June 2011

Issue Brief

Identifying, Monitoring, and Assessing Promising Innovations: Using Evaluation to Support Rapid-Cycle Change

Marsha Gold, David Helms, and Stuart Guterman

ABSTRACT: The Center for Medicare and Medicaid Innovation (Innovation Center) was created by the Affordable Care Act to identify, develop, assess, support, and spread new approaches to health care financing and delivery that can help improve quality and lower costs. Although the Innovation Center has been given unprecedented authority to take action, it is being asked to produce definitive results in an extremely short time frame. One particularly difficult task is developing methodological approaches that adhere to a condensed time frame, while maintaining the rigor required to support the extensive policy changes needed. The involvement and collaboration of the health services research community will be a key element in this endeavor. This issue brief reviews the mission of the Innovation Center and provides perspectives from the research community on critical issues and challenges.

★ ★ ★ ★ ★

The mission of The Commonwealth Fund is to promote a high performance health care system. The Fund carries out this mandate by supporting independent research on health care issues and making grants to improve health care practice and policy. Support for this research was provided by The Commonwealth Fund. The views presented here are those of the authors and not necessarily those of The Commonwealth Fund or its directors, officers, or staff.

For more information about this study, please contact:

Marsha Gold, Sc.D.
Senior Fellow
Mathematica Policy Research
MGold@Mathematica-Mpr.com

To learn more about new publications when they become available, visit the Fund's Web site and register to receive email alerts.

Commonwealth Fund pub. 1512
Vol. 12

OVERVIEW

The Center for Medicare and Medicaid Innovation (Innovation Center), authorized in Section 3021 of the Affordable Care Act and located in the Center for Medicare and Medicaid Services (CMS), seeks to promote innovation in health care payment and delivery.[1] It has a legislated mandate:

> to test innovative payment and service delivery models to reduce program expenditures…while preserving or enhancing the quality of care furnished to individuals…(under Medicare, Medicaid, and the Children's Health Insurance Program). In selecting such models, the Secretary shall give preference to models that also improve coordination, quality, and efficiency of health care services furnished…
>
> The Secretary shall select models to be tested…where the Secretary determines that there is evidence that the model addresses a defined population for which there are deficits in care leading to poor

clinical outcomes or potentially avoidable expenditures.[2]

To support the Innovation Center's goals, the legislation provides $10 billion in funding from 2011 to 2019 and enhanced authority to waive budget neutrality for testing new initiatives.[3] The intent is to allow quicker and more effective identification and spread of desirable innovations, with the goal of ultimately modifying Medicare, Medicaid, and the Children's Health Insurance Program in ways that support program-wide change.

Though the strategic focus for the Innovation Center is still under development, there have been clear signals that its focus will be broad, with an emphasis on transformative change to address the "triple aim" of improving the quality of care, reducing cost growth, and enhancing population health.[4,5]

Achieving this will be challenging and the time frame demanding in the face of historical experience in which years elapse between the origination of an idea and the process of designing, implementing, and evaluating.

This issue brief focuses on three critical requirements the Innovation Center must address to meet its objectives:

1. Focusing on change that matters;
2. Documenting innovation to support effective learning and spread;
3. Generating the evidence needed to support broad-based policy change.

Tensions between competing goals can be reduced by anticipating them and thoughtfully designing the way innovations are tested and evaluated in the Innovation Center. Different trade-offs may be appropriate for innovations at different stages or with different potential risks and rewards. Collaboration among researchers, innovators, and policymakers about how best to address different goals and potential tensions is needed to enhance the innovation center's overall prospects for success.

INTRODUCTION

The Center for Medicare and Medicaid Innovation (Innovation Center), as authorized by the Affordable Care Act and located in the Center for Medicare and Medicaid Services (CMS), seeks to promote innovation in health care payment and delivery. To support the Innovation Center's goals, the legislation provides $10 billion in funding from 2011 to 2019 and enhanced authority to waive budget neutrality for testing new initiatives. The intent is to allow quicker and more effective identification and spread of desirable innovations, with the goal of ultimately modifying Medicare, Medicaid, and the Children's Health Insurance Program (CHIP) in ways that support program-wide change.

