pay more attention to findings from settings similar to their own than to other data.

Summary of the Recommendations

OVERVIEW OF THE REPORTS BACK SESSIONS

The September 13-15, 2005 symposium was designed to elicit a collaborative exchange of ideas to improve the science base for health care and public health quality improvement research and evaluation. During the symposium, three breakout sessions were held to give participants the opportunity to generate recommendations. For the first and second breakout sessions, participants were asked to attend the group related to their area of expertise. The five homogeneous groups were: research design experts, peer reviewers and journal editors, research training directors, experts in disparities in health care quality, and users of quality improvement research, meaning stakeholders and evidence synthesizers. For the third breakout session, the purpose of the session was to generate overarching recommendations and, for this, participants were sorted into heterogeneous groups by the first letter of their last name. Each breakout session was led by a facilitator who was assisted by a recorder taking notes at a flipchart. Facilitators and recorders are listed at the end of this document. The facilitator and recorder conferred to produce a five-minute oral report for their session. The reports for the homogeneous breakout sessions are combined for the summaries below.

The charge to the breakout groups for the first breakout session was to recommend solutions to the challenges described in the presentations at the clinical microsystems level and, for the second breakout session, to recommend solutions to the challenges at the clinical microsystems and health systems levels. In addition, the second session's purpose was to note similarities in the first and second sessions' recommendations, and to discuss the roles for various players (journal editors, funders, researchers, end-users, and others).

In introducing this charge to the breakout sessions, Dr. Dougherty noted that this symposium might transform into a conference on developing better QI strategies, but the field will not get there unless we do better QI evaluations. She noted that attendees might feel constrained during the first breakout session by wanting to address the issue of resources and by the discussions having been on the clinical microsystem, and that higher levels of the health care system and

bigger challenges for design would be coming later in the symposium. Dr. Dougherty asked participants to address these other levels and to generate ambitious ideas, but to keep the focus on how to evaluate QI projects. We were asking for solutions to the challenges participants had heard and for big ideas.

Dr. Arnold Milstein moderated the first reports back session, and Dr. Shawna Mercer moderated the second session.

Reports from the Homogeneous Breakout Groups

DISPARITIES RECOMMENDATIONS

Dr. Chin and Dr. Dougherty reported that this was a passionate group that generated eleven major points:

1. What is the appropriate definition for disparities, what constitutes a significant disparity, and which disparities should we take action upon? It would be valuable to set clear expectations for this kind of work. The group did not come to a conclusion for this, but suggested that narrowing the gap between blacks and whites, for example, or raising the floor for the lowest measures of care might be ways to address disparities.
2. The group noted that it is hard to find control groups for disparities research.
3. We need to develop a conceptual model or logic model for disparities. We need to know what causes disparities, in terms of individual factors and larger organizational and societal factors, so that mechanisms can be identified and then tested in interventions. It is important to look at where disparities issues fit within QI. IOM includes equity in its definition of quality, but equity has not been highlighted in other outcomes work. If the Joint Commission on Accreditation of Healthcare Organizations developed a definition of equity, it could be applied to hospitals. QI projects could be required to examine the reduction of disparities.
4. The U01 was suggested as a funding mechanism to establish a network of researchers to develop a logic model for disparities reduction within QI and outside of QI. (The NIH U01 mechanism is a cooperative agreement supporting a specified project. This is a type of grant in which NIH has substantial scientific and programmatic involvement with the institution receiving the award.)
5. Context and structure: there can be disparities within organizations or across organizations. Where the disparity exists and the investigators' model of disparities will affect what the intervention is, such as a provider-level intervention or an organization-level intervention.
6. Data requirements and research question requirements: should racial and ethnic identifiers be included in data collection? Journals and funders should examine whether disparities questions can be required aspects of projects. Another possibility would be requiring community-level needs assessment or a CBPR-type process for funded and published projects. We do not know what pay for performance effects on disparities will be.
7. "Bang for the buck"—one needs to look at difficult settings where there are fewer resources and be willing to take risks. There might be a good payoff from working in such settings.
8. We think that disparities happen in underfunded settings so we need to pay attention to these settings to reduce disparities.

9. The group would like disparities reductions reports from NIH's disparities activities.
10. Other suggestions for increasing a disparities focus within QI were to identify good examples of such work and give awards to people.
11. Finally, a new challenge was identified: how does public health disparities reduction interact with QI?

RESEARCH TRAINING DIRECTORS' RECOMMENDATIONS

Dr. Francis reported that this was a small but diverse group that included individuals involved in training predoctoral students, postdoctoral fellows, physicians, and non-physicians, along with a funder of training programs and one research fellow. This group made eight points:

1. We want to encourage diversity in the training environment and to engage a variety of practitioners, locations, and disciplines. Quality improvement involves messy, "real world" settings, and trainees need to gain experience in such settings, particularly those that may not have a robust infrastructure for data collection.
2. Quality improvement is a topic that crosses disciplinary lines, not a discipline per se. The output of academic training is an independent investigator versed in the methods of their particular discipline. This creates an immediate tension between the goals of the academy and the real-world needs of a health system or community for improvement.
3. There was considerable discussion of whom to train and when in their careers (e.g., pre- or post-doctorate) to train them. This may be better suited for continuing education after one has gained experience in a discipline. This training should include clinicians and practitioners and not just individuals who will engage in QI research.
4. Related to the issue of the proper time to learn QI research skills is how to fit this into one's career trajectory. Finding job security in a traditional discipline is challenge enough. How willingly will individuals seek out non-traditional trajectories, for which the return on investment is less clear? We mused that persons working in this area would need a "portfolio" as opposed to a "curriculum vitae" to be marketable - e.g., short resumes with attached case studies and testimonials about the impact of their QI projects. Some recommended, tongue-in-cheek, that this sort of research be conducted only after one gets tenure.
5. Pursuing this type of research requires a nurturing team and an environment of respect. Part of this is learning to be a good partner with the system you are studying and not overwhelming it with data requests and publications. Building trust means not publicly exposing the system's limitations without some mutually negotiated agreement. Sometimes the organizations with which you are working need to maintain control over the data and dissemination.
6. One wants to make sure that trainees do not become overburdened with service-related demands ("scut" in clinical vernacular). The needs of trainees should be balanced with the needs of the organization.
7. Organizational skills are every bit as important as technical and methodological ones in undertaking such applied work.
8. Just as this conference presented QI implementation case studies from both health systems and community-based perspectives, we need to recognize the overlap of public health and health systems research.

DESIGN EXPERTS' RECOMMENDATIONS

Dr. Leviton and, on Dr. Francis Chesley's behalf, Dr. David Introcaso reported that there is no way to do justice to the richness of the discussion among the design experts. The *challenges* that were identified by the group included the following points.

