

Expert Panel Meeting on Disease Management Outcomes Measurement

Summary Report

Contract Number HHS-100-03-0019
Task Order 07EASPE000099

Prepared for:
The Office of the Assistant Secretary for Planning and Evaluation
U.S. Department of Health and Human Services

Prepared by:
RAND Corporation

August 2008

Executive Summary

Disease management (DM) refers to a system of coordinated health care interventions and communications to help patients address chronic disease and other health conditions. It seems an intuitively plausible approach for addressing rising healthcare costs and the need for improved quality of care in the U.S. However, in spite of the rapidly growing penetration of DM in the commercially insured market, there is no conclusive evidence that DM reduces cost, and limited evidence that it improves quality of care.

To clarify the issues surrounding the implementation and evaluation of DM interventions for public sector programs—Medicare and Medicaid—the Office of the Assistant Secretary for Planning and Evaluation (ASPE) commissioned an expert panel meeting on January 16, 2008. The objective of the meeting was to convene subject matter experts to discuss measurement of the impact of DM on health outcomes and costs of care and implications for integrating DM in public sector programs. A primary goal of the meeting was to consider the variety of DM programs and identify strategies to improve the assessment of their impact on individuals with multiple chronic conditions. Nine experts representing the private sector, public sector, and academia agreed to participate in the Expert Panel (Exhibit A). This report summarizes the proceedings of the Expert Panel meeting.

Exhibit A:

Expert Panel Members

David Atkins, MD	Chief Medical Officer, Center for Outcomes and Evidence, Agency for Healthcare Research and Quality (AHRQ)
Melanie Bella, MBA	Senior Vice President for Policy and Operations, Center for Health Care Strategies (CHCS)
Ahmed Calvo, MD, MPH	Director and Chief Medical Officer, HRSA Health Disparities Collaboratives, Center for Quality, Office of the Administrator Health Resources and Services Administration (HRSA)
Jeffrey Lemieux	Senior Vice President, Center for Health Policy and Research America's Health Insurance Plans (AHIP)
Ariel Linden, DrPH	President, Linden Consulting Group
Linda Magno, MPA	Director, Medicare Demonstrations Program Group, Centers for Medicare & Medicaid Services (CMS)
Nancy McCall, ScD	Chief Scientist, RTI International
Gordon Norman, MD, MBA	Executive Vice President and Chief Science Officer, Alere Medical, Inc
Paul Wallace, MD	Medical Director for Health and Productivity Management Programs; Senior Advisor, The Care Management Institute and KP-Healthy Solutions The Permanente Federation, Kaiser Permanente

Evolution of DM

DM traces its origins back to the 1970s with the concept of prospective medicine. The movement progressed in the 1980s as diabetic patients were encouraged to engage in self-care activities. In the 1990s, the term “disease management” was coined to describe a strategy to increase pharmaceutical sales by improving medication adherence. DM programs are rapidly growing in popularity in the commercially insured market. At the same time, they are being adopted in many Medicaid programs and tested under pilot and demonstration projects for certain Medicare populations.

DM has evolved into a complex and diverse industry that offers comprehensive management of chronic conditions. The focus has shifted from single-disease programs to an integrated approach that targets all of a patient’s health problems. Programs have also expanded from targeting patients with very high cost and risk to addressing the needs of broader populations, with approaches ranging from mass communication technology and call-center based outreach to more intense approaches such as home visits. DM is being embedded in patients’ primary care sites, as envisioned by the Chronic Care Model; becoming a service provided by integrated delivery systems that combine insurance and provision of care; and being offered by external providers that can be free-standing companies or parts of health plans.

The increasing uptake of DM intersects with many of today’s fundamental issues in health policy, such as

- the patient/provider relationship;
- cultural changes that are needed to evolve from medical-centered to patient-centered care;
- the role of health information technology (HIT) and how to leverage and foster its use; and
- how best to align incentives and organizational fabric to ensure that resources are used efficiently and effectively.

A particular challenge is the integration of DM with the patient’s usual site of care, as cultural barriers and communication problems may interfere with leveraging the advantages of DM.

The Challenge of Measuring Program Impact

Determining which DM program provides value under which conditions requires 1) defining “value”, as the impact of DM programs can be measured along numerous dimensions, and 2) identifying the specific measures and metrics that will be used to capture performance along those domains. The measurement strategy must reflect the intended use of the information and resource implications of data collection and report preparation and must provide sound and actionable information. However, in practice, experience has shown that measures selection is driven by negotiations about performance targets rather than a desire to obtain a multi-faceted view of DM impact.

While measurement science has evolved, it continues to trail the growing complexity of the programs. Measurement approaches continue to be disease focused rather than patient-centric. Bridging this gap requires measures that capture how well a program has met a patient’s need, rather than standardized measures. For example, rather than measuring only whether each patient with diabetes has received an HbA1c test, a more appropriate strategy may employ measures that reflect whether the five most important health issues for the individual patient have been addressed.

For public sector programs where many patients have multiple co-existing conditions and complex non-medical needs, disease-focused measurement is a particular challenge. The lack of uniform measures (and their definitions) across entities presents obstacles for measurement. Improved alignment would facilitate data collection, because providers would not have to follow different data collection protocols for different entities, and interpretation of results would be consistent.

The Challenge of Attribution

Determining the impact of a DM program requires attributing changes in the selected measures to the intervention itself, rather than to other changes such as secular trends or market-level changes. This need implies making a prediction of what would have happened in the absence of the program. Randomized controlled trials are the most precise method to achieve this end, but they are rarely feasible in operational settings, because they are expensive, time consuming and suffer from threats to generalizability. Attribution approaches are needed that minimize common threats to validity, such as selection bias and regression to the mean, but provide a reasonable balance between scientific rigor and feasibility.

A consensus seems to be emerging in principle between researchers, who advocated the most rigorous and demanding evaluation approaches, and DM program operators, who argued for practical and efficient approaches, that an equivalent comparison group is needed to ensure proper attribution, even though there is still substantial disagreement over what equivalency means. The Disease Management Association of America (DMAA) has proposed voluntary industry guidelines, and the research community has argued for the use of rigorous non-experimental methods. While it is not clear yet which approaches will be universally adopted for which purposes, discussion of the different approaches and more research into their relative performance are needed. The current heterogeneity of evaluation methods stands in the way of comparing programs and learning which approaches work best under which conditions.

The debate about evaluation methods has also fueled a broader discussion about the value of DM programs beyond the narrow target of cost reduction. Different stakeholders attach different importance to different domains of value, and a holistic evaluation approach needs to be able to express such a multi-faceted conception of value. This broader definition of value may not be as easy to communicate as return on investment (ROI) but will allow for a richer assessment of the impact of DM.

Characteristics of Public Sector Programs

Most of the experience with DM has been derived from a commercially insured population, a fact that has shaped the development and refinement of programs. It is likely that the particular characteristics of the Medicare and Medicaid populations will affect how well DM will work and that both operations and content of the DM programs will have to be retooled. The medical needs of patients in Medicare and Medicaid programs are considerably more involved and variable than those of commercially insured patients. Patients in Medicare and Medicaid programs commonly receive treatment for multiple chronic diseases, and general frailty and end-of-life issues may limit treatment options and determine priorities.

Socioeconomic issues like health literacy and economic concerns and limitations in access to care may also interfere with the operations of DM programs in public sector populations. A

specific challenge for Medicaid programs is that patients may move in and out of eligibility, which poses a problem for program effect as well as evaluation, as it becomes difficult to define the point at which a patient can be considered engaged in a program.

