Conference:
Physician-Level Interventions:
What Works to Improve Quality of Care

Principal Investigator: Daniel Wolfson
Team Members: Christine Cassel, Ann Greiner, Eric Holmboe
Organization: American Board of Internal Medicine
Dates of Project: September 30, 2011 – September 29, 2012
Grant Award Number: 1R13HS020565-01A1

I. Structured Abstract

On November 18, 2011, the American Board of Internal Medicine (ABIM) sponsored a conference, entitled *Physician-Level Assessment & Recognition Programs: What Works?* The conference was held in Washington, DC, and comprised a mix of health services researchers, policymakers, and private sector leaders. ABIM commissioned two health services researchers to begin researching and writing a paper on existing market, regulatory, and professional standard-setting interventions on the individual physician level; their initial research findings were presented at the conference. The conference participants debated the implications of the emerging research, discussed how research and policy can interact in a more effective way, and provided comments and suggestions to guide the next stage of research.

II. Purpose

Following the 2001 report by the Institute of Medicine, *Crossing the Quality Chasm: A New Health System for the 21st Century*, there has been heightened interest on the part of public and private payers to address quality issues that stem from a federal healthcare system widely recognized to be disorganized and uncoordinated. This fractured system has prompted an emerging multitude of creative and diverse initiatives to address quality and cost problems affecting healthcare, with an increasing number targeting individual physicians because of their central role in healthcare decision making. On the private side, these interventions include efforts to collect and report physician-level clinical data and patient experiences, incentive programs (including pay for reporting/performance), and efforts to redesign care delivery and the care team (e.g., Accountable Care Organizations and the Patient-Centered Medical Home). The federal government has leveraged its market and regulatory clout in these areas as well, and the Affordable Care Act expanded more upon existing federal physician-directed efforts.

Although the quantity of research is growing, there remains limited research about the efficacy of these market and regulatory interventions. Historically, health service researchers interested in physician assessment have worked in parallel but not necessarily overlapping domains. For examine, economists analyzing the effects of pay-for-performance interventions have tended to publish in econometric or policy-related journals, such as *Health Services Research*, whereas those interested in innovation in testing

have published elsewhere, such as *Journal of Computerized Adaptive Testing* or *Applied Measurement in Education*. This project brought together different bodies of research literature that shared the common aim of documenting innovation and examining the efficacy of physician-level quality improvement efforts. Extending the research across types of assessment areas provided a richer, more comprehensive picture of the different kinds of interventions and helped identify what aspects of the various approaches held the most promise for improving physician quality.

This project focused on the development and dissemination of research that examines the evidence related to a targeted number of interventions focused on enhancing physician performance. More specifically, this grant was used to convene a conference of experts and health policymakers to examine and discuss initial research examining physician-level quality interventions and provide insight for future research.

III. Scope

Federal and state governments, as well as private employers by proxy of their health plans, have embraced a variety of innovative interventions – pay for performance, public reporting, patient-centered medical homes – in their attempt to enhance quality and curb the growth of healthcare spending. Because of their central role in driving medical decision making and the potential to dramatically impact quality and cost, an increasing number of these quality interventions have been targeted at individual physicians. Substantial public and private dollars are expended to design, implement, and incentivize various physician-level interventions. Research is needed to inform policy decisions so that they may be evidence based, to the extent possible.

The overall aim of the project was to examine and evaluate the efficacy of various clinician-directed interventions and to inform future strategies and decisions of public and private payers and peer regulators. However, the more limited project scope of the grant was to provide a forum for a select group of health policymakers, health services researchers, and leaders in the private and public markets to discuss the current environment of physician quality initiatives, share best

practices of quality improvement interventions, and start an open dialogue about the inherent tension between health service researchers and health policymakers.

IV. Methods

The overall goal of the project was to produce an initial document within 6 months that delineated the current environment of physician-level quality interventions: what was known and not known about a specified set of recognized and well-defined market, regulatory, and professional standard-setting interventions. ABIM selected two well-known, highly experienced health services researchers. The researchers reviewed and analyzed empirical, quantitative studies, reviewing the following interventions directed at the individual physician level: assessment and feedback, pay for performance, certification, public reporting, and pay for reporting. These interventions were selected because they reached the largest number of physicians, were both widely embraced and varied in approach, and were likely to be further implemented under directives in the Affordable Care Act. In many cases, their implementation and maintenance were supported by federal dollars. A key challenge was comparing across different intervention types, particularly due to their different specific aims (beyond the general goals of improving the quality of care and/or reducing cost) and diverse metrics.

ABIM also developed a conference plan, inviting a diverse mix of health policy researchers, health policymakers, and leaders from the private and public sectors, with the goal of debating the emerging research and discussing whether the strategies currently utilized to enhance the quality of physician performance provided the best approach to advance healthcare reform. The conference proceedings were written and distributed to conference attendees to guide and encourage future research. Following the conference, ABIM asked all meeting participants to complete an online conference evaluation survey, to rate the panelists and the presentations, and to provide detailed comments. The survey was sent to each participant, with a reminder and a link sent again at a later point in time.