1. We have difficulty with nested question and theories that apply to the levels of explanation. These range from the local market for insurers and payers to institutions to units and microsystems to patient level characteristics. Currently, the theories are more elaborated for the individual behavior level than they are for organizations and larger systems.
2. There are multiple decision makers with vastly different needs for information about QI studies, ranging from policy makers to specific practitioners. The gold standard of the RCT needs to be matched by something that is legitimized as an alternative and that captures the dimensions of interventions that we feel to be most important.
3. Even for systematic reviews, people find that they do not know how to classify various types of interventions. For example, in diabetes interventions, one can look at case management. On its own, case management does not make a huge difference, but when one looks at specifics such as a nurse providing different medications, this part can make a big difference in the whole intervention.
4. Another challenge is that there is always an interaction between the specific study and the context in which it was conducted, and we need to understand the interaction and do further testing to verify impressions.
5. There were many other issues the group considered to be important, such as the unit of analysis.

The key *recommendations* from the group were that:

1. There was general agreement that we need to change the research platform for this type of study to have a better marriage of practice (including decision makers for different settings) with the research question.
2. We need a more agile set of funders that can take advantage of natural experiments or fund major platforms of research so that many types of questions can be asked of one data set.
3. We need to legitimize the discovery process and work on lines of inquiry that can generate hypotheses and then test them.
4. We need to brainstorm about the mechanisms by which QI occurs, and we need to specify mediators and moderators of changes in quality.
5. We need to brainstorm about new methods for the study of mediators and moderators.
6. We have some ideas about important elements in QI studies, such as the idea that high staff turnover will impair quality improvement efforts.
7. There was much discussion of a compendium or toolkit or toolbox. There is great cumulative knowledge, but we are not fully exploiting what has already been done.
8. There was discussion of the issues concerning systems and process and leadership. Study design needs to include data collection on not only the diagnosis and treatment of a disease but also on social interactions, such as between patient and provider.
9. We need to be more interdisciplinary. We need to be appreciative of Greenhalgh et al.'s work on innovation processes in this reference: Greenhalgh T, Robert G, Macfarlane F,

Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review. *The Milbank Quarterly*. 2004;82(4):581-629.

10. The group noted the need for a shared taxonomy/glossary and databases.

PEER REVIEWERS' AND JOURNAL EDITORS' RECOMMENDATIONS

Dr. Wagner and Dr. Molla Donaldson reported that this group made a series of observations and recommendations.

1. There was strong consensus on the need for a taxonomy/glossary of QI terms which would allow, for example, for labeling types of QI projects and would include descriptions of phases of translational research. Such a taxonomy will be important for reviewers and users of QII.
2. The group discussed the challenges for researchers in getting their grants funded and their papers published. On a more detailed level, one needs to be able to convey in grant applications enough detail about the system changes to be implemented that reviewers can assess their value while allowing for the adjustments in design that will inevitably occur
3. There is an unavoidable tension between presenting both quantitative and qualitative research methods in an understandable way for reviewers who are not familiar with them and fitting this description within the 25-page limit in grant applications and the 2,700 word limit of some print publications.
4. Researchers need to be more aware in writing proposals and manuscripts of the multidisciplinary perspectives of grant and manuscript reviewers. Indeed, papers in this field require transdisciplinary review, perhaps by editorial committees. Dr. Wagner joked that the group briefly considered genetic engineering of reviewers.
5. Two challenges are of particular importance in developing this field as a science. First, most QIIs take place in a limited geographic or organizational context. In both proposals and manuscripts, researchers must address the generalizability of the study design and findings. If successful, is it sustainable? Whether successful or not, what interventions, and in what combination were important, and which were not? To which populations, conditions, or organizations might these interventions apply?
6. Second, a science of QII, like other sciences, needs to build a body of evidence about effective and ineffective interventions and the conditions in which they have been used. The beauty of randomized controlled trials is, of course, that potentially confounding variables are either eliminated by exclusion and inclusion criteria or differ insignificantly in control and experimental groups. In addition, a very restricted set of interventions are held constant. These two conditions, however, are not always (or even usually) possible in QII. This means it is critical to address variations in populations, organizations, and evolving implementation strategies in proposals and manuscripts.
7. In order to build a coherent evidence base, researchers also need to put new projects and new findings in the context of previous findings. One suggestion was that funders and peer reviewers could do more to promote grant proposals and papers that would contribute to cumulateness.
8. There is a need for broader inclusion of users in the planning of improvement projects, in study sections, during studies, and in the editorial review process. Requests for applications (RFAs) and requests for proposals should reflect the broader view of what the end users seek to achieve by these projects.

9. Researchers also need to be more aware of the final end-users of the research; that is, that particularly in this field, publication is not an end in itself or even just a step in further research. The purpose of QII is to discover and deploy useful tools for improving health care.
10. The usual brief summary of methods in publications may not meet users' needs. Users who want to implement change may want and need more detail, but space limitations exist, particularly in peer-reviewed print journals. Possible solutions are splitting articles, online appendices, and splitting write-ups across journals on different topics. One wonders whether it is better to lump into one article or split one's findings into related publications. Other suggestions were that individual articles include the role of theory and how to operationalize theory in an intervention. In addition, articles could include more on an intervention project's history and on the theoretical constructs and assumptions underlying the development of the specific interventions.
11. It also would be valuable to have a publication vehicle to describe early results. Perhaps this could be covered in a special journal section dealing with works in progress. It would be valuable for authors to explain how early studies led to the design of the intervention that was used, to include information on process evaluation, and provide guidelines for others in testing or replicating an intervention.
12. Two different kinds of databases could be very useful. The first is a registry of ongoing studies. Perhaps it would make sense to make registration of studies a condition of publication. Second, it would be valuable to have a data repository of published studies. With appropriate agreements, such data might be used for secondary analysis.
13. A journal's impact factor is important to publishers. The impact factor is based on the number of citations a paper receives, which is a measure of a paper's use to researchers. Might there be a revised (or separate) impact factor that would reflect the study's impact on practice or policy? Similarly, might there be an impact factor on outcomes and on decreases in disparities?
14. Finally, Dr. Donaldson noted that there is a lot of focus on genomics now; in fact, there are 83 "-omics" terms indexed in PubMed. Dr. Donaldson joked that if you want to capture the attention of funders and policy makers, it might be good to change the name of the field to "improve-omics."

USERS' RECOMMENDATIONS

The users group included practitioners, managers, people at the front lines of QI, evidence synthesizers, and people representing patient advocacy groups. Dr. Atkins and Dr. Briss reported on these sessions. The overriding theme of the first breakout session was to encourage paying attention to the issue of spread. In other words, it is important not only to look at whether an intervention works, but also to examine what it takes for something to spread and take hold in a broader environment.

1. The first recommendation was to recognize the importance of engaging clinical users as well as patients and communities in framing questions and interpreting results. PBRNs have an important role in generating clinically important questions from the practitioner level.
2. It is important to realize that the question is not what works, but what works for whom. Understanding effect modifiers is important. This challenge of assessing adaptability and generalizability is more complex in QI research than in many types of clinical studies and

creates challenges in assessing the extent to which work will translate across specific populations and settings.