Though the strategic focus for the Innovation Center is still under development, there have been clear signals that its focus will be broad, with an emphasis on transformative change to address the "triple aim" of improving the quality of care, reducing cost growth, and enhancing population health.

FOCUSING ON CHANGE THAT MATTERS

The Affordable Care Act provides $10 billion in funding to support the Innovation Center's goals. This is a substantial amount, but it is less than 0.1 percent of projected federal Medicare and Medicaid spending through the end of this decade and a much smaller proportion of the projected $32 trillion in total health spending over the same period.[6] This small percentage stands in contrast with the much higher proportion of industry revenues devoted to research and development in the pharmaceutical industry and in other industries in which innovation is a central focus, like technology and communications.[7] Neither this level of funding nor available staff is likely to be sufficient to invest in all the innovations that might be considered, so priorities must be set. Priority setting is a policy rather than a research decision, but research can help lead to better decision-making. Input from the research community is therefore an important element from the beginning of the innovation process.

To achieve its statutory goals, the Innovation Center must identify as priorities those innovations that have the potential to achieve demonstrably large positive impact on quality and costs, as measured by a combination of improved outcomes and reduced costs. Innovations can be successful either by generating large gains over a relatively small population or smaller gains over a large one. The relative merits of the gains that may be achieved by different strategies vary and may depend on their administrative costs and whether they are fixed or vary with the size of the population. In any case, the research community can contribute to the determination of the potential net gains from alternative pilots, as well as to the development of measures that can be used to monitor and assess the performance of those pilots.

DOCUMENTING AND LEARNING FROM INNOVATION

The Innovation Center's success depends not just on developing and implementing innovations but on the ability to monitor and evaluate innovations to provide evidence of their success and information to encourage widespread adoption. This is unlikely to occur without clear articulation of the essential logic of an innovation, how it is intended to operate, and—perhaps most important—the results it is expected to produce and how success can be recognized. It is important to document the context in which an innovation was tested and assess how important that is to its success or applicability elsewhere so that those who may be considering it have an explicit understanding of the potential gains and associated costs they may experience.

Careful Planning and Clear Definition of Success

An innovation's goals must be expressed in concrete, measureable terms that are linked to a time frame that provides a basis for monitoring performance and determining success. Essential elements of success include an explicit understanding of the activities needed to generate the anticipated outcomes; how the activities are logically connected to outcomes; the environment and context; and any potential obstacles and how they will be addressed.

Unless innovations are well defined and their connections to desired outcomes are well understood from the start, it will be difficult to achieve success. Even if positive outcomes are achieved, it will be difficult to assess the relevance of those results to other settings and to replicate them throughout the health system. Unfortunately, such clarity is often lacking or limited, with critical design elements and site-specific characteristics unstated, key details driving success or failure potentially omitted or unrecognized, and the likelihood of success low because interventions are insufficient in scope or scale to achieve their intended effects.[8] These limitations frequently can be traced to the lack of necessary data systems and measures and the need for methods that can produce more flexible and timely, accurate analysis.

Tracking Implementation and Performance

Innovations rarely remain fixed over time. Key features are likely to be modified as experience grows or problems emerge. Time frames may depart from those anticipated. Objective short-term indicators of implementation success provide a basis to judge whether midcourse refinements may be valuable. Documenting what actually was implemented versus what was initially sought is critical for interpreting the lessons from testing and providing the basis for future spread. Case studies of implementation experience also can be invaluable to others that may seek to replicate or build upon what was learned from a given experience.