These particular characteristics imply that public sector DM programs will have to be adapted for those populations and will likely be more intense and costly. Nevertheless, the greater burden of disease in those populations means that the potential financial return could be large, especially if one takes the long run perspective of sustaining Medicare's finances; the potential benefit could be considerable as well. Realizing those opportunities will require creatively addressing challenges of implementing and operating DM programs for those populations and more research into what is working, for whom, and at what costs.

The Potential of DM for Public Sector Programs

Results from Medicare demonstrations and pilot programs have so far largely failed to demonstrate cost savings or improved outcomes, and some realism may be necessary in terms of setting expectations for these programs. Many state Medicaid programs have adopted DM interventions, either as demonstrations or as routine operations, typically driven by the goal of containing cost. Unfortunately, few programs have undergone rigorous evaluations, as states commonly lack the resources to conduct them. A particular challenge is the lack of a platform that would allow states to pool their data and conduct cross-program research with the goal to learn from each other and identify best practices.

The limited evidence for success of DM in Medicare and Medicaid programs thus far challenges the assumption that DM has a positive impact on cost and quality of care. This finding should not lead to the conclusion that chronic care cannot be improved, but should stimulate the development and implementation of innovative approaches to improve care in those complex populations. To be successful, those approaches will have to be patient-centric, multi-dimensional, and coordinated with the patients' usual source of care.

Conclusions

Early enthusiasm about the promise of DM has given way to skepticism and a growing realization that better evidence is needed to identify effective and efficient programs that adequately address the complex medical and non-medical needs of patients in public sector programs. Research must get beyond the question of whether DM works and begin to address richer and more contextual questions that explore the value of DM and how its value can be measured and expressed. With better information, an approach might emerge that encompasses the patient, is multifaceted and intelligent, and has rich interventions.

To design such interventions, more research is needed that is rigorous yet feasible. Although randomized control trials may not be the solution, we can continue to learn from doing more observational research, more post hoc analyses, and more integration of data by marrying claims data with care management and patient self-reported data in order to get a richer picture of what might work, for which population, at what cost.

There is still a long road ahead of us to determine the promise of DM and other approaches that attempt to improve chronic care among populations enrolled in public sector programs. Pursuing a research agenda to answer this question will require a multi-stakeholder effort, because data and knowledge will have to be pooled and shared. It will also require funding,

which is likely to become the key constraint in light of shrinking research budgets, particularly in the public sector.

Expert Meeting on Disease Management Outcomes Measurement

1. Introduction

In the face of increasing health care costs, evidence of the need for system-wide health care quality improvement, and an aging population, disease management (DM) seems an intuitively appealing way to improve the coordination and quality of care, and ultimately improve health outcomes among the chronically ill. In broad terms, DM refers to a system of coordinated health care interventions and communications to help patients address chronic disease and other health conditions. Commercial health plans and large employers are embracing this strategy, with 96% of the top 150 commercial U.S. payers offering some form of DM service¹ and 83% of more than 500 major U.S. employers using programs to help individuals manage their health conditions.² Public purchasers of health care services are testing the waters: CMS has launched the Medicare Health Support (MHS), a large demonstration to evaluate DM,³ and several states are offering DM programs under Medicaid. However, DM has not been universally embraced. The Congressional Budget Office has concluded that there is insufficient evidence that DM reduces health care spending.⁴ A recent RAND review of the evidence found that, while DM seems to consistently improve processes of care and disease control, no conclusive evidence exists that it would improve long-term outcomes and reduce cost.⁵

In light of this ongoing debate and in order to further the Department of Health and Human Services (DHHS)' understanding of the issues surrounding the implementation and evaluation of DM interventions for public sector programs, the Office of the Assistant Secretary for Planning and Evaluation (ASPE) commissioned The RAND Corporation (RAND) to convene the Disease Management Outcomes Measurement Expert Panel meeting on January 16, 2008, at the RAND Offices in Arlington, VA. The meeting was moderated by Soeren Mattke, M.D., D.Sc., of RAND. The purpose of the Expert Panel discussion was to provide ASPE with an overview of the key issues and to inform the development of a research and policy agenda in this area.

2. Convening the Expert Panel

A. Selection of Expert Panelists

The objective of the meeting was to assemble experts to discuss measurement of the impact of DM on health outcomes and costs of care for public sector programs. A primary goal of the meeting was to consider the variety of DM programs and identify strategies to improve the assessment of their impact on individuals with multiple chronic conditions. Professionals from a variety of relevant fields and areas of expertise were approached to participate in the panel, including experts from the following areas:

¹ Matheson, D., Psacharopoulos, D., Wilkins, A. (Boston Consulting Group report 2006) Realizing the Promise of Disease Management: Payer Trends and Opportunities in the United States. Available at:

http://www.bcg.com/publications/files/Realizing_the_Promise_of_Disease_Management_Feb06.pdf. Accessed July 18, 2006.

² <http://www.hewittassociates.com/Intl/NA/en-US/AboutHewitt/Newsroom/PressReleaseDetail.aspx?cid=2037>. Last accessed July 18, 2006.

³ Available at: <http://www.cms.hhs.gov/CCIP/>. Accessed July 18, 2006.

⁴ Congressional Budget Office. An Analysis of the Literature on Disease management programs. October 2004. Available at: <http://www.cbo.gov/showdoc.cfm?index=5909&sequence=0>. Accessed July 18, 2006.

⁵ Mattke S, Seid M, Ma S. Evidence of the Effect of Disease Management. Is \$1 Billion a Good Investment? Am J Managed Care 2007; 13: 670-676

- Academia
- Direct Providers / Clinicians
- Employer Consultants
- Federal Government
- Integrated Delivery Systems
- Private Organizations / Industry
- Statistics/Quantitative Methods
- State Medicaid Programs

A total of nine (9) experts representing the private sector, public sector, and academia agreed to participate in the Expert Panel (Exhibit A).

Exhibit A: Expert Panel Members

Expert Panel Members	Affiliation
David Atkins, MD	Chief Medical Officer, Center for Outcomes and Evidence, Agency for Healthcare Research and Quality (AHRQ)
Melanie Bella, MBA	Senior Vice President for Policy and Operations, Center for Health Care Strategies (CHCS)
Ahmed Calvo, MD, MPH	Director and Chief Medical Officer, HRSA Health Disparities Collaboratives, Center for Quality, Office of the Administrator Health Resources and Services Administration (HRSA)
Jeffrey Lemieux	Senior Vice President, Center for Health Policy and Research America's Health Insurance Plans (AHIP)
Ariel Linden, DrPH	President, Linden Consulting Group
Linda Magno, MPA	Director, Medicare Demonstrations Program Group, Centers for Medicare & Medicaid Services (CMS)
Nancy McCall, ScD	Chief Scientist, RTI International
Gordon Norman, MD, MBA	Executive Vice President and Chief Science Officer, Alere Medical, Inc
Paul Wallace, MD	Medical Director for Health and Productivity Management Programs; Senior Advisor, The Care Management Institute and KP-Healthy Solutions, The Permanente Federation, Kaiser Permanente

B. Background Materials

Prior to the Expert Panel meeting, panel members were provided with the following background information to assist them in their preparation for the meeting:

- The Disease Management Association of America's (DMAA) Outcomes Guidelines Report 2007 ⁶
- A journal article entitled, "Guidelines for Analysis of Economic Return From Health Management Programs" ⁷
- A journal article entitled, "Evidence for the Impact of Disease Management: Is \$1 Billion a Year a Good Investment" ⁸

Panel members also were provided with a discussion guide that outlined the main issues to be addressed during the meeting. The guide included questions intended to prompt and focus the discussion.