V. Results

First, it is fair to say that high-quality evidence to support any of these interventions is scant, making robust comparison impossible. Second, with that caveat in mind, some of the strongest evidence supported the role of professional identity and standards as powerful motivators to participate in performance improvement activities. The group discussed the promises and challenges of physician-level interventions against the background of these tentative results.

The Research Challenge

There was broad consensus at this conference that we don't really know whether efforts to assess individual physicians are improving quality. This uncertainty prompted a call for more research into the impact of programs such as assessment and feedback, P4P, certification, public reporting and pay for reporting, with a focus on more detailed information about why particular interventions were succeeding or failing – from a failure to communicate findings to physicians to the weakness of incentives.

Alongside the call for more research, however, came a series of suggestions that the traditional medical research model may be ill suited to evaluating these interventions.
Many argued that the level of rigor researchers bring to evaluate the success or failure of medical interventions may be unnecessary and even counterproductive, especially given the tension between policymakers' need for fast feedback about program successes and failures and researchers' preference for complex models and longer time frames. There were counterarguments, however, that the validation offered by traditional peer-reviewed research is too valuable to abandon and that research standards need to be preserved. From both sides, there was widespread agreement that the research community needs to be more creative and experiment with innovative new designs, particularly those that take advantage of the time- and space-collapsing properties of modern technology.

This discussion prompted ideas for how to improve health services research. One was the creation of a PubMed equivalent to make it easier to search what already exists in the field. A second was for a "national evidence accelerator" that would allow researchers to get their work to colleagues far faster in a mode of virtual peer review that could coexist with the traditional

track of peer-reviewed publication in major journals. This could also provide an avenue for clinical systems and insurance companies (who hold much of the relevant data about health system interventions but frequently lack an inclination to publish it) to make their data available to health systems researchers, who could then more rigorously evaluate potentially promising data.

The dialogue also led to the idea—borrowed from the legal context—that research regarding health system interventions be evaluated by a 'civil' standard (i.e., more likely than not to be successful) than a 'criminal' one (i.e., proven successful beyond a reasonable doubt). A number of participants also advocated for a greater reliance on case studies to support the adoption of particular interventions, such as the suggestion that policy setters accept a series of positive case studies as sufficient evidence to embrace a particular intervention.

The uncertainty regarding the impact of interventions aimed at individual physicians caused a number of participants to suggest that the individual is not the proper unit of analysis. Instead, they suggested, we need to examine how interventions influence the systems in which physicians operate. Though we may lack confidence about the effect of interventions on individual physicians, there is ample reason to believe that such incentives can motivate change in systems.

Motivation: Is There a Disconnect?

The meeting featured a rich discussion about the factors that motivate physicians and whether existing efforts to inspire improved performance through assessment, reporting, and payment properly accounted for what we know about motivation. Motivation can be intrinsic or extrinsic, and efforts to shape physicians' choices through financial incentives and penalties are classic examples of extrinsic motivation. Although there is no doubt that money is a motivator for physicians—as it is for most people—numerous conference participants argued that the evidence demonstrates that intrinsic motivators such as professionalism and a culture of peer pressure have the deepest and most persistent influence on physician behavior. One example is the powerful impact of sharing performance data internally in medical group practices, where quality improves without financial incentives. Another example of this influence is the willing response

of board-certified physicians to the significant requirements they must meet in order to maintain their certification. The vast majority of physicians voluntarily undertake re-certification activities, and they report that their primary motivation for doing so is to satisfy their sense of professional identity rather than to meet external expectations or requirements.[i]

An expert in motivational theory discussed the three major intrinsic motivators characterized in self-determination theory: mastery, purpose, and autonomy. Of these, he particularly emphasized the importance of autonomy, not only in initiating change but in maintaining it, especially as financial incentives for desired behaviors diminish or disappear over time. In fact, financial incentives can have the perverse effect of "crowding out" intrinsic motivation for desired behavior, converting acts traditionally performed for altruistic reasons into commercial transactions, with blood donation a classic example. Under this "crowding out" phenomenon, the elimination of rewards causes performance to revert to levels even lower than when the incentives were originally instituted.

Like all people, physicians are more likely to embrace potential changes that reinforce their sense of mastery, purpose, and impact/effectiveness, and numerous participants stressed the critical importance of that physician embrace. There are techniques that can amplify and harness that sense; for example, to engage physicians in quality improvement efforts, the goals must be clearly communicated to them, and feedback should be provided as quickly as possible. This way, physicians can internalize and take ownership of desired behaviors and reduce the feeling that they are being controlled or manipulated by outsiders. As discussed more fully in the following section, there was no disagreement that the current reality—in which physicians operate among an opaque set of incentives of which they may be only dimly aware—falls far short of the conditions required to achieve true acceptance and ownership.