Supporting Timely Measurement

The success of rapid-cycle change depends on measurement—capturing and feeding back timely data on change after the launch of an innovation that allows fine tuning of the project, early insights on additional questions for analysis, and ongoing communication and the potential to learn from failure and success. Prior demonstrations highlight the challenges in securing timely data. For example, in the Medicare Physician Group Practice demonstration, financial results were

not available for almost a year after the initial performance year; data to inform quality bonuses took even longer (Exhibit 1).[9]

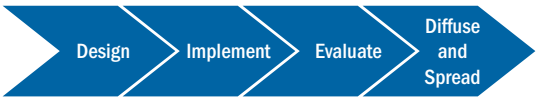
Making use of data generated naturally in the course of administering an innovation on a real-time basis can lessen delays. For example, sites often will have real-time information on use of services and hospital admissions, registries that may document who was eligible for an innovation or served by it, and data on patient feedback. Sponsors of Vermont's all-payer medical home demonstration say their ability to leverage existing administrative processes to capture data was critical to reducing providers' costs of participation and enhancing the timeliness of information feedback.[10] Effective use of such data is likely to require advanced planning. In a different effort, evaluators provided sites with a workbook tool for generating measures, including definitions of numerators, denominators, and the included population.[11] An alternate strategy that can enhance data quality and consistency is to work with payers and providers to aggregate data they receive in a centralized fashion and feed it back to providers in a consistent and timely way so that they can monitor and manage what they are learning from their efforts at innovation.

However data flow occurs, the process for data exchange and the format and content of reports should be decided up front and structured so data are useful for providers. Analyzing the data before implementation also can help with setting benchmarks and intermittent milestones.

Investing in Shared Metrics and Documentation

Developing the capacity to assess innovations also requires a concerted effort to develop metrics and documentation.[12] To facilitate this process, CMS should identify common variables that are needed across all sites testing specific kinds of innovations and standard metrics that will facilitate aggregation and comparison of performance across sites. This includes outcome metrics relevant to all innovations and critical data that identifies design elements included in particular innovations, the settings in which they are employed, and other variables relevant to their success. Evaluations that include structured study of implementation typically address such concerns, but they have not historically included the kind of timely feedback that the Innovation Center likely will require.

Exhibit 1. Time Line for the Medicare Physician Group Practice Demonstration



Design: Obtain authority from Congress, develop design, select sites, get Office of Management and Budget approval for waiver (2000–2005)

Implement: Implement program and collect operational data (2005–2010)

Evaluate: Evaluate initial effects (2006), two-year results (2009), and final outcomes (2011–2012 and beyond)

Diffuse and Spread: Disseminate effective payment practices nationwide (2011–2012 and beyond)

Source: Adapted from a presentation by Mark McClellan at a Roundtable on Methods for Identifying, Designing, Monitoring, and Evaluating Innovations, Washington, D.C., Nov. 17, 2010.

Targeting Learning to Achieve Stakeholder Buy-In

If the lessons of an innovation speak to the interests of diverse users and stakeholders, widespread adoption is more likely. Successful replication of innovations will require addressing the concerns of critical participants. For example, providers will want to understand the operational demands of any innovation, how their revenue streams will be affected, and whether change will help or hinder them in achieving institutional goals. To complement the analysis of results, case studies from objective researchers can provide important insight into the key factors contributing to successes and challenges from the perspectives of multiple stakeholders.

Setting Realistic Expectations

Implementation almost always takes longer than expected, with modifications occurring along the way. The larger the scope of an innovation, the greater the complexity of the organization, and the more units or organizations involved, the more time is likely required for ramp-up. Personnel must be recruited or trained, approvals obtained, and participants defined and recruited. Delays may occur because of personnel

change (e.g., loss of the champion or key source of leadership and support), competing organizational priorities that limit access to resources, or new issues that require design modifications. If Medicaid or CHIP is involved, state policymaker buy-in and approvals may be required and time may be lost reconciling different concerns that may exist in cross-state demonstrations.

Although careful planning prior to implementation is always required, pilots and demonstrations are not conducted in controlled environments and the implementation process must allow for adjustment to contingencies as they arise. This requires flexibility on the part of the entities and individuals directly involved in the innovation but also on the part of researchers responsible for evaluating initiatives and policymakers who will be acting on the results.

Emphasizing Clarity of Objectives and Timeliness of Implementation

Organizations are more likely to be able to implement innovations that are clear and simple. Successful innovation can be enhanced by avoiding unnecessarily complex elements or requirements and by limiting standardized features to those most essential to success and common analysis of cross-site activity. In any case, the objectives must be set in a way that all stakeholders understand and agree.