C. Main Issues

During the Expert Panel meeting, panelists considered and discussed the following five main issues pertaining to the measurement of the impact of DM on health outcomes and costs of care and the potential for DM for public sector programs:

- **DM Overview**
 - DMAA definition
 - Variety of programs that are labeled DM
- **Outcomes Measurement Issues**
 - Performance measures
 - Attribution; commonly used methodologies and the strengths and weaknesses of various quasi-experimental, non-randomized research designs (e.g., selection bias, regression to the mean)
- **Characteristics of Medicare & Medicaid Populations**
 - Multiplicity of conditions combined with social needs
 - Implications for program content and operations
- **Realistic Achievements of DM for Medicare and Medicaid populations**
- **Future Directions**
 - Implications for research and policy agenda

⁶ **Disease Management Association of American** (2007). DMAA Outcomes Guidelines Report, Volume 2. Washington, DC

⁷ **Serxner S, Baker K, Gold D.** Guidelines for Analysis of Economic Return from Health Management Programs. The Art of Health Promotion July/August 2006

⁸ **Mattke S, Seid M, Ma S.** Evidence of the Effect of Disease Management. Is \$1 Billion a Good Investment? Am J Managed Care 2007; 13: 670-676

3. Panel Findings

A. How Did DM Interventions Evolve?

Overview

DM can trace its origins to the 1970s' concept of prospective medicine, where some health services researchers began evaluating risk in populations, which gave way to the formalized process we now know as health risk appraisal (HRA). This trend continued in the 1980s, when patients with diabetes were encouraged to perform self-care and measure their own glucose levels. The self-care movement continued to accelerate through the 1980s.

The term “disease management” was coined in the 1990s by the Boston Consulting Group, which promoted it as a value-added strategy for the pharmaceutical industry. The activity called DM began as a plan to improve medication adherence, because chronically ill patients were taking only about half the drugs prescribed to them. The idea was to increase adherence and thus increase sales and profits for pharmaceutical companies, while also ensuring better disease control.

Consequently, the industry has broadened from a focus on medication adherence to comprehensive management of chronic conditions and from single-disease programs to an integrated approach that targets all of a patient’s health problems. Programs have progressed from targeting very high-cost patients to addressing the needs of broader populations. The most recent development is a shift to offering services along the full continuum of care that range from health promotion and disease prevention programs, to classical DM that targets distinct chronic conditions, to case management for high-risk individuals irrespective of the underlying condition.

In parallel to the growing complexity in service offerings, the delivery models for disease and other care management have become more varied. DM is being embedded in patient primary care sites, as envisioned by the Chronic Care Model⁹; becoming a service provided by integrated delivery systems that combine insurance and provision of care; and being offered by external providers that can be free-standing companies or parts of health plans. The range of services now ranges from mass communication approaches via mailings, telephonic broadcasting and the Internet, to call-center based outreach, to more intense approaches, such as home visits.

If you’ve seen one disease management program, you’ve probably seen one disease management program.
Panelist

The fastest growth and the greatest diversity of offerings can be found in the commercial insurance market, and, increasingly, employers contract directly with vendors. DM has also become a routine benefit under several Medicaid programs, and the Medicare program has a long history of conducting pilot and demonstration projects to evaluate alternative program configurations.

With those developments, the term “disease management” has become increasingly more misleading, because it is the patient—not the disease—that is managed. Thus, the term “population health improvement (PHI)” might describe the interventions more appropriately.

⁹ See http://www.improvingchroniccare.org/index.php?p=The_Chronic_Care_Model&s=2 (accessed March 27, 2008) for a detailed description.

The programs have also become more patient-centered and reflective of an individual patient's needs rather than focused on managing a narrow aspect of care.

The current Disease Management Association of America (DMAA) definition of population health improvement explains that “population health aims to improve the health status of the target population, and can reduce health inequities among population groups.” Population health encompasses the realization that a range of physical, environmental, and socioeconomic factors contribute to health. By successfully managing health influences on individuals, population health endeavors to affect the complete physical, mental, and social well-being of a target population.¹⁰

This evolution in our conception of DM is exemplified by the Medicaid DM programs, which were initially very disease-focused but have evolved into a comprehensive care management umbrella under which specific diseases are targeted. There is an attempt to build programs based on the Medical Home concept. The medical home concept is an approach to providing comprehensive health care in which the primary care provider and patient form a partnership to ensure the patient's care needs are identified and fulfilled. Programs are also becoming more responsive to the social, economic and cultural needs of patients. For example, the Indiana Chronic Disease Management Program offers a broad range of educational materials and self-management support for Medicaid enrollees with various chronic conditions as well as personalized case management services for high-risk individuals. Those patient-directed services are combined with efforts to disseminate evidence-based practices to Medicaid primary care providers.

The growing complexity of DM has also brought new attention to long-existing challenges for Medicaid programs. For example, better screening will identify unmet needs for specialty care, which tends to be a bottleneck for the Medicaid population. Payment systems are sometimes not well aligned with care needs, as is the case for carve-outs. Also, many patients move in and out of Medicaid eligibility, which creates challenges to continuity of care. As a consequence, states are struggling with the question of how to optimally configure DM programs and how to align DM with Medicaid reform efforts.

Expert Panel Observations

As DM programs evolve and increase in complexity, how they interact and integrate their efforts with primary care providers becomes more challenging

The services external DM vendors offer are sometimes complementary and sometimes contradictory to the services primary care providers deliver in their practices. Providers may receive helpful guidance that identifies gaps in care and opportunities for improvement, but also contradicting messages or recommendations that are not suitable for a given patient. They must also deal with a great variety of DM organizations, all using different formats and styles for relaying information. Thus, there is a fundamental tension between the economies-of-scale that can be achieved by offloading activities to a centralized service and the desire to make care patient-centric. Ideally, the interaction would be person-based and customized, for example, through care managers who visit provider practices but can leverage the sophisticated data and decisions support systems of a large DM organization. However, it is not clear that such a model would be cost-effective.

¹⁰ Norman, G. (2008). HHS/ASPE Expert Meeting on Disease Management Outcomes Measurement. January 16, 2008, Arlington, VA.

There are also cultural barriers in that primary care providers are often reluctant to grant an external entity a role in the care process and to move to a team-based model of care. However, greater integration between external DM organizations and primary care is clearly needed. Achieving this integration will require that primary care providers learn to trust DM organizations to be a resource rather than a challenge or a distraction. Similarly, DM organizations must become more responsive to the needs and concerns of primary care providers in order to develop stable and productive partnerships.

As DM programs evolve in terms of increased complexity and costs, a key challenge is determining a cost-effective mixture of resources that best supports healthcare providers

Determining the mix of services and resources that provides the best value, i.e., can support and extend the scarce resource of the primary care providers' time, remains a challenge. The Chronic Care Model, for example, may offer a way to increase the value of DM programs. An attractive feature of the model is that it supports both physicians and patients with necessary resources and preserves the patient-provider relationship. Maintaining the appropriate mix of resources available to the clinician, including less-expensive personnel such as mid-level providers, promotes a team-based approach to care. Focusing on allocating resources to the patient promotes cost savings and effective care. This effort is particularly important with care for chronic conditions, allowing patients to spend more of their time at work or at home than at the doctor's office.