P4P

Much of the conversation at the conference focused on P4P, a major focus for public and private payers. Evaluating the effectiveness of generic P4P is particularly challenging, because the larger concept of paying for performance encompasses a host of individual programs with different

components, and the same physicians can be subject to multiple approaches. Suggested reasons why it might not work to improve quality included the following: (1) P4P program incentives are often very weak and get lost in the shuffle of all of the incentives that are already built into the healthcare system; (2) the measures used are not meaningful or motivating to physicians; (3) the measures are not well linked to clinical outcomes; and (4) execution problems, such as feedback mechanisms that physicians do not find useful, exist. Another participant suggested that we have unrealistic expectations for P4P and should not expect a bonus system designed to reward individual physicians to address the systemic problems of our healthcare system.

One medical educator suggested that, if P4P were evaluated by the standards used to assess educational interventions, they would fail. The goal of the interventions—overall and continuous quality improvement and transformation of clinical practices to achieve it—is in fact overwhelmed and masked by requirements to measure and report discrete and disconnected bits of data. As a result, physicians do not regard these programs as significant to their practice of medicine. Indeed, many physicians are unaware that they are even part of an intervention program supposedly influencing their behavior.

Although many were skeptical about the value of P4P and other financial incentives, a number of attendees believed that such interventions are in fact necessary to achieve quality improvements. These observers argued that financial incentives such as P4P are needed to get the attention of most doctors and open the door to educating them about quality improvement. (A number of participants said the same about maintenance of certification.) A health plan representative suggested that her organization's P4P programs have been effective, primarily because of an acute focus on ensuring physician awareness of the program and an opportunity for physicians to review and accept the program.

Impact on Physicians

Having considered the challenges confronting those who want to know what works to improve quality and the ways in which our current efforts in quality improvement succeed or fail in leveraging what we know about physician motivation, the group discussed how the field should

react. A number of participants discussed the need to change the physician job description to persuade physicians that they bear an obligation both toward their patients as they work their way through the entire healthcare system and toward the population as a whole. One government representative suggested that, if we can change what excellence means for a physician, and include a commitment toward care coordination and resource use accountability, many of the problems facing the healthcare system would disappear. This process would start at the medical education level, where one participant said we need to convey that medicine is a "team sport" and the "team is accountable for patient outcomes."

Others expressed concern about the burdens placed on physicians through these attempts to measure individual performance, especially considering the unclear evidence regarding their effectiveness. For example, one participant suggested that it would be appropriate to "hit the pause button" on P4P until we can learn more about whether and how existing efforts are working, including any unintended consequences. At the same time, he and other commenters supported moving forward with efforts to promote organized systems of care, which can appropriately enable improvement and speed the cultural change required to reshape the physician job description. A related suggestion to ease the burden on physicians involved the prospect of compensating them for the increased efforts involved in quality improvement programs by reducing other sorts of demands. For example, one participant relayed how an insurer agreed not to conduct a utilization review as a reward for improved quality performance.

With broad support for the benefits of moving toward more organized systems of care, some in the group expressed severe skepticism about the fate of the small practice and its ability to adjust to the increasing demands of these physician interventions. This view was not universal: one health plan representative said that, contrary to initial expectations, small practices had succeeded in the plan's quality improvement activities. Others believed that there were solutions to some of the hurdles these practices face. A number believed that virtual integration and organizational supports could help small practices adjust by allowing them to benefit from some of the economies of scale that larger systems experience. When the conversation turned toward whether small practices should be formally discouraged, another participant suggested that

payers should focus on setting appropriate payment policies and leave physicians to sort out practice structure.

Post-Conference Survey

The responses to the post-conference survey were generally positive, with 90 percent of respondents agreeing or strongly agreeing that the meeting provided valuable information on physician-level assessment and recognition programs. The respondents felt that the initial research was "too limited" but allowed for "provocative and creative thinking about the implications," while others felt that hearing comments about the limitations of the study "didn't help us figure out…a path forward for driving provider change" and "limited the utility of the session." There was overall agreement that there is a need for creative solutions to quickly disseminate research results while maintaining the integrity associated with peer review. The discussion panelists were all rated positively, and by far the most popular discussion was about the theories behind motivation and human behavior, with one respondent saying, "it was one of the most interesting sessions I have ever attended." The survey asked respondents to list follow-up activities they would like to see happen, which generated many responses, including the creation of an open virtual library or forum to share activities, tools, strategies, and evidence used to bring about improvement and to increase transparency.

VI. List of Publications & Products

The final paper is being resubmitted for publication to the *Annals of Internal Medicine*.

[i]Lipner RS, Bylsma WH, Arnold GK, Fortna GS, Tooker J, Cassel CK. Who is maintaining certification in internal medicine--and why? A national survey 10 years after initial certification. Ann Intern Med [Internet]. 2006 Jan 3;144(1):29-36.