Timeliness in the implementation process is also important. Momentum can be critical to organizational success; once organizations are poised for action, delays can be very damaging to underlying stakeholder support. Delays can be minimized by streamlining processes between the announcement of an initiative and its implementation and by developing common procedures and approaches that work across a variety of innovations.

GENERATING THE EVIDENCE NEEDED TO SUPPORT BROAD-BASED POLICY CHANGE

The Affordable Care Act enhances the authority of the Secretary of Health and Human Services to modify payment and selected program policies for the pilots being conducted by the Innovation Center.[13] However, the ultimate goal is to encourage better ways of financing and providing care throughout the health system, many of which are expected to require a shift away from the current fee-for-service payment methods under which providers are paid now, not only by Medicare and Medicaid, but also by other payers.[14] While the Secretary has the authority to make changes in Medicare without going back to Congress, she must be convinced change is warranted by its demonstrated potential to improve quality and the CMS actuary must be willing to certify that, at a minimum, it will not add to program costs. Considering the level of concern about the federal budget, costs are likely to be a major focus and generating definitive evidence of the effects on program costs is likely to be a particular challenge.

One key question to consider is what standard of evidence is likely to be required to support such decision-making and how evaluations should be structured to generate it. This issue is critical to the design and conduct of effective evaluations, and it will be an important factor in the Innovation Center's ability to carry out its mission.

Historical Context

Historically, the effectiveness of an intervention has been assessed using relatively rigorous research methods that evaluate the actual (versus intended) effects of demonstrated program change on desired outcomes, such as the triple aims of better health, better care, and lower costs. This typically has involved independent evaluation by contracted researchers employing several basic elements, including:

- Careful definition of the target population and how it is to be assessed for purposes of judging success;
- One or more comparison populations or control groups to serve as a benchmark for indicating what might have happened in the absence of the change;
- Metrics defining the outcomes of interest and how they change over time, which often require new forms of data collection or unique data files

developed from existing claims or other program data; and

- Long time frames designed to distinguish immediate effects from more stable, longer-term effects. Five-year time frames have been common, though some initiatives have been assessed more rapidly.

The evaluation designs seek to distinguish true effects of an innovation from those that can be explained by other factors like secular trends, changes in patient mix, or other contextual change. In other words, they try to isolate the impact of the innovation compared with what would have been expected to occur in its absence.

The size of the target population and the control group is an important design factor. Large populations are helpful in developing statistically valid estimates of effectiveness and in distinguishing subpopulations most likely to benefit from the innovation. The involvement of large populations, however, typically adds to the cost of an evaluation.

Timeliness is another important factor. The design, development, implementation, and evaluation of an innovation can be a lengthy process. Exhibit 2 illustrates the time line for Medicare's Physician Group Practice demonstration, which was a model for the Medicare Shared Savings Program created in the Affordable Care Act. The more than 10-year time line is not dissimilar from the experiences of other demonstrations. The time line can be shortened by developing clear goals for new pilots and explicit criteria for participating and streamlining the decision-making process, and establishing standardized metrics for monitoring performance from data already available from claims and other sources. In addition, the methodology for identifying promising initiatives, monitoring performance, and evaluating results should be examined for its ability to meet the needs of a process intended to produce rapid change.

The Need for Timeliness and Rigor

The legislation establishing the Innovation Center seeks to accomplish rapid-cycle change in health care delivery. This will require the ability to shorten the time needed to identify, develop, and assess innovations with sufficient rigor to provide definitive evidence that they can improve the quality of care while reducing costs. Such expectations will require some modification in the process that traditionally has been used to develop demonstrations and new methodological approaches for assessing the performance of health care delivery systems and policies.

Planning and Coordination

The Innovation Center gives CMS great flexibility to test potential policy changes. Effective use of such authority will require streamlining the process for developing and implementing pilot projects. It is important not to cut corners and take shortcuts that would threaten the validity of the process and to lay out a clear and consistent approach that can be accomplished with a minimum of unnecessary delay.