Health information technology is both an opportunity and a challenge for DM organization

The slow adoption of health information technology (HIT) tools, such as electronic health records (EHR) and electronic prescribing systems, is believed to create significant obstacles to improving care and reducing cost, in particular in primary care settings. HIT has the potential to facilitate the integration of primary care and DM by improving the flow of information between health care providers. This technology would greatly improve a process that today largely relies on the fax machine. Theoretically, DM organizations could become a valuable partner for the primary care provider and offer HIT solutions, such as patient registries, electronic health records, and decision support. DM could become a change agent to alter the way we think and manage work flow more efficiently and effectively in clinical practice.

However, the mere availability of HIT tools will not suffice, as experience suggests that practitioners vary in their adoption of these resources. To some degree, this variation is a consequence of technology problems.

Enhanced interoperability, standardized formats for content, and increasing transportability are needed to improve and further facilitate this process.

Another challenge is that many of the EHR systems

were essentially developed from a billing system and focus more on the physician than on the patient; thus, they lack the functionalities built into systems for population-based care. HIT

To me, one of the real leanings is that every time you hire an information technologist you probably should go to lunch with an anthropologist. What I mean by that is that it's about work flow and it's about recognizing that information technologies allow us to imagine doing things in wholly different ways. But people don't just automatically change their work, and it takes a great deal of support and leadership and time in order to get people to evolve their work.

Panelist

adoption also requires cultural change and fundamental changes in workflow, both of which are much harder to bring about than fixing the technology, implying that much time and effort is required to go from implementation of HIT to true adoption.

Summary

The DM industry has evolved from a strategy to increase pharmaceutical sales to a complex and diverse industry. The increasing uptake of DM also intersects with many of today's fundamental issues in health policy, such as the patient/provider relationship; cultural changes that are needed to progress from a medical centered way of care to a patient-centered way of care; the role of HIT and how to leverage and foster its use; and how best to align incentives and organizational fabric to ensure efficient and effective resource utilization. Thus, it is important to determine which DM approach provides value under what conditions for what population.

B. How Can and Should the Impact of DM be Measured?

The Issue

Determining which DM program provides value under which conditions, requires that we first define “value,” as the impact of DM programs can be measured along numerous dimensions. Those impacts are, for example, cost and utilization of care, non-medical cost (e.g., health-related productivity loss), quality of care, disease control, health status, satisfaction (patient and provider), and behavioral change. Which of those domains are included in an assessment of value depends on the particular context of a program. For example, impact on health-related productivity may be very important to an employer but would lack salience for the elderly Medicare population. It is also important to understand how the domains will be weighted and prioritized, as programs may not affect every domain equally. For example, a program may yield a negative return on investment (ROI) but have a positive impact on clinical quality.

The second issue after deciding which domains should be considered in evaluating a DM program is to identify the specific measures and metrics that will be used to capture performance along those domains. This decision requires weighing trade-offs and considerations, including access to data, cost of data collection, and intended use of the results.

Time-to-effect is an important consideration in selecting measures for DM programs, as some changes will materialize faster than others. Thus, realistic expectations as to when certain changes will occur need to be built into decisions. For example, evaluating a diabetes DM program after the first year will be unlikely to identify changes in rates of long-term complications, like blindness and renal failure, implying that such health status measures should not be included.

Measurement science has evolved greatly but many significant gaps remain. Measures for selected specific clinical conditions like heart failure and diabetes have received much attention, whereas other conditions, like stroke and arthritis, have received less attention. Better measures are needed to capture care for patients with multiple comorbidities and complex conditions because their range of care needs may not be appropriately reflected in disease-specific measures. To illustrate, many quality measures capture whether patients are treated with drugs that prevent long-term complications from a disease, such as beta-blockers to reduce mortality in coronary artery disease. However, co-existing conditions may create contraindications to these drugs, such as COPD in the case of beta-blockers. Also, the decision is more difficult in a patient with multiple conditions to initiate a treatment that will not reduce symptoms in the short-run but mainly improve prognosis in the long-run with potential short-term side effects.

Measures that address cross-cutting issues like care management and care coordination also have gaps. Further, operational definitions of measures must be standardized across entities. Such standardization would facilitate data collection, because providers would not have to follow different data collection protocols for different entities, and interpretation of results would be easier, because measures that are collected and reported by different entities would have the same underlying definition. An important example of such alignment is agreement between CMS and the Joint Commission on common definitions for a set of hospital measures.

A key issue that remains for many measures is risk adjustment so that differences in the underlying population are appropriately accounted for, especially when measures are used for purposes such as pay-for-performance or contract performance evaluation.

Expert Panel Observations

Measures selection is traditionally driven by contractual obligations

Decisions on measures are determined by the need to meet the performance targets specified in DM services contracts. Historically, the focus has been on financial targets, but more recently, other domains, such as patient satisfaction or quality of care, have been used. The business context continues to drive the measurement strategy, and measures are still being selected through a negotiation process between DM program operator and purchaser rather than through a scientific process.

This *ad hoc* process does not yield a comprehensive and sound measurement strategy, in turn preventing the industry from learning what is working and what is not working and how to change programs to make them work better. What is needed is more nuanced thinking about measures that reflect the changes expected from program implementation; when changes are likely to be observable; how change is defined; and how the results are fed back into the program. Implementing such a comprehensive measurement strategy may require involving an independent third party to avoid potential conflicts of interest and to allow for true learning as opposed to disputes about reconciliation.

Data availability remains an important consideration for measures selection

Data collection can be expensive. Organizations must consider the cost of retrieving the necessary data when developing a measurement strategy, and the cost of data collection must be in realistic proportion to the cost of running the program. Two extreme approaches should be avoided: allowing readily available data to drive the measurement strategy, and collecting too much data that may not even be useful for program evaluation.

Data availability and integration differ widely among organizations. While some organizations have sophisticated electronic health records that facilitate retrieval of clinically detailed information, others may be confined to claims data. Thus, the implementation of a measurement strategy must be based on the capabilities and infrastructure of the respective organization. Overall, improvement in providers' data capabilities allows more sophisticated measurement approaches.

In the measurement of clinical processes, we should move from a disease-specific paradigm to one that is more patient specific

Clinical process measures reflect the degree to which care is provided in accordance with professional guidelines and practice standards. These standards are, by their very nature disease or condition specific, reflecting the traditional focus on the disease, rather than on the patient. While the approach to chronic disease care has become more patient-focused, the approach to measurement has not yet followed

When we think about our current approach to clinical measurement—measurement of clinical processes—we're still very, very much burdened by the legacy of the last 100 years, where we still think about diseases almost exclusively. It's very logical; it just happens to not be all that useful in understanding the clinical course of the people who really drive most of the resource consumption.

Panelist

suit. Catching up requires emphasizing measures that capture how well a program has met a patient's need over using standardized measures. For example, rather than measuring only whether each diabetic has received an HbA1c test, one could envision measures that reflect whether the patient's five most important health issues have been addressed.

Patients in public sector programs often have multiple co-morbidities, presenting unique measurement challenges

The issue of patient-centered versus disease-focused measurement is related to the challenge of evaluating programs that are delivered to populations with multiple comorbidities and complex non-medical needs. For example, the average age of the Medicare population in DM is over 75 years; the U.S. Preventive Services Task Force guidelines may not fully apply to this age group because recommendations are often limited to younger persons. Research has also shown that for individuals with multiple comorbidities, guidelines may conflict. Thus, selecting measures that do not reflect the underlying population's true needs may actually worsen or lead to inappropriate care and unnecessary services with little expected effect on long-term outcomes. Better measures that capture the care needs of such populations are clearly needed. Non-medical needs or medical needs unrelated to the condition that a DM program is intended to manage might be a patient's most pressing and rate-limiting problem. For example, it may be necessary to address cognitive barriers in the form of limited health literacy or severe depression before attempting to manage a patient's diabetes. A truly patient-centered measurement approach must be able to capture the effectiveness of the overall approach, not just how often certain clinical tests are ordered.