3. The group recommended creating a toolbox to provide not one fixed intervention package but a more flexible set of options that users can adapt to their skills, needs, setting, resources, etc. One challenge is that we need a better shared taxonomy so that we can determine common components among interventions that would allow users to more consistently assess lessons learned. Toolboxes would make it easier to replicate and implement best practices.
4. Several recommendations were related to spread. We need to understand and support research networks so we can understand how intervention spread works. We need to understand how to support a network in terms of technological needs, but especially in terms of the social aspects of what makes a network or learning community work. We need to understand the incentives that affect willingness to participate in research and to undertake change. These are incentives that relate to patients and community as well as practitioners. The current model for dissemination and translation in terms of journal articles and online material is not adequate for promoting spread. We need additional options. Maybe we need better case studies (something like what one sees in the *Harvard Business Review*) or face-to-face interaction in the collaborative model. This needs to be built into education and training. In other words, “willingness to change” needs to be built in to clinical education. We also need to train researchers on how to work with practices and communities. This will require an understanding of how to build this capacity and to consider who is going to be responsible for building this capacity. Would it be government, funders, or the health systems themselves?
5. We need partnership building leading to partners/networks that survive across projects. We need to engage users during projects and understand the perspectives of various actors.
6. We need to develop conceptual models (and small “t” theories of change, which address how these interventions work and what one expects to change). Theories need to be grounded in the reality of implementation and to reflect that these efforts are not a grand, academic exercise.
7. We need to address data collection and synthesis issues (i.e., what works and for whom). We need better measures of outcomes and better data on intermediate variables, economic costs, and measures of potential harm.
8. We need to capture and share information on harder-to-get-at information such as tacit knowledge and experiential learning (i.e., the “grey” information that does not usually get into publication).
9. Communications need to be segmented according to intended audience.
10. There need to be incentive structures that encourage everyone to participate and that encourage researchers to disseminate.

Discussion period after the first set of reports back

Moderator: Dr. Arnold Milstein

Comments during this session focused on the causes of observed disparities in care, legitimacy of available designs, and the role of journals in QII evaluation dissemination.

A participant asked whether researchers had measured patient preferences to distinguish between differences due to disparities and differences due to preferences, particularly in regard to surgery and antibiotics. To address disparities in a meaningful way, there is a need to identify effective and efficient ways to measure preferences. Dr. Dougherty responded that a literature review by the IOM found that patient preference is not one of the strongest predictors of disparities.

One participant noted that within the design experts' group, there was discussion of increasing the legitimacy of a variety of methods. Often, there is skepticism about non-RCT methods. He asked if members of the peer review group had considered what would make research proposals or the submission of papers to journals with non-RCT methods more acceptable and what it would take to convince people that there is value in these methods and that the findings are generalizable and real. Dr. Wagner indicated there has been a certain degree of progress in this area in terms of nonrandomized studies and studies with qualitative components not being *automatically* rejected, but the medical community still needs to pay attention to how submissions are written in terms of including detailed description of the methods. He recommended being as numerative as possible when writing up qualitative results.

Another participant encouraged others to write articles that include both qualitative and quantitative components. *Annals of Family Medicine* may start a format in which the details of a project are in an online appendix and the printed article will focus on the lessons learned. They may also have sets of three companion papers, including a quantitative article, a qualitative article, and a synthetic article on lessons learned across the qualitative and quantitative methods. The journal is exploring these options. However, if one is going to pursue mixed methods, one needs to do a good job on both the quantitative and qualitative pieces and bring them together well. Others noted that there are good exemplars of in other fields, such as criminal justice and education.

A participant asked if the field is “letting the journals off too easy and the users off too hard.” To do both qualitative and quantitative research well within the same project, one either has to have a great deal of money or ask small questions which can be addressed rigorously, but which may not be as important. She asked how we can benefit from smaller-scale projects that address important questions but do not have a large enough budget to have the highest level of rigor. This is what important user communities are asking: do these projects have to stay outside journals and the learning sphere?

REPORTS FROM THE HETEROGENEOUS BREAKOUT GROUPS

The core planning group met to consider the recommendations from the homogeneous breakout groups, and sorted them into these three topics:

1. What can we do to get our act together as a field (such as creating toolkits, guidelines, methods)?
2. Communicating with/engaging others who are not us.
3. “Growing the tent”/Recruiting others outside our tent.

Two heterogeneous breakout groups were asked to address the 1st topic, one to address the 2nd topic, and two groups to address the 3rd topic.

Dr. Mercer reported back on topic 1: What can we do to get our act together as a field (such as creating toolkits, guidelines, methods)?

This group noted the importance of a taxonomy. It is important to define terms, including what QI is and what it is not. However, we have struggled with the issue of definitions in other burgeoning fields and do not want to exclude anyone by not incorporating their discipline's perspective so there needs to be a lot of caution when proceeding on a taxonomy. We want to be careful about how we define and use terms and we need to know who needs to be at the table. When we are working with a group in a messy, real world situation, we need to define what we mean so we can come to the table with ideas, but we do not want to do it in a way that is exclusive.

We noticed that Deming, for example, has not yet been mentioned in the course of this meeting, and this reminds us that we need to have a sense of history and to keep track of advances in the field to date.

There is a certain messiness when you are doing a project in the real world. One needs to bring in all involved parties at the earliest stage possible, such as bringing in evaluators and users during project planning. We need to think about how the information is going to be used and think of users' needs at the start. We need to think about internal and external validity from the beginning of a project. We may not be able to address both from the beginning, but we do need to think about this context from the outset of a study.

Other areas we discussed included noting that trainees are important. We need to develop future researchers. Some might be from institutions with no background in QI and we should think about how they could use the symposium participants as mentors. We need to give them real world, messy experience in projects.

People need to be able to measure disparities in all their projects.

If projects are to be successful, one needs to consider demands, interests, and incentives of all people involved, including researchers and participants.

We kept coming back to the need to expand our tent to work with others. We need to think about how we fit with others. While we want to define what we are doing, we do not want to be exclusive and separate ourselves from health services research and other areas. We need to think about how we work with people in other fields. Funding agencies need to work together collaboratively and to participate in the process, but not drive the process. Maybe we could expand this to the Centers for Medicare and Medicaid Services and other organizations as well.

Dr. Fine reported back on topic 1: What can we do to get our act together as a field (such as creating toolkits, guidelines, methods)?

In our group, our theme was a focus on research, and one important point that was made was that the definition of health research across the Department of Health and Human Services (HHS) should include what we are talking about at this conference.

We discussed improving or adapting the peer review process for funders and funding mechanisms such that the process better serves QII and is more user-friendly. For example, peer review panels should include transdisciplinary representation. We discussed that there is a need to find a timely way to provide funding for acute events to facilitate their study. We noted that the VA has been wrestling with this with some successes.

When an agency wants to expand the type of research it covers, there should be special training to educate study section members in their new broadened role.

We need to continue to encourage cooperation across the HHS funding agencies. We noted that the VA is a model for QI and AHRQ understands the issues, but some in the group believed that this is less true at NIH.

We should make sure funding can go to anyone who is capable, not just academics. The disease-specific orientation of NIH institutes is a dilemma for this field which cuts across diseases.