Advanced planning is particularly important in this context. The establishment of clear and consistent goals for each initiative and a transparent and coordinated mechanism for approving potential pilots can not only reduce the time needed to assess effectiveness but also help ensure they will, in fact, be effective. A key factor is the ability to provide an infrastructure for supporting new initiatives, so that the data needed for CMS to monitor the performance of pilots and for the pilot participants to manage the initiatives and gauge their own performance are available on a timely basis and in a useful format.

Better coordination among the key stakeholders in the process—both within and outside government—is also important. Many parties are involved in developing the innovative strategies that the Innovation Center will test, and the approval of the Office of Management and Budget, which often has been difficult to obtain, will still be necessary to conduct the pilots. Implementation will involve CMS and the participating sites, but also—in multipayer

initiatives—may include Medicare, Medicaid, private payers, and other stakeholders in the communities and at the national level. Evaluation of pilots will involve the Innovation Center and other CMS components, as well as the Office of the Secretary, which will be responsible for attesting to the quality of Innovation Center pilots, and the Office of the Actuary, which will be responsible for certifying cost-saving potential.

Assessing Performance

With standardized evaluation procedures and better data, the time frame for evaluation could be shortened. This can be accomplished without cutting corners or sacrificing rigor, but by relying on an ongoing stream of information to monitor projects and make midcourse corrections as well as reaching definitive conclusions about effectiveness.

Making assessments over a shorter time period inevitably raises questions: will the effects observed over a shorter period be borne out over the long term? The assessment of pilots implemented by the Innovation Center must take this into account and balance the desire to have results quickly with the need to have an accurate picture of how these pilots work and the results they are likely to produce over time.

Unless a change is very dramatic, its effects may not be immediate, so early assessments can result in discarding potentially promising innovations that would be proven effective if given more time. Conversely, some innovations may appear successful initially but the effects may be short-lived or offset by gaming or unintended results that are not apparent until more time has elapsed. Different outcome measures also have inherently different time frames. Policymakers must consider these risks when applying

Exhibit 2. Illustrative Time Line: Medicare Physician Group Practice Demonstration

Required Activity by Phase	Relevance to Innovation Center Context
PRE-IMPLEMENTATION	
Congressional Mandate (2000)	The Innovation Center can proceed without explicit congressional approval but will still need to define priorities.
Design (2001 to 2003)	Aim is to encourage "bottom-up" planning but the Innovation Center will still need to decide which features to focus on within each priority area and how to structure metrics and criteria for success.
Site Selection (August 2003)	The Innovation Center will need to establish standards for participation in pilots.
Waiver Approved (October 2004)	No new waiver is required, but there will be an internal process for approving flexible delivery and payment policies.
IMPLEMENTATION	
Official start to demonstration (April 2005)	
Five-year demonstration ends March 2010	Less time is required for effects that are expected sooner. Rapid feedback can give indications of whether the innovation appears to be working as planned or needs fine tuning. Even shorter term outcomes, however, are likely to require continued monitoring, particularly if short- and long-term effects differ in important ways.
EVALUATION	
Report of first-year results but no quality or expenditure data available to include (2006) Report available with data on first two years of the demonstration covering April 2005–March 2007 (2009) Final evaluation (expected 2011/2012)	Evaluations can build in early feedback loops and timely designs that support midcourse corrections and generate lessons for refinement and spread. Doing so requires mechanisms for collecting, processing, analyzing, and distributing data that are not currently in place. Appropriate balance needed between short- and long-term evaluation of progress and performance.

Note: Comments regarding how this experience would apply to the new CMMI mandate are the authors' alone.
Source: Presentation by Mark McClellan at Roundtable on Methods for Identifying, Designing, Monitoring, and Evaluating Innovations, Washington, DC, November 17, 2010.

the experiences of the Innovation Center nationally through changes in Medicare and Medicaid policy and whether the potential gains from adopting a fast-tracked policy change outweigh the downside risk of adjusting the policy should subsequent longer-term evaluation warrant changes. The contributions of the health services research community will be extremely valuable in this area.