Measures need to provide actionable information

The ultimate goal for DM is to improve health with an efficient use of resources. Since meaningful changes in health outcomes do not materialize quickly, intermediate measures are needed to determine early on whether the program has the intended effect. Traditionally, process measures that capture how well clinical care is being provided in accordance with professional guidelines and practice standards are used for this purpose. Usually, a substantial body of research links those processes to clinical outcomes. Yet an important issue for public sector programs is to better understand the linkage between process measures and outcomes in a frail population with multiple comorbidities. For example, providing an HbA1c test every year may not really affect the outcomes for this population. A process-focused measurement strategy may not be able to capture the complexity of care.

I think an area of really future need is to try to understand the linkage between process measures and outcomes in the frail population, in the sick population. I don't think we're there yet. I don't think we know if providing a hemoglobin A1C test every year or twice a year is really going to affect the outcomes for these really sick people.

Panelist

An alternative might be measures for intermediate outcomes that reflect the degree to which a disease is under control. Those can be clinical parameters, like blood pressure and LDL cholesterol, or utilization events—such as hospitalizations, emergency room visits, and readmissions—that indicate deteriorations in health status and quality of life and a failure of outpatient care to adequately control them.

It is also important to improve our understanding of how measures are correlated and interrelated. As the number of measures that are being used to evaluate a program increases,

it becomes harder to interpret the results, since findings can be inconsistent or even contradictory. A better understanding of how measures interrelate would facilitate the construction of composite measures that summarize the information embedded in multiple measures in an intuitive way. Finally, more knowledge of interrelation between measures might help to identify practice patterns that work—or do not.

Summary

The complexity of DM programs has evolved over time. Today's programs are more complex than their predecessors, which focused on a particular disease or a particular aspect of care, like medication adherence. This level of complexity creates a particular measurement challenge for interventions offered to patients in public sector programs, who often have multiple co-existing conditions and complex non-medical needs. As the growing complexity of the programs surpasses the measurement science, sound and actionable measures are often lacking. In addition, measurement needs to be reoriented from a provider- or disease-focus to a patient-focus, i.e., it should reflect not simply whether certain tests and services were delivered but also whether a patient's specific needs were met.

More research and development are needed to create measures that can realistically be implemented and will reflect what an intervention program is attempting to achieve. In other words, the measurement strategy must be aligned with objectives and context. Finally, more research is needed to improve our understanding of how measures are interrelated to help distill actionable information out of the measures, in the form of, say, composites or intermediate outcomes.

C. How Can Changes in Performance Measures be Attributed to a DM Intervention?

The Issue

After selecting measures, determining the impact of a DM program requires assessing the proportion of change in the selected measures that can be attributed to the intervention itself and the proportion that is likely attributable to other factors, such as secular trends or market-level changes. This process requires predicting (or measuring) what would have happened in the absence of the program. The most valid method of doing this is the randomized controlled trial, which ensures a reference group that is identical to the intervention group in all other aspects but the program. But controlled trials are very expensive and time-consuming and therefore rarely feasible outside of demonstration and research projects. They can also suffer from limited generalizability, because interventions are carried out by a dedicated staff on a highly selected group of patients; thus, findings from such trials do not always predict the real-world experience very well. Therefore, to evaluate operational programs, attribution approaches are needed that provide a reasonable balance between scientific rigor and feasibility. The determination of what constitutes a reasonable balance also depends on the context. For example, compared with employers, who make annual benefits selections, the Federal Government has a longer-term perspective, and its decisions have a much greater impact on the marketplace, leading to a need for more rigorous approaches.

These attribution approaches must deal with various threats to validity, e.g., confounding factors that may lead to biased estimates of program effects. The two most prominent threats in DM evaluation are *selection bias* and *regression to the mean*.

Selection bias means that the treatment group is not a representative sample of the underlying population. It might differ in ways that the evaluator can observe and thus adjust for, such as age and sex, but it might also differ in unobservable characteristics, such as motivation to engage in self-management activities. Some differences can be partially observable for the evaluator. For example, disease severity might be accurately determined based on clinical data, but it often must be approximated from past service utilization, if clinical data are not available. In the presence of selection bias, one can not determine the proportion of changes in the selected performance measures that are attributable to differences in the population that participated in the program and which are attributable to the program itself.

Regression to the mean refers to the observed tendency of outliers to revert to the mean over time. For example, patient A may have very high health care costs in a given year because of a severe exacerbation of his/her disease (for example, acute myocardial infarction resulting in cardiac surgery) However, in subsequent years, patient A's costs would likely decrease with or without a DM program, because such catastrophic events tend to be non-recurring. As many high-cost events in health are non-recurring, it is difficult to distinguish such regression to the mean from program effects. A sound attribution strategy must be able to deal with these and other threats to validity to ensure that the true effect of the DM program is reflected.

Expert Panel Observations

There is convergence on the issue of attribution in principle

Traditionally, researchers have tended to argue that only controlled trials can provide a conclusive answer to the question of whether a DM program was successful, whereas program operators have insisted that any strategy that was more demanding than a simple

before-and-after comparison was not feasible for routine program evaluation. More recently, a consensus seems to be emerging, in principle, that an equivalent comparison group is needed (and sufficient) to ensure proper attribution, although substantial disagreement remains over the meaning of equivalency.

I think it's really important for us to realize that there's a tension here between what, in a sense, is a research agenda and what's an operational agenda. And are there ways that we can actually live within that tension and reconcile and learn from it.

Panelist

Most notably, the DMAA has issued guidelines for DM program evaluation; two editions have been published, and a third is pending. The guidelines explicitly recommend the use of equivalent control groups where possible but acknowledge that a high standard of equivalency, as can be achieved in controlled trials, is not feasible for operational purposes. Instead, the DMAA proposes a modified pre-post evaluation approach that uses the part of the population that is not eligible for the program (often referred to as non-purchased or non-chronic group) as a reference group. The trend among this reference group would serve as the benchmark trend for the DM program group. In addition, the DMAA guidelines recommend adjusting for the historical differences in trend between the program and comparison group, because cost and utilization in a population with chronic conditions tend to have a different trajectory than in a non-chronic population. The guidelines also set forth explicit criteria for which patients to include in the evaluation and support the use of risk adjustment to further account for differences between program and comparison groups.

While the research community recognizes that the guidelines are a step in the right direction, many have argued for more rigorous evaluation designs and maintain that such designs can be used for operational purposes. A variety of such designs that do not require randomized control groups have been proposed and tested, such as the difference-in-differences approach or propensity score matching. Typically, those designs require individual-level statistical modeling, whereas the DMAA guidelines would require comparisons only of group means.

Heterogeneity of methods makes comparisons difficult

Various attribution approaches are differentially immune from bias; thus, the use of the method will have substantial impact on the estimates of program effect. In the early 2000s, private-sector DM programs commonly reported a ROI of 4:1 or even 8:1, i.e. savings of \$4 or \$8 for every dollar spent, when the evaluation was based on a pre-post comparison of program participants. Estimates that were derived with more rigorous methods were often much less favorable and may even have suggested that the program had no effect on health care cost. The challenge is that DM providers have entered contractual arrangements that guarantee high ROIs. It is difficult for them to explain to clients that such ROIs can no longer be achieved when more rigorous evaluation methods are used. Similarly, clients have become accustomed to an expectation of high ROI and may not consider an offer from a vendor that promises a lower but more realistic ROI. Until the sensitivity of the results to the choice of methods is better understood, vendors will resist adopting more rigorous and more uniform methods that would facilitate comparing programs and identifying best practices.