We touched on theory—small, medium, and large—and suggested that it would be nice to have funding for theory development by itself, and not just as part of another project.

There were discussions about toolkits, methods, and websites to communicate better about methods and instruments that work well and as a resource for finding them. The VA has one (<http://www.hsrd.research.va.gov/queri/implementation/>), but we need more. Electronic journals are another way to address that need, especially in that they can accept longer papers. One issue for guidelines for publications is to describe the details of an intervention and its context.

One negative point was raised: there were several votes for *not* having a registry of studies. Participants felt that at this point in the field, a registry would not be a useful thing.

Dr. Francis reported back on topic 2: Communicating with/engaging others who are not like us. We are interested in engagement with end users, but we had few end users at the session and in the meeting, so we need to take this dialogue and this kind of conference to end users. Leif Solberg is working to organize just such a meeting.

A second goal that we could set for ourselves is to stop whining. We have been talking about QII from a scarcity paradigm as if we are not understood, there are not enough resources, and there are not enough people doing this kind of work. However, if you look at people who have been successful in doing QIIs, such as administrators and people in operations, they did not bemoan the vagaries of grantsmanship or the lack of academic cachet, but instead sought out the people who had the evidence as well as those willing to support the work. We need to engage with health systems and community leaders and thereby access the people and resources available. Thus, we need to change from a scarcity metaphor to an abundance mentality.

The third issue we discussed was a short-term to mid-term goal. The increasing availability of electronic health information through electronic medical records and other data repositories raises concerns about subject protections and the need to work with IRBs to ensure accountability. In the past, when we considered something QI, as opposed to research, it was

believed that IRB and HIPAA requirements did not pertain. Current IRB structures and processes are not well adapted to QII research, and there is considerable discussion on designing mechanisms more appropriate to such work. Ongoing work on this topic is being funded by the Hastings Center.

For mid- to long-term goals, we discussed that we need new multivalent models of research in which one is doing complex interventions and including end users in study design and even on study sections.

Dr. Green reported back on topic 3: “Growing the tent”/Recruiting others outside our tent. As our broad goal, we looked to expanding the tent in two directions: to end users and to a wider range of disciplines and fields. We identified short-term objectives under each of those goals. Also, we have some intermediate objectives, such as incorporating more of the continuous quality improvement/short loop feedback perspective in schools of public health and other training programs, and more emphasis on process or mediator variables in courses on evaluation, as well as better use of natural experiments. We identified a second intermediate objective: the need to embed QI into health care and public health. Strengthening translational research may be a means to that end.

We identified two major long-term goals: 1) expanding the tent to a wider range of, and more deeply involved, end users, and 2) expanding the tent to more disciplines to incorporate their models, theories, methods, and experiences into the tent. Regarding expanding the tent to include end users, this has to happen early (rather than waiting until the tent is in final order) and we should include non-governmental organizations (NGOs) as partners in developing the tent. We identified 12 specific short-term needs regarding working with NGOs:

1. The need for engagement of users in developing a taxonomy of questions and prioritizing those questions.
2. The need to develop meaningful collaborations.
3. The need to embed users in all phases of the research process.
4. The need for support to develop collaborations.
5. The need for partnership requirements to be built into RFAs.
6. The need to develop standards, guidelines, and criteria for what constitutes a genuine partnership.
7. The need for follow-up and maintenance with partnerships after research.
8. The need to involve end-users in the peer review process.
9. The need to develop training about partnerships for partners.
10. The need to understand how to get along with academics.
11. The need to address how this field will work with the academic promotion and tenure system.
12. The need to assess “what is in it for me?” for each partner.

Regarding the need to incorporate a wider range of disciplines, this group had three examples of short-term objectives:

1. We want to involve evaluators whose work has been in other sectors.
2. We want to include more on QI in HSR training programs and in health policy and health management programs.

3. We had an inconclusive discussion about including health insurers in partnerships. Maybe there is a conflict of interest there.

Dr. Dougherty reported back on topic 3: “Growing the tent”/Recruiting others outside our tent. We chose part of the “expanding the tent” charge to focus on how to develop transdisciplinary theory for QI (but included other topics in our discussion). Everyone agreed that we do need a transdisciplinary theory and practice for this. But the question was, “how do we develop that?”

1. The group recommended having conferences with a focus on users’ needs (such as what does an MD need to know from an anthropologist, a psychologist, or an organizational specialist, and how can we bring their disciplinary languages together?).
2. We examined who should be in the tent. We started making a list of individuals, which seemed overwhelming, so we turned to listing major organizations that should be included. People listed social science lobbying groups and the NIH’s Office of Behavioral and Social Sciences Research and suggested looking to these groups to identify specific people. We need to provide some incentives for getting people to participate, and we need to look at how research teams are structured to make sure that all participants are recognized as peers.
3. Form follows funding so it is up to funders to require transdisciplinary teams. However, funders are influenced by others so we need to reach out to others who can be influential. We may need to make some sort of campaign to make QI important to people, following the model of the IHI 100,000 lives campaign. We might want to bring in AcademyHealth. The tent should not just be focused on the researchers and the evaluators.
4. We need to engage users to bring in a multidisciplinary perspective and we need to be able to use natural experiments.
5. Agencies can be a role model for transdisciplinary work. RWJF is a role model for including a lot of different disciplines who each have something to add.

Plenary Session and Discussion

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 asked a few people in the planning committee to comment on the reports back session regarding the big issues about which we would like to hear cross-talk or underlying or fundamental issues that we have not gotten to yet. He asked them to mention if there are any big underlying questions that have not been addressed in the conference so far.

Dr. Orleans emphasized that one of the most important cross-cutting issues the field faces is disparities, and this should be addressed in every one of the initiatives with which we move forward. Another issue that we should continue to be addressed is consumer involvement or patient involvement. Also, Dr. Francis’s point about linking to operations experts is an important one.

Dr. Wagner summarized that the recurring theme of the conference is that there seems to be an emerging field but there does not seem to be an infrastructure in the funding community and in the academic community to support this emerging field.

Dr. Dougherty hoped to discuss how do we organizationally build a field, move this field, and have a systems perspective. It is clear that health care QI has been slow to adopt a systems perspective, and that has been reflected at this conference. Each person may have thought about what they might do to build the field but not necessarily how they might work with others to build the field. She asked for feedback from the user's perspective. The VA is a model, perhaps because within the VA one is aware of the patients and caretakers every day in that they are part of the VA system. AHRQ is different because it is a research agency. The AHRQ mission is to improve health care for all Americans, and AHRQ needs to work with the people who are doing that work through health plans. Dr. Dougherty would like to hear from people in that field to find out their needs.

Dr. Green noted that the business model for continuous quality improvement continues to be used in the business world, but other than continuing discussion of short-loop feedback cycles, the push within the field of QI for respectability and the pull from outlets for publication result in a tendency for the RCT to emerge as the driving methodology in this area, and that emphasis pushes against the short-loop feedback cycle as the area of emphasis in this field. Dr. Green appealed for ways to keep that on the agenda in the field. Dr. Green was happy to hear that Dr. Brian Mittman and Dr. Martin Eccles are going to include information on process measures, intermediate variables, and moderator variables in articles in *Implementation Science*.