Standards of Evidence and Their Related Risks

The standard-of-evidence issue involves making judgments on the trade-offs of risks from different types of errors in interpreting pilot outcomes. There is a risk of judging change (e.g., an intervention or innovation) to be effective when it is not. This type of error has obviously adverse implications: it can lead to the propagation of a model of payment and care that has no advantages relative to the current system or is perhaps worse. Another type of error creates the opposite result: rejecting an innovation as a failure when it is actually effective. This type of error can be very harmful as well because it delays or obstructs the implementation of effective initiatives that can improve the current system.

Historically, most evaluations have been designed with the goal of limiting the risk of the first type of error. Some criticize this approach as overly conservative and insensitive to the second type of error, particularly when the objective is to find effective alternatives to the current system. There are risks from both types: moving too slow to encourage effective innovation or too fast to institutionalize innovation that may falsely believed to be effective. The appropriate way to balance the two approaches varies with the context and the potential impact of each type of error. Changes that have greater potential to harm patients or add significantly to program costs must be guarded against. Where the gain-to-risk ratio is more favorable, an approach that leans toward proceeding with new approaches may be warranted, with policy fine-tuned as additional information is generated.

Evidence and Policy Change

Since the Affordable Care Act gives the Secretary (working with the CMS actuary) authority over decisions that previously were the responsibility of Congress, it is useful to review the standards of evidence Congress historically has applied to authorize a change in program policy.

A review of Medicare history shows that Congress often has enhanced important policy changes without solid evidence to support such changes. For instance, Medicare competition demonstrations were still being evaluated when program-wide authority for the Medicare risk contracting program was enacted in 1982.[15] Congress enacted the Medicare hospital prospective payment system and changed national Medicare policy on hospital payment, citing New Jersey's existing work with diagnosis-related group–based payment to support the feasibility of change. But the details of the New Jersey system tested varied substantially from the Medicare model that was put in place so the national change, in fact, was based on relatively limited testing.[16]

Other changes, such as the introduction of a resource-based relative value schedule for physician payment, were not tested as much as built on research to define key parameters of the payment model and expert vetting involving a range of stakeholders. Some evaluations that have shown positive results (such as competitive bidding for durable medical equipment) have never been implemented globally because of organized opposition.[17] In certain cases evaluations have proven negative—as with Medicare's cost contracts that were found to increase program costs—but the programs have been retained because they serve other valued objectives.[18] Important changes in Medicare, like the authority for accountable care organizations, were enacted with relatively limited empirical support.[19]

This history argues against applying standards of evidence that are so technically rigorous that they impede real progress in improving the performance of the health system, which requires change on many dimensions.[20] At the same time, clear and technically

defensible standards of evidence to support major changes in program policy can serve CMS well in its mission to reform the payment and delivery systems. Standards provide a way of navigating politically contentious debates over change and provide the guidance necessary to appropriately target limited resources.

CONCLUSIONS AND IMPLICATIONS

Timely evaluation that is targeted to important concerns can help identify the kinds of innovations likely to make a big difference and support policymakers to better structure the way they test innovations to enhance the ability to learn from such testing. Evaluation also can help answer the questions anticipated to arise in applying the lessons from testing to support program-wide policy change that will institutionalize incentives to improve health care delivery and value.

It is important to keep in mind three conclusions from assessments of past experiences dealing with evaluating finance and organizational changes. First, implementation itself is important. The evidence that alternative policies can be adopted and are feasible can be a powerful lever for change. The Innovation Center appears well-suited to developing such evidence, by building systems that efficiently document the feasibility of innovation in forms that can be shared.

Second, the quality of evidence likely to be generated by testing innovations will vary. Testing and evaluation practices likely to encourage high-quality evidence include: 1) clearly articulated models developed to assess program logic, including feasibility and plausibility; 2) ongoing measurement that provides information on relevant intended and unintended outcomes associated with the innovation; 3) appropriate analysis that reinforces confidence that change can be legitimately attributed to the innovation rather than other causes; and 4) information on context and implementation experience to help others determine whether the innovation is likely to be appropriate in their setting and how to proceed.