Attribution methods need to be sensitive to the context

Attribution methods vary widely in their complexity. Thus, they need to be chosen based on the intended purpose. Rigorous statistical evaluations that do not require random assignment to treatment and control can probably generate estimates that are close to the ones that

randomized trials would yield. Those approaches can be utilized in a research context, for example to conduct cross-program comparisons to determine differential effects on different populations. They could also be used for so-called book-of-business evaluation, in which the entire client base of a DM program operator is analyzed, or to evaluate the long-term impact on patients in a public sector program. But those designs are still too demanding for the routine reporting requirements of DM providers that need to provide the quarterly account-level reports typically required in the private sector market: More streamlined approaches are needed for such applications. Nevertheless, more research would be desirable into the divergence among estimates derived with the different methods, to get a sense of the bias introduced by simplifying the evaluation. Such information would make the tradeoffs between scientific rigor and feasibility more transparent and help to inform the choice of methods.

There is a wealth of untapped information on program differences

With the increasing penetration of DM and the growing diversity of programs, a huge laboratory has emerged that would allow, in theory, comparative analyses between programs and populations. Observational research designs could be used to take advantage of the natural variation among programs to determine what works best under which conditions. While such post hoc analyses have

When you actually have data on the whole population that's increasingly reflective of their transit through care, you can begin to trade off selection bias or other kinds of observations and we can begin to free ourselves of some of the constraints.

Panelist

been rejected by the scientific community in the past, a paradigm shift has occurred, which has led to the recognition that research based on so-called realistic or observational designs can yield important insight about intervention effects under real-world conditions.

Program evaluations need to go beyond measurement of cost savings to a broader conception of value

A simplistic focus on impact on direct medical cost does not allow the value of a DM intervention to be fully captured. A more holistic approach is needed, as the value proposition of DM programs is more complex than exclusively measuring cost, especially because truly cost-saving interventions are relatively uncommon in medicine. Moreover, some interventions/programs may increase cost but demonstrate a positive impact on the health status of a population for a reasonable investment. Thus, the DM community should move away from attempts to give a simple yes-no answer to the question of whether a program worked, based on direct medical cost and provide a more nuanced understanding of its effects on various endpoints.

Such answers would also allow purchasers of programs to select an intervention that suits their unique needs, as different purchasers will have different perspectives of what constitutes value. Employers might view effects on health-related productivity losses as critical, while productivity would not be an important issue for a program that is providing services to the elderly Medicare population. As another example, some employers might be perfectly comfortable with a program that increases medical cost, if they think the investment in the health of their workforce will be justified by non-tangible returns, while others might focus on short-term financial goals.

Different stakeholders will also view the question of value differently. Clinicians view value differently from patients, who in turn interpret value differently from payers. Patients may not

experience any quantifiable economic benefits from DM programs but may perceive improvements in quality of care.

To clinicians, DM programs may improve the efficiency of medical practices, because, for example, the social support offered by the DM program allows them to spend less time on the phone, but these clinicians receive little direct

I think we need to still get our arms around whether the value proposition here is saving money, or whether there are broader definitions of value, and adding value. I think that part of what many of us have alluded to is that it's more complex than just saving money, and we want to understand what are the additional determinants that go into that.

Panelist

benefit from reduced hospitalizations. In contrast, the purchaser will see value in a general reduction in costs or ROI.

Such considerations should not distract from the fact that resource implications play an important role in embracing innovations like DM. While improving quality of care is an important goal, payors must also consider the sustainability of the system. In addition, to make informed choices about inevitable tradeoffs, payors need reliable information on the impact a given investment will have on a particular dimension.

A broader conception of value will also make results more credible

Using a variety of endpoints rather than merely cost savings also allows checking for consistency of results. For example, if a congestive heart failure program reported substantial cost savings with significant reductions in hospital admission rates, the question should be raised whether the savings estimates are accurate.

Tracking endpoints over time will provide visibility for future program effects. For example, if medication compliance and use of preventive services increased in a diabetes program in the first year, and disease control parameters like HbA1c and blood pressure levels improved in the second year, it is plausible to assume that cost savings would materialize in subsequent years. Conversely, a program that reported cost savings without evidence for change in care patterns or parameters of disease control might have difficulties explaining its results.

Summary

The pressure to change the suboptimal *status quo* in chronic disease care has created a dynamic and innovative industry and pressure on purchasers to adopt DM programs. In some cases, new DM approaches were not evaluated rigorously before they were implemented in programs. The adoption of scientifically sound evaluation methods also lagged behind. Prevailing evaluation methods in the industry tended to provide very high estimates for DM-related cost savings, further fueling the enthusiasm. But more recently, skepticism about those estimates has emerged, and a discussion about evaluation methods that would yield more realistic results has begun. The DMAA has proposed voluntary industry guidelines, and the research community has argued for the use of rigorous non-experimental methods. While it is not clear yet which approaches will be universally adopted for which purposes, the various approaches are being discussed, and more research into their relative performance is needed.

The debate about evaluation methods has fueled a broader discussion about the value of DM programs beyond the narrow target of cost reduction. Different stakeholders attach different importance to different domains of value, and a holistic evaluation approach needs to be able

to express such a multi-faceted conception of value. This broader definition of value may not be as easy to communicate as ROI but will allow for a richer assessment of the impact of DM. Finally, there is a limited understanding of which DM programs work well under which conditions. More research is needed to compare and contrast programs to identify best practices.

D. What are the Implications of the Particular Characteristics of Medicare and Medicaid Populations for DM?

The Issue

Most of the experience with DM in general and population-based DM in particular has been derived from a working-age, commercially insured population. The experience with this population has shaped the development and refinement of programs. As DM is increasingly piloted and used in Medicare and Medicaid populations, the question arises whether and to what degree the interventions can be simply extended to patients in public sector programs or whether they need to be adapted. It is likely that the particular characteristics of the Medicare and Medicaid populations will affect DM's effectiveness and that both operations and content of the DM programs will have to be retooled.

Expert Panel Observations

Among Medicare and Medicaid patients, co-morbidity is the rule, not the exception

In contrast to a commercially insured population, patients in Medicare and Medicaid programs have a very high rate of co-morbid conditions. Medicare patients commonly receive treatment for multiple chronic diseases, and general frailty may limit treatment options and determine priorities. End-of-life issues play an important role, and seemingly minor issues, like hearing impairment or poor vision, can interfere with DM programs. Nevertheless, segments of the Medicare population can closely resemble the privately insured population, with relatively good health status, high literacy, and willingness to self-manage chronic disease. Certain segments of the Medicaid populations are affected by high rates of disability and a high prevalence of mental health disorders. In particular, the latter presents substantial obstacles to successful DM, as it will be very difficult to educate a depressed or schizophrenic patient about diabetes self-management.

Social issues have a strong influence on the needs of the Medicare and Medicaid populations

DM relies heavily on the patient's ability to interact with the program operator and to engage in the intervention. These two preconditions are not always met in Medicare and Medicaid populations. First, commercial programs make use of telephonic and Internet-based communications. Medicaid patients may not have access to a regular phone or may be difficult to reach, and both groups are typically less accustomed to using the Internet. Second, limited health or general literacy and limited English skills are more common among Medicaid patients than in the general population. Third, for some Medicaid and dually eligible populations, economic concerns may override health concerns.