Dr. Francis indicated that while we have been talking about developing a new field, it is imperative not to overlook the importance of generalism. Creating a separate specialty and fragmentation is part of the academic paradigm, but it is one of the barriers to doing QIIs well in health systems and public health venues. Overspecialization may make a person less capable of having an impact where practice occurs. We do not want to become so overspecialized and jargon-ridden that we lose contact with others.

Mr. Chapel opened the floor to comments from all participants.

One participant noted that in regard to engaging the user, it is important to be observant and participatory, and to allocate energy to listening. It is also important to be aware what operations look like and to spend time in the clinical microsystem and the operations context. This participant saw this meeting as the first part of a conversation.

One participant wanted to discuss whether NIH views translational research in the same way as other entities do. The old concept of translational research as “bench to bedside” and getting drugs to market is very different from what has been discussed at this meeting. Is there a need to better interface with that effort in order to accomplish more?

A participant commented that the notion of a field coming together is very important but very tricky. There is no clear definition of what the field really is. The definition does not depend on the methods used, because they are drawn from other fields. The field could be defined by its

goal of improving health care, and this suggests a marriage of patient safety and quality improvement worlds. We should think through the semantics of whether this is quality improvement research, research to improve quality, and whether there are implications of the semantics. Regarding whether there are any voices we have not heard from today, he also noted the absence of manufacturing, meaning those who use quality improvement in a variety of areas and have introduced it to their fields, from this meeting. We should try to learn from those outside the health care field.

A participant noted a paper by Thomas Gilmore of the University of Pennsylvania on field building in philanthropy. Dr. Gilmore uses the example of groups paying more attention to end of life care and makes the point that how the moral imperative is presented is an important part of energizing a field. We should think about what is the moral imperative that has been expressed throughout this meeting. One example is winning the tobacco wars. We are selling ourselves short if we do not make the case for the importance of health care QI in terms of the moral imperative. Analogies to business are useful, but may be overdrawn, particularly for state departments of health or nonprofit organizations where the constraints are different from those in businesses. Sometimes analogies can harm an effort because, for example, one does not run a grass roots effort the way one runs General Motors. We should think about pattern matching in order to determine where such analogies are helpful and where they are not helpful.

A participant remarked that, as a consumer advocate, her goal was to improve health outcomes for everyone. She cautioned against putting too much emphasis on trying to develop this area as an emerging field. She felt that that will take care of itself.

A participant thanked the planners for including consumer organizations and asked participants to seek more partnerships with consumer organizations in designing plans and programs because groups such as hers are advocates. As such, they can advocate for what makes sense to them in having quality health care. They can be champions for this effort if they are included as partners.

A participant advocated including operations people in these efforts. He described an example from his health plan of how partnerships can work. At HealthPartners, they have a new internal grant program with approximately \$250,000 to spend each year on projects requiring partnership between a researcher and an operations person. They have a fast-track review process in which the applications are reviewed and funded within a month of submission.

A participant noted that this effort may not require a new field. What we are talking about is trying to produce predictable change in human behavior in a variety of settings and under a variety of constraints. If we conceptualize the field as the study of human behavior, we have something that already is a scientific field.

Next Steps, Parting Remarks

Dr. Orleans summarized that the organizations that sponsored this symposium are collaborating so we can think about how to do a better job of accelerating quality improvement research. We also are examining how to best disseminate the results of this meeting. She and others from RWJF were very pleased to see their grantees in attendance. The grantees are doing leading work and RWJF will be making an effort to synthesize their findings in order to contribute to the

cumulative evidence from the field. RWJF also is applying quality improvement to reduce disparities in health care and soon will launch a number of new initiatives in this area. Lori Melichar of RWJF has been leading an effort to improve the quality of QI research. RWJF is considering sponsoring a network among its grantees and perhaps among a larger group. Finding a common taxonomy will be critical to this effort, and it must be embedded in the larger forces of health care QI and pay-for-performance initiatives, as well as performance measurement and reporting initiatives. The use of a common measure of system-level supports for quality care is needed to introduce QI into the health care improvement “machinery.” RWJF has funded the National Committee for Quality Assurance (NCQA) to adapt the ACIC (Assessing Chronic Illness Care) measure developed by Edward Wagner, Russell Glasgow, and colleagues into a HEDIS-type measure. This measure assesses the presence of system supports for Chronic Care Model-based care for a range of chronic conditions. The most recent version of this NCQA measure is being validated for use in practices of varying sizes, ranging from small one-to-two physician office practices to larger health plans and preferred provider organizations. Ideally, such a measure could be adopted into the National Quality Forum’s core performance measurement set, as well as into other national “pay for performance” measurement sets. RWJF also is funding a new effort led by the Association of American Medical Colleges, with significant input from the Improving Chronic Illness Care national program to build CQI into the core medical curriculum. Finally, Dr. Orleans noted that federal agencies must work to better address the needs of end-users. She suggested that funders should spend more time with operations personnel and study end-users in their own environments. HealthPartners is planning a meeting that will bring an operations system perspective into sharper focus—this meeting may be a starting point for such efforts.

Dr. Mercer noted that CDC has been involved in conferences such as these because the agency is interested in projects related to translational research. CDC’s vision statement is “healthy people in a healthy world” and it seeks to attain this through all of its research portfolio, its direct support of public health services, and through surveillance. As CDC works with state and local health departments, it is able to observe and determine their needs from the front lines. She mentioned the importance of partnerships and the need to have a seamless interface between public health and the health care system. The issue of external validity is very important to CDC. At this point, quality improvement as a term does not resonate at CDC as much as the concept of translation of research into practice, but it is recognized that the terms have areas of commonality. The issue of disparities also is important to CDC. Quality improvement terminology is creeping into the vernacular of some parts of CDC, such as those involved in developing business case models. And there are some opportunities for QI funding at CDC in areas such as obesity and diabetes where there is close networking between the public health and health care systems.

Dr. Francis pointed out that the VA has a new chief research and development officer, Dr. Joel Kupersmith, who has written a position paper²⁹ proposing an expansion of effectiveness and implementation research through the creation of a public-private partnership. Employers, health plans, pharmaceutical and device companies, and provider organizations all stand to benefit from the implementation of evidence-based practice, and the support of such work would constitute a very small percentage of their total expenditures. Interagency collaboration is also important,

although many structural barriers exist (e.g., coordinating funds transfer and funding cycles is difficult).

Dr. DeVoto stated that NIH is actively involved in research and funding of health care QI, mostly within the context of individual institutes. But, also there are cross-NIH efforts, such as a special interest group that includes individuals at AHRQ, CDC and other federal agencies as well. The special interest group was formed to pool resources and to learn from one another and work together on health care quality improvement. The National Library of Medicine dedicates a lot of resources to health services research, and the Health Services Research Interest Group is working with the library to compile a list of the RFAs and program announcements from across NIH related to health services research and this will be posted on the web.