Third, there are inherent trade-offs involving flexibility, timeliness, and the ability to generate rigorous evidence that will enhance the confidence policymakers have about the effects of policy change. There is an important distinction between rigor and rigor mortis. Methodological rigor is extremely important in distinguishing initiatives that are useful and can be propagated throughout the health system to good ends from those that "wish only to preserve the status quo."[21] But decisions about methodological rigor must not stifle all attempts to improve the health system on the grounds that no data are good enough and no risk is worth taking. Risks—albeit informed risks—must be taken to improve the health system and avoid the ever-intensifying pressure, not only on federal and state governments but also on businesses and households, as a result of increasing health spending without concomitant improvements in quality and outcomes.

Tensions between competing goals can be reduced by anticipating them and thoughtfully designing the way innovations are tested and evaluated in the Innovation Center. Different trade-offs may be appropriate for innovations at different stages or with different potential risks and rewards. Collaboration among researchers, innovators, and policymakers about how best to address different goals and potential tensions is needed to enhance the Innovation Center's overall prospects for success.

NOTES

[1] See page 306+ of the bill posted at http://docs.house.gov/energycommerce/ppacacon.pdf.

[2] The legislation provides 20 specific examples for consideration though the list is not meant to be restrictive. See Exhibit 2 in S. Guterman, K. Davis, K. Stremikis, and H. Drake "Innovation in Medicare and Medicaid Will Be Central to Health Reform's Success," *Health Affairs,* June 2010 29(6):1188–93.

[3] Guterman, Davis, Stremikis et al., "Innovation in Medicare," 2010.

[4] CMS, "The New Center for Medicare and Medicaid Innovation" Press release November 16, 2010, Available at http://innovations.cms.gov/news/press-releases/pr110910.shtml, accessed January 5, 2010.

[5] D. M. Berwick, "Introduction to the CMS Center for Medicare and Medicaid Innovation—and innovations.cms.gov," Nov. 26, 2010, available at http://innovations.cms.gov/blog/introducing.shtml.

[6] Centers for Medicare and Medicaid Services, "National Health Expenditure Projections 2009–2019," Sept. 2010, available at http://www.cms.gov/NationalHealthExpendData/downloads/NHEProjections2009to2019.pdf.

[7] Congressional Budget Office. *Research and Development in the Pharmaceutical Industry* (Washington, D.C.: CBO, Oct. 2006).

[8] See, for example, J. E. Mahoney, "Why Multifactorial Fall-Prevention Interventions May Not Work: Comment on 'Multifactorial Intervention to Reduce Falls in Older People at High Risk of Recurrent Falls,'" *Archives of Internal Medicine,* July 12, 2010 170(13):1117–19. Mahoney emphasizes the relevance of definition content, process, and targeting as part of evaluating interventions that aim to improve patient outcomes.

[9] Presentation by Gregory Pope, RTI, at the Roundtable. The delay required six months for a claim to be 98 percent complete, an additional one to two months for data to reach the CMS data center, three months for claims acquisition and analysis, and one to two months for CMS review.

[10] Comments by Jim Hester as part of the Roundtable.

[11] D. Esposito, E. F. Taylor, and M. Gold, "Using Qualitative and Quantitative Methods to Evaluate Small-Scale Disease Management Pilot Programs," *Population Health Management,* Feb. 2009 12(1):9–20.

[12] D. M. Berwick, "Developing and Testing Change in Delivery of Care," *Annals of Internal Medicine,* April 15, 1998 128(8):651–56.

[13] The legislation also removed a requirement that innovations tested in the Center be assessed as budget-neutral prior to their testing. A. Cassidy, "The Fundamentals of Medicare Demonstrations," Background paper no. 63 (Washington D.C.: National Health Policy Forum, July 22, 2008); and Guterman, Davis, Stremikis et al., "Innovation in Medicare," 2010.