Public sector programs tend to place much more emphasis on face-to-face interventions and have developed more expertise in comorbidity management with serious emotional and mental illness comorbidities such as schizophrenia and bipolar disorder. Public sector programs tend to have greater emphasis on social and economic dimensions of health, focusing on transportation, access to care, language barriers, and cultural sensitivities. Public sector programs also tend to work with safety net providers whose care does not largely overlap with private providers.

Panelist

Particular features of insurance coverage in public sector programs interfere with DM

Care continuity is a common challenge among Medicaid populations, who may move in and out of program eligibility. While analyses have shown that cost and utilization tend to be higher for patients who move in and out of eligibility, states cannot easily continue to provide services through eligibility gaps. Eligibility gaps typically affect younger Medicaid recipients whose eligibility is solely based on low income but are not common in the dually-eligible, disabled, and long-term care segments. The eligibility gaps also pose a problem for program evaluation, as it becomes more difficult to define the periods during which a patient can be considered engaged in a program. Many states do not enroll dually eligible patients in DM, as most of the benefits would accrue to the Medicare program, yet DM is not available to fee-for-service (FFS) Medicare patients outside of demonstrations and pilot programs.

Medicare patients, in contrast, enjoy comprehensive and stable coverage of health care. But care coordination remains a significant issue, particularly in the FFS segment. On average, FFS beneficiaries have six providers and often no designated primary care provider, leading to diffused accountability. Seasonal migration may further exacerbate this problem.

Access issues may interfere with DM as well

Related to the issue of insurance, patients in public sector programs face issues with access to care that may limit the effect of a DM intervention, because cooperation with the patients' regular provider is a key success factor. Many providers will not accept Medicaid patients for routine care. Among participating Medicaid providers, office visits may focus on the acute and most pressing problems, leaving little time for patient education and guidance on managing chronic conditions. Access to specialty care may be even more limited. While Medicare patients typically do not have financial restrictions on access, those living in rural areas, like all rural dwellers, face problems with physical access to care, particularly to specialty care.

Obviously, all these patients, by the fact that they have insurance, have some level of financial access to health care. That's a very different issue, though, about whether they really have access to health care in their community in terms of the availability of providers.

Panelist

Those particular characteristics have important implications for program design

At a very basic level, recruitment of Medicaid and Medicare patients into a program cannot rely solely on call-center operations as is commonly done in programs for the commercially insured population. Contact points that allow for face-to-face encounters, like assessment centers or home- and practice visits, may be required. The same is obviously true for program operations. Lessons from Medicaid programs and Medicare demonstrations, like the MHS, have shown that mass communication technology and telephonic outreach need to be combined with direct contact. While various models have been implemented, not enough is known yet to determine which mix of services works best in which population. Similarly, the content of the interventions, which often focus on improving medical care for isolated chronic conditions in programs for commercially insured patients, had to be modified to address the multiplicity of medical and social problems that are common in Medicare and Medicaid populations.

The complex nature of patients also provides for greater opportunities to intervene

The upside of the complex nature of problems faced by patients in public programs is that the opportunities to make a difference in their lives and to increase efficiency are substantial. Spending is typically highly concentrated, with about 5% of the patients consuming 50% of the resources, and potentially avoidable high-cost events, such as hospital admissions and

readmissions for chronic conditions, are common. But identifying the patients and the conditions that are likely to benefit from a DM intervention is important, as several groups, such as the severely cognitively impaired, may not be amenable to this type of intervention. Predictive modeling may help with this task.

DM intervention in public sector programs may need to consider different targets than the programs for commercially insured patients, which typically aim at improving disease control to avoid high-cost exacerbations. Greater opportunities could possibly be found in areas where medical and social needs intersect, such as end-of-life care. If long-term disease control and secondary prevention were the target, one possibility might be to enroll 55-65 year olds (the near elderly) in Medicare DM programs, because early management of chronic conditions may yield long-term benefits.

The beauty of Medicaid and Medicare is that the complexity of the populations and the financial levers that are available mean that there should be a lot of attention and a lot of opportunity to enhance some of the measures. I think the caution is that we not get too ahead of ourselves with some of the programs and the claims made about the programs without the measurement and the evidence to back that up.

Panelist

The cost-benefit calculus for public sector programs is likely to be different

The need for more intense outreach and more customization implies that public sector DM programs will be considerably more expensive on a per member per month basis than their private sector counterparts. However, the possible return is also higher because of the much greater burden of disease and higher spending levels. It is not yet clear how to configure programs to achieve reasonable value for the money.

Summary

The medical and non-medical needs of patients in Medicare and Medicaid programs are considerably more complex and variable than those of commercially insured patients. Social and basic access issues may dominate concerns over chronic care management. This difference implies that programs will have to be adapted for those populations and will likely be more intense and costly. Nevertheless, the greater burden of disease in those populations means that the potential financial return could be large (especially if one takes the long run perspective of sustaining Medicare's finances) as would the potential welfare gain. Realizing those opportunities will require a lot of creativity to address challenges of implementing and operating DM programs in those populations. More research is needed on what is working, for whom, and at what costs.

E. What can DM Realistically Achieve for Medicare and Medicaid Populations?

The Issue

In spite of the rapidly growing penetration of DM in the commercially insured market, there is surprisingly little evidence for the effect of large-scale population-based programs. A recent RAND review identified only three rigorous evaluations of such programs but a substantial body of research on smaller-scale interventions that are typically designed and operated by providers at a single site. Overall, the review showed consistent evidence that DM could improve quality of care and disease control in chronic conditions, but no conclusive evidence that it could reduce cost. This finding contradicts publicized reports of DM vendors, which commonly tout substantial cost savings.

In the public sector, numerous demonstrations and pilot programs have been conducted in the Medicare population and many states have adopted DM for the Medicaid population to cover conditions like asthma, diabetes, heart failure, coronary artery disease, and COPD. Some states are also focusing on schizophrenia and AIDS. Medicaid DM interventions can include a variety of program segments, such as the FFS population, Medicaid Managed Care, and specialized carve-outs, but not the dual eligible segment, because savings would accrue to Medicare. Conditions targeted in Medicare programs have focused on discrete chronic conditions such as CHF, diabetes, and end-stage renal disease (ESRD), as these conditions were thought to identify high risk cohorts and populations. Some programs targeted only high-cost populations but in all cases tried to get programs to focus on the whole person. Elements of these programs include a spectrum of different interventions including, but not limited to, physician home visits (house calls), care management, and remote monitoring.

As results from Medicare demonstrations and pilot programs have so far failed to show cost savings and improvements to outcomes, some realism may be necessary in terms of setting expectations for these programs. This lack of results may not be surprising because the burden of disease in this population is great and the opportunity to provide additional care is substantial. It is possible that the needs of this population are too complex to be fixed with a simple DM design and that more is needed in terms of integration and multi-faceted interventions to fundamentally change care experiences and ultimately outcomes. If such interventions can be crafted at reasonable cost, savings may be achievable albeit over a very long time horizon.

Expert Panel Observations

Medicare demonstrations have not yet demonstrated cost savings

The Medicare program commonly conducts demonstration and pilot projects to test innovative ways to organize, finance, and deliver care for its beneficiaries. Typically, the key criterion to evaluate the success of those projects is that they are at least budget-neutral. This criterion is often explicitly required in

Congressionally mandated projects, such as the Benefits Improvement and Protection Act (BIPA) or MHS. For the MHS, the goal was even more ambitious as the vendors initially agreed to a target to reduce spending by 5% net of their fees, but the target was later reduced to budget neutrality.