Dr. Dougherty thanked everyone for their contributions during the symposium, noted that it had been a great success, and adjourned the meeting.

The Core Planning Group

The September 13-15, 2005 symposium was planned by:

Denise Dougherty, Ph.D.
Agency for Healthcare Research and Quality

David Atkins, M.D., M.P.H.
Agency for Healthcare Research and Quality

Peter A. Briss, M.D., M.P.H.
Centers for Disease Control and Prevention

Barbara J. DeVinney, Ph.D.
Project Manager

Emily DeVoto, Ph.D., M.S.P.H.
National Institutes of Health/Office of the Director

Molla Sloane Donaldson, Dr.P.H.
National Institutes of Health/National Cancer Institute

Lawrence J. Fine, M.D., Dr.P.H.
National Institutes of Health/National Heart, Lung, and Blood Institute

Joseph Francis, M.D., M.P.H.
Department of Veterans Affairs

Lawrence W. Green, Dr.P.H.
University of California at San Francisco

Shawna L. Mercer, M.Sc., Ph.D.
Centers for Disease Control and Prevention

C. Tracy Orleans, Ph.D.
Robert Wood Johnson Foundation

Participant List

David Abrams, Ph.D.

Director
Office of Behavioral and Social Sciences
Research
National Institutes of Health
Bethesda, MD

Kathryn Ahlport, M.S.P.H.

Program Manager
Burroughs Wellcome Fund
Health Research Alliance
Research Triangle Park, NC

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

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

Anne-Marie J. Audet, M.D., M.Sc.

Vice President
Quality Improvement and Efficiency
The Commonwealth Fund
New York, NY

James B. Battles, Ph.D.

Senior Service Fellow for Patient Safety
Center for Quality Improvement
and Patient Safety
Agency for Healthcare Research and Quality
Rockville, MD

Arne Beck, Ph.D.

Director of Research
Clinical Research Unit
Kaiser Permanente
Aurora, CO

Steven Berman, M.A.

Executive Editor
Joint Commission Resources,
Publications
Joint Commission Journal on Quality and
Patient Safety
Oakbrook, IL

Carolyn Berry, Ph.D.

Associate Research Professor
Center for Health and Public Service
Research
New York University
New York, NY

Joseph Betancourt, M.D., M.P.H.

Director
Disparities Strategies Center
Institute for Health Policy
Massachusetts General Hospital
Boston, MA

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

Chief
Community Guide Branch
Centers for Disease Control and
Prevention
Atlanta, GA

Peter I. Buerhaus, Ph.D., R.N., FAAN

Senior Associate Dean for Research
School of Nursing
Vanderbilt University
Nashville, TN

Ned Calonge, M.D., M.P.H.

Chief Medical Officer
and State Epidemiologist
Executive Director's Office
Colorado Department of Public Health
and Environment
Denver, CO

Alan James Roy Cameron, Ph.D.

Executive Director
Centre for Behavioural Research
and Program Evaluation
University of Waterloo
Waterloo, Ontario
Canada

Thomas J. Chapel, M.A., M.B.A.
Senior Health Scientist
Office of the Director
Office of Strategy and Innovation
Centers for Disease Control and Prevention
Atlanta, GA

Martin P. Charns, D.B.A., M.B.A.
Director
VA Center for Organization, Leadership,
and Management Research
Boston VA Healthcare System
Boston, MA

Francis D. Chesley, Jr., M.D.
Director
Office of Extramural Research, Education,
and Priority Populations
Agency for Healthcare Research and Quality
Rockville, MD

Yen-pin Chiang, Ph.D.
Health Scientist Administrator
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality
Rockville, MD

Marshall Chin, M.D., M.P.H.***
Associate Professor of Medicine
University of Chicago
Chicago, IL

Carolyn Clancy, M.D.
Director
Agency for Healthcare Research
and Quality
Rockville, MD

David Colton, Ph.D., M.P.A.
Information Specialist
Commonwealth Center for Children and
Adolescents
Staunton, VA

Lawton Cooper, M.D., M.P.H.
Medical Officer
Division of Epidemiology
and Clinical Applications
National Heart, Lung, and Blood Institute
Bethesda, MD

**Margaret Coopey, M.P.S., M.G.A.,
R.N.**
Senior Health Policy Analyst
Center for Outcomes and Evidence
Agency for Healthcare Research
and Quality
Rockville, MD

Benjamin Crabtree, Ph.D.
Professor
Department of Family Medicine
The Robert Wood Johnson
Medical School
University of Medicine and Dentistry
of New Jersey
Somerset, NJ

Frank Davidoff, M.D., MACP
Executive Editor
Annals of Internal Medicine
Institute for Healthcare Improvement
Wethersfield, CT

Kelly J. Devers, Ph.D.
Associate Professor
Department of Health Administration and
Family Medicine
Virginia Commonwealth University
Richmond, VA

Barbara DeVinney, Ph.D.*
Project Manager
Office of Behavioral and
Social Sciences Research
National Institutes of Health
Christiansburg, VA

Emily DeVoto, Ph.D., M.S.P.H.*
Senior Advisor for Public Health Policy
Office of Medical Applications of Research
National Institutes of Health
Rockville, MD

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

Senior Scientist for Quality of Care Research
and Policy
Outcomes Research Branch
Division of Cancer Control
and Population Sciences
National Cancer Institute
Bethesda, MD

Denise Dougherty, Ph.D.*

Senior Advisor
Child Health and Quality Improvement
Office of Extramural Research, Education, and
Priority Populations
Agency for Healthcare Research
and Quality
Rockville, MD

Marian F. Earls, M.D., M.T.S., FAAP

Medical Director
Access II/III
Guilford Child Health, Inc.
Greensboro, NC

Martin Eccles, M.D., M.B.

Professor of Clinical Effectiveness
Centre for Health Services Research
University of Newcastle upon Tyne
Newcastle upon Tyne
England

Elizabeth Edgerton, M.D., M.P.H.

Director of Clinical Prevention
Center for Primary Care, Prevention,
and Clinical Partnerships
Agency for Healthcare Research and Quality
Rockville, MD

Paula Einhorn, M.D., M.S.

Medical Officer
Division of Epidemiology
and Clinical Applications
National Heart, Lung, and Blood Institutes
Bethesda, MD

Rachael Evans, M.P.A.

Health Science Specialist
Health Services Research
and Development
Department of Veterans Affairs
Washington, DC

Lawrence J. Fine, M.D., Ph.D.*

Scientific Group Leader
Clinical Prevention and Translation
National Heart, Lung, and Blood Institute
Bethesda, MD

Ann Barry Flood, Ph.D.

Professor of Community
and Family Medicine
Center for Evaluative Clinical Sciences
Dartmouth Medical School
Hanover, NH

Steven Fox, M.D., M.P.H., S.M.

Center for Outcomes and Evidence
Agency for Healthcare Research and Quality
Rockville, MD

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

Associate Director
Office of Research and Development
Health Services Research and Development
Department of Veterans Affairs
Washington, DC

Constance Fung, M.D., M.S.H.S.