[14] See S. Guterman and S. C. Schoenbaum, "Getting from Here to There in Payment Reform: Necessary Practices and Policies," *Journal of Ambulatory Care Management,* Jan./March 2010 33(1):52–57.

[15] Communication with Randall Brown, Mathematica Policy Research, Feb. 2010.

[16] M. Gold, K. Chu, S. Felt et al., "Effects of Selected Cost-Containment Efforts: 1971–1993," *Health Care Financing Review,* Spring 1993 14(3):183–225.

[17] M. Likes, "The Battle Continues over Medicare Competitive Bidding Program," *HealthWatch,* July 2, 2010, available at http://thehill.com/blogs/healthwatch/medicare/107001-battle-continues-over-medicare-competitive-bidding-program.

[18] M. Sing, R. S. Brown, and S. C. Hill, "The Consequences of Paying Medicare Managed Care Plans Their Costs," *Inquiry,* Summer 1998 35(2):210–22.

[19] J. K. Iglehart, "Assessing an ACO Prototype: Medicare's Physician Group Practice Demonstration," *New England Journal of Medicine,* Jan. 20, 2011 364(3):198–200.

[20] See K. Davis, C. Schoen, and K. Stremikis, *Mirror, Mirror on the Wall: How the Performance of the U.S. Health Care System Compares Internationally, 2010 Update* (New York: The Commonwealth Fund, June 2010).

[21] D. M. Berwick, Administrator CMS Remarks at "Workshop on Issues Related to Accountable Care Organizations" Federal Trade Commission, CMS and HHS Office of the Inspector General, Baltimore MD October 5, 2010. http://www.ftc.gov/opp/workshops/aco/index.shtml. Berwick added that "authenticity matters" and "we just do not have time for games anymore."

About the Authors

Marsha Gold, Sc.D., is a senior fellow at Mathematica Policy Research. Dr. Gold has run several projects monitoring the implementation of Medicare managed care. In addition to Medicare Advantage, Dr. Gold's research focuses on managed care markets more generally and on Medicaid managed care and health care access. Over her career, Dr. Gold has been particularly interested in ways to make research findings and evaluation techniques relevant to and accessible to policymakers and other important audiences. Dr. Gold earned her doctorate from the Harvard School of Public Health.

David Helms, Ph.D., is the director of LMI's Center for Health Reform. LMI is a nonprofit strategic consulting firm in Washington, D.C. Until 2011, Dr. Helms was president of AcademyHealth, the national society for the fields of health services research and health policy analysis. Dr. Helms' career has focused on advancing health care improvements through planning, program analysis, and development of new initiatives to expand coverage for the uninsured. Through his leadership of the Robert Wood Johnson Foundation's demonstration programs to expand health coverage, he has become a leading national expert on health reform options. Dr. Helms received his Ph.D. in public administration and economics from the Maxwell School of Citizenship and Public Affairs at Syracuse University.

Stuart Guterman is vice president for Payment and System Reform at The Commonwealth Fund and executive director of the Commonwealth Fund Commission on a High Performance Health System. His prior positions include: director of the Office of Research, Development, and Information at the Centers for Medicare and Medicaid Services; senior analyst at the Congressional Budget Office; principal research associate in the Health Policy Center at the Urban Institute; and deputy director of the Medicare Payment Advisory Commission (and its predecessor, the Prospective Payment Assessment Commission); and chief of institutional studies in the Health Care Financing Administration's Office of Research. He can be e-mailed at sxg@cmwf.org.

Acknowlegments

This paper was developed as part of a project funded by The Commonwealth Fund to AcademyHealth, with a subcontract to Mathematica Policy Research. It builds on the discussion at a roundtable of experienced evaluators and users of evaluations that met on November 17, 2010. The authors are particularly indebted to Mark McClellan, Gregory Pope, Randall Brown, and Robert Berenson for their presentations at the roundtable, but the discussion in this paper represents the thoughts of the authors alone and does not necessarily reflect the views of their affiliated organizations, other presenters or participants in the roundtable, or their affiliated organizations.

Editorial support was provided by Deborah Lorber.


THE
COMMONWEALTH
FUND