Assistant Professor of Medicine
Department of Medicine/Health
VA Greater Los Angeles Healthcare
System/RAND
Santa Monica, CA

Theodore G. Ganiats, M.D.

Professor and Interim Chair
Family and Preventive Medicine
University of California at San Diego School of
Medicine
La Jolla, CA

Daniel S. Gaylin, M.P.A.

Senior Vice President and Director
Health Survey, Program,
and Policy Research
National Opinion Research Center
Washington, DC

Rosemary Gibson, M.S.

Senior Program Officer
The Robert Wood Johnson Foundation
Princeton, NJ

Alan S. Go, M.D.

Research Scientist II and Senior Physician
Division of Research
Kaiser Permanente of Northern California
Oakland, CA

Marsha Gold, Sc.D.

Senior Fellow
Mathematica Policy Research, Inc.
Washington, DC

Junius Gonzales, M.D.**

Division of Services and Intervention Health
National Institute of Mental Health
Bethesda, MD

Lawrence Green, Dr.P.H.*

Adjunct Professor
Epidemiology and Biostatistics
School of Medicine and Comprehensive Cancer
Center
University of California at San Francisco
San Francisco, CA

David Haggstrom, M.D., M.C.R.

Cancer Prevention Fellow
Applied Research Program
National Cancer Institute
6130 Executive Boulevard, MSC 7344
Bethesda, MD

Michael I. Harrison, Ph.D.

Senior Research Scientist
Organizational and Systems Research
Center for Delivery, Organization,
and Markets
Agency for Healthcare Research and Quality
Rockville, MD

LeRoi S. Hicks, M.D., M.P.H.

Instructor
Health Care Policy
Harvard Medical School
Boston, MA

**M. Carolina Hiestrosa,
M.A., M.P.H.**

Executive Vice President
Programs and Planning
National Breast Cancer Coalition
Washington, DC

Richard Hirth, Ph.D.

Associate Professor
Health Management and Policy
University of Michigan School
of Public Health
Ann Arbor, MI

Charles J. Homer, M.D., M.P.H., FAAP

President and Chief Executive Officer
National Initiative for Children's
Healthcare Quality
Cambridge, MA

David S.P. Hopkins, M.D., M.P.H.**

Staff Scientist
Community Guide Branch
Centers for Disease Control and Prevention
Portland, OR

Susan Horn, Ph.D.

Vice President, Senior Scientist
Institute for Clinical Outcomes Research
International Severity Information
Systems, Inc.
Salt Lake City, UT

Kelly Hunt, M.P.P

Research Officer
Research and Evaluation
Robert Wood Johnson Foundation
Princeton, NJ

David M. Introcaso, Ph.D.

Evaluation Officer
Office of Performance, Accountability,
Resources, and Technology
Agency for Healthcare Research and Quality
Rockville, MD

Maulik Joshi, Dr.P.H.

President and Chief Executive Officer
Delmarva Foundation for Medical Care, Inc.
Easton, MD

Douglas Kamerow, M.D., M.P.H.

Chief Scientist
Health, Social, and Economics Research
RTI International
Washington, DC

Emmett Keeler, Ph.D.**

Senior Mathematician
The RAND Corporation
Santa Monica, CA

Catarina I. Kiefe, M.D., Ph.D.

Professor and Director
Division of Preventive Medicine
University of Alabama at Birmingham
Birmingham, AL

Lawrence Kleinman, M.D., M.P.H.

President and Chief Executive Officer
Quality Matters, Inc.
Allentown, PA

Uma R. Kotagal, M.B.B.S., M.Sc.

Vice President for Quality and
Transformation
Department of Clinical Effectiveness
Cincinnati Children's Hospital Medical
Center
Cincinnati, OH

Bruce E. Landon, M.D., M.B.A.

Associate Professor of Healthcare Policy
Harvard Medical School
Boston, MA

Carole Lannon, M.D., M.P.H.

Co-Director
The Center for Children's Healthcare
Improvement
University of North Carolina at Chapel Hill
Chapel Hill, NC

Scott Leischow, Ph.D.

Senior Advisor for Tobacco Policy
Office of the Secretary
U.S. Department of Health
and Human Services
Rockville, MD

Laura C. Leviton, Ph.D.**

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

Terry Lied, Ph.D.

Statistician
Office of Clinical Standards and Quality
Quality Improvement Group
Center for Medicare & Medicaid Services
Baltimore, MD

Nancy Maher, Ph.D.

Portfolio Manager, Rehabilitation Research
Office of Research and Development
Department of Veterans Affairs
Washington, DC

Jill Marsteller, Ph.D., M.P.P.

Assistant Professor
Health Policy and Management
Johns Hopkins Bloomberg School
of Public Health
Baltimore, MD

Geraldine McGlynn, M.Ed.

Director
Center for Information Dissemination
and Education Resources
Boston Veterans Affairs Healthcare
Systems
Boston, MA

Linda McIvor, M.S., M.H.S.

Program Manager
QUERI and SDP Programs
Health Services Research & Development
Department of Veterans Affairs
Washington, DC

Lori Melichar, Ph.D.

Economist
Department of Research and Evaluation
The Robert Wood Johnson Foundation
Princeton, NJ

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

Health Scientist
Office of the Chief Science Officer
Centers for Disease Control and Prevention
Atlanta, GA

Arnold Milstein, M.D., M.P.H.

Natural Healthcare Thought Leader
Health and Group Practice
Mercer Human Resource Consulting
San Francisco, CA

Brian S. Mittman, Ph.D.

Senior Social Scientist
Center for the Study of Healthcare
Provider Behavior
VA Greater Los Angeles Healthcare System
Sepulveda, CA

Patricia Dolan Mullen, M.P.H., Dr.P.H.

Professor
Division of Health Promotion
and Behavioral Sciences
University of Texas School of Public Health
Houston, TX

Alvin I. Mushlin, M.D., Sc.M.

Chairman
Department of Public Health
Weill Medical College of Cornell University
New York, NY

Jack Needleman, Ph.D.

Associate Professor
Department of Health Services
University of California at Los Angeles School
of Public Health
Los Angeles, CA

**Mary P. Nix, M.S., M.T.(A.S.C.P),
S.B.B.**

Service Fellow
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality
Rockville, MD

Donna Gore Olsen

State Coordinator
Family Voices of Indiana
Indianapolis, IN

C. Tracy Orleans, Ph.D.*

Senior Scientist/Senior Program Officer
Department of Research and Evaluation
The Robert Wood Johnson Foundation
Princeton, NJ

Steven Ornstein, M.D.**

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

Elizabeth Ozer, Ph.D.

Associate Professor of Pediatrics
University of California at San Francisco
San Francisco, CA

Debra Joy Perez, Ph.D., M.A., M.P.A.

Program Officer
The Robert Wood Johnson Foundation
Princeton, NJ

Barry Portnoy, Ph.D.

Senior Advisor for Disease Prevention
Office of Disease Prevention
National Institutes of Health
Bethesda, MD

Lloyd P. Provost, M.S.

Senior Fellow
Institute for Health Care Improvement
Austin, TX

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

Dean
Office for Diversity and Community
Partnership
Harvard Medical School
Boston, MA

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

Professor
Health Services Research
VA Greater Los Angeles Healthcare System
North Hills, CA

Mary C. Ruhe, R.N.**

Department of Family Medicine
Research Division
Case Western Reserve University
Cleveland, OH

Susanne Salem-Schatz, Sc.D.

Principal
Health Care Quality Initiatives
Newton, MA

Karen M. Sanders, M.D.

Acting Chief
Office of Academic Affiliations
Veterans Health Administration
Richmond, VA

Rob Sanson-Fisher, Ph.D.

Professor of Health Behavior
School of Medicine
University of Newcastle
Newcastle, New South Wales
Australia

**Stephen C. Schoenbaum, M.D.,
M.P.H.**

Executive Vice President for Programs
The Commonwealth Fund
New York, NY

F. Douglas Scutchfield, M.D.

Peter P. Bosomworth Professor
College of Public Health
University of Kentucky
Lexington, KY

Kaveh G. Shojania, M.D.

Assistant Professor
The Ottawa Hospital - Civic Campus
University of Ottawa
Ottawa, Ontario

Lisa A. Simpson, M.B., B.Ch., M.P.H.

Professor and All Children's Hospital
Endowed Chair for Child Health Policy
Department of Pediatrics
University of South Florida
St. Petersburg, FL

Jean Slutsky, P.A., M.S.P.H.

Director
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality
Rockville, MD

Jane Smith, M.S.C.

Deputy Editor
British Medical Journal
London
England

Leif I. Solberg, M.D.

Associate Medical Director for Care
Improvement Research
Health Partners Research Foundation
Minneapolis, MN

Robert Spengler, Sc.D.

Director
Office of Public Health Research
Centers for Disease Control and Prevention
Atlanta, GA

Theodore Speroff, Ph.D.

Director
Center for Patient Healthcare Behavior
Vanderbilt/Veterans Affairs
Tennessee Valley Healthcare System
Geriatric Research, Education and Clinical
Center/Health Services Research
Nashville, TN

Mark Splaine, M.D., M.S.

Co-Director
Center for Evaluative Clinical Sciences
VA National Quality Scholars Program
Dartmouth Medical School
Hanover, NH

Kurt C. Stange, M.D., Ph.D.**

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

Cheryl B. Stetler, Ph.D., M.A.

Evidence-Based Practice and Evaluation
Consultant
QUERI Program
Veteran Health Administration
Amherst, MA

David P. Stevens, M.D.

Vice President
Clinical Care Improvement
Association of American Medical Colleges
Washington, DC

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

Assistant Professor
Division of Pediatrics
Medical University of South Carolina
Charleston, SC

Stephen H. Taplin, M.D., M.P.H.

Senior Investigator
Applied Research Program
National Cancer Institute
Rockville, MD

Neil Thakur, Ph.D.

Assistant Director for Implementation
Health Services Research and Development
Department of Veterans Affairs
Washington, DC

Martina Vogel-Taylor, M.T.

Senior Advisor for Disease Prevention
Office of Disease Prevention
National Institutes of Health
Bethesda, MD

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

Director
The Center for Health Studies
Group Health Cooperative
MacColl Institute for Healthcare Innovation
Seattle, WA

Paul Wallace, M.D.

Executive Director
Care Management Institute
Kaiser Permanente
Oakland, CA

Abraham Wandersman, Ph.D.

Department of Psychology
University of South Carolina
Columbia, SC

Gayle Weaver, Ph.D.

Associate Professor
School of Allied Health
Division of Rehabilitation Sciences
University of Texas Medical Branch
Galveston, TX

Robin Weinick, Ph.D.

Institute for Health Policy
Massachusetts General Hospital
Boston, MA

Barbara Wells, Ph.D.

Health Services Researcher
Division of Epidemiology
and Clinical Applications
National Heart, Lung, and Blood Institute
Bethesda, MD

Sankey Williams, M.D.

Sol Katz Professor of Medicine
Associate Editor
Annals of Internal Medicine
University of Pennsylvania
Philadelphia, PA

Steven H. Woolf, M.D., M.P.H.

Professor
Family Practice, Epidemiology
and Community Health
Virginia Commonwealth University
Richmond, VA

- * Core Planning Group
- ** Presenters
- *** Commentators

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

- ¹ Institute of Medicine. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: National Academy Press; 2003.
- ² Institute of Medicine. *To Err is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 1999.
- ³ Institute of Medicine. *Priority Areas for National Action: Transforming Health Care Quality*. Washington, DC: National Academy Press; 2003.
- ⁴ Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press; 2001.
- ⁵ 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.
- ⁶ Agency for Healthcare Research and Quality. National Healthcare Disparities Report. http://www.qualitytools.ahrq.gov/disparitiesreport/download_report.aspx. Accessed June 10, 2004.
- ⁷ Agency for Healthcare Research and Quality. National Healthcare Quality Report. http://www.qualitytools.ahrq.gov/qualityreport/download_report.aspx. Accessed June 10, 2004.
- ⁸ 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.
- ⁹ 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.
- ¹⁰ Mohr JJ, Batalden P, Barach P. Integrating patient safety into the clinical microsystem. *Qual Saf Health Care*. 2004;13(Suppl II):ii34-ii38.
- ¹¹ 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.
- ¹² Balas AE, Boren S. Managing Clinical Knowledge for Health Care Improvement. *Yearbook of Medical Informatics 2000*. Stuttgart, Germany: Schattauer; 2000.
- ¹³ Wilson EO. *Consilience: the unity of knowledge*. New York: Kopf; 1998:59.
- ¹⁴ 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.
- ¹⁵ 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.
- ¹⁶ 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.
- ¹⁷ Wilbur, K. *Sex, Ecology, Spirituality*. 1995/2000, Boston: Shambhala Publications, Inc.
- ¹⁸ Wilbur, K. *A Brief History of Everything*. 1996, Boston: Shambhala Publications, Inc.
- ¹⁹ Stange KC, Miller WL, McWhinney IR. Developing the knowledge base of family practice. *Fam Med*. 2001;33(4):286-297.

-
- ²⁰ Thomas P, Griffiths F, Kai J, O'Dwyer A. Networks for research in primary health care. *BMJ*. 2001;322:588-590.
- ²¹ Thomas P. *Integrative Primary Health Care: Leading, Managing, Facilitation*. London: Radcliff Press, 2006.
- ²² 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.
- ²³ 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.
- ²⁴ 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.
- ²⁵ Ovretveit J, Bate P, Cleary P, et al. Quality Collaboratives: Lessons From Research. *Quality And Safety In Health Care*. 2002;11:345-351.
- ²⁶ 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.
- ²⁷ 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.
- ²⁸ Novotny TE, Seigel MB. California's Tobacco Control Saga. *Health Affairs*. 1996;15:58-72.
- ²⁹ 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.