We don't have final results on all of these projects. We have interim results on all of them, and we're getting more in daily. . . Looking at treatment group versus control group spending, across all of these, virtually none of the sites is covering even half of their fees. This goes across the board, except we have four sites out of thirty-five where we have some glimmers of hope.

Panelist

The Centers for Medicaid & Medicare Services (CMS) has experimented with a great variety of approaches to improve chronic care, ranging from telephonic interventions to remote monitoring to home visits. Those interventions have targeted a variety of populations, such as patients with distinct chronic conditions, high-cost patients, and dual eligibles. Largely, those programs proved to be disappointing, at least on the basis of the currently available data that reflect only interim evaluations for some projects. Only four of 35 tested interventions have shown some promise; three programs in the Medicare Coordinated Care demonstration have demonstrated that they may be budget neutral, and one DM demonstration for dual eligibles has shown some success but it has been reduced in size and is a narrowly targeted program.

Medicare demonstrations have also failed to show effects on quality of care

In addition, Medicare demonstrations and pilot projects with DM have generally not led to significant improvements in quality of care. One reason for this finding may be that the baseline level, at least as determined by the commonly used disease-focused indicators, is relatively high, making it difficult to achieve significant improvements over the status quo. Subgroup analyses within individual projects have sometimes suggested beneficial effects on selected patient groups but have not yet revealed any consistent patterns that could be used to develop targeted interventions. There is also the risk that intense *post hoc* analysis might over-interpret the data. No attempts have been made thus far to pool data across projects to investigate whether common patterns emerge, but CMS is considering building a database that would inform such analyses.

There are no conclusive results for the impact of DM in Medicaid programs

Numerous DM interventions have been used and are being used in State Medicaid programs, either as demonstrations or as routine operations, typically driven by a desire to contain cost. Unfortunately, few formal evaluations of those interventions have been conducted, as states commonly lack the resources to conduct them. While vendors often estimated substantial savings, internal analyses and external evaluations tended to be less favorable. For example, a recent report by Mathematica Policy Research that integrated the experience of several states stated that most programs did not show any effect on cost in the short evaluation timeframe, even though some directional improvement was seen. This finding is consistent with results of an HRSA HIV/AIDS management program that failed to show an effect after two years but proved successful after four. A particular challenge is the lack of a platform that would allow states to pool their data and conduct cross-program research with the goal to learn from each other and identify best practices.

Greater levels of integration may be required

Lessons from the Medicare Group Practice Demonstration seem to indicate that a greater level of integration may be needed to change chronic care. Those interventions, which were situated at large group practices with sophisticated electronic medical record capabilities and integrated telephonic outreach with care coordination approaches at the practice level, tended to show promise. Thus, a key difference from predominantly external DM approaches is that they were driven by the patients' primary care providers and integrated in the patients' usual site of care. Similarly, Kaiser Permanente sees some success with its attempts to use the Chronic Care Model to improve office-based care and to combine it with telephonic outreach.

The limited success of public sector DM programs offers opportunities for learning

The limited evidence for success of DM in Medicare and Medicaid programs thus far challenges the assumption that DM has a positive impact on cost and quality of care. But this finding should not lead to the conclusion that chronic care cannot be improved. The results should be used to look ahead and devise new approaches, as there is still a great need to change the current suboptimal and costly care patterns.

In a sense, demonstration projects invert the fundamental question of how to improve care for chronic conditions in populations with complex needs and instead ask whether a pre-conceived approach, like DM, will work. More fundamental questions need to be asked, such as what the

As someone who's sort of trained to be skeptical about evidence, I still want to say that I hope we recognize perhaps the limitations of what we can learn from this evidence, and think about how to, as people suggested, learn as much as possible. You know, I would hope the message isn't that we can't improve the care of chronic diseases, but it's just that we need to be more thoughtful about the targeting and the models.

Panelist

to

relative roles of patient and provider should be in chronic care, how to achieve lasting behavior change, and how best to manage care over transitions and hand-offs. Answers to such questions may allow a portfolio of interventions to be crafted that can address the various needs of chronically ill patients. Another area that may offer important lessons is the challenge of implementation, which may have contributed to the limited impact of some demonstrations.

Moving forward, the focus needs to include multi-dimensional interventions

Policymakers, patients, and providers are confronted with an unacceptable status quo in chronic disease care, resulting in pressure to achieve change. In such a situation, it is tempting to look for the “magic bullet”, a well-defined, practical and scalable solution that will improve care and contain cost without the need to fundamentally reorganize the health care system. So far, the success of such proposed solutions has been elusive. Great hope has been placed in prevention and DM, and more recently on HIT, risk reduction, and the Medical Home. It is likely that these newest fixes, when used in isolation, will also fail to produce the hoped-for results, as suboptimal care for chronic diseases is a complex problem with deep roots in the fundamental ills of the health care system and the difficulty among some patients to sustain behavior change. New approaches are necessary to address or at least patch those fundamental ills, even if only for a selected subgroup of patients. For this, we need more research into what works for whom under which conditions.

It's really the evaluation of that multiple interventions happening simultaneously that I would encourage . . . The reason is that if we spend all our time evaluating single interventions I think the basic evidence on those is already in there and it would be a waste of money to do these things over and over in small little incremental analyses only.

Panelist

4. Conclusions

A. What Do We Know About the Promise of DM for Public Sector Programs and Where Are the Gaps in Knowledge/Evidence?

DM has been met with great expectations and sometimes enthusiasm, as it seems an intuitively plausible approach to cure the twin ills of the health care system—inadequate quality of care and rapid cost growth. Its premise is that by managing chronic conditions tightly according to professional guidelines and accepted standards of practice, it can improve the health and well-being of patients, while at the same time containing cost growth by avoiding high-cost exacerbations. Evaluations by DM program operators appeared to support this optimistic view and commonly reported substantial savings.

However, those estimates were often based on flawed evaluation methods. Current evidence in the scientific literature suggests that DM may lead to better control of chronic diseases due to improved processes and better quality of care, but little evidence exists that DM programs can improve long-term outcomes and reduce medical cost. A particular problem is the very limited published evidence in the peer-reviewed literature for the large population-based programs that are currently considered by private and public purchasers. Evaluations of Medicare demonstrations also tended to show little effect of the DM interventions.

This lack of evidence for an effect of past DM programs may be due to their simplistic approach. The needs of the complex populations in public sector programs and the deficits in managing chronic conditions will require a well-crafted solution. Over time, DM programs have evolved and become very complex to include integrative and multifaceted approaches to enhance the care experience, which in turn should improve patient outcomes. Today, these programs are more intricate than the simple disease-specific interventions that narrowly focused on medication adherence. Unfortunately, we do not know what approach to DM works best for whom, at what cost, and consequently we do not yet quite know what an optimal DM intervention would look like. We also need to keep in mind that DM can make only an incremental contribution to healing the fundamental ills of the U.S. health care system, which include lack of integration and lack of HIT adoption.

The Expert Panel provided some advice on how to move forward with assessments of DM's impact. We need to get beyond the black and white question of whether DM works to richer and more contextual questions that explore the value of DM and how value can be measured and expressed. For this advancement, more and better measures are needed. We also need to better understand how to improve care for patients who are frail and have multiple chronic conditions and for patients who have both social and medical needs. Moving forward, this undertaking will require a multi-stakeholder effort because it will require collaboration and sharing of knowledge and data among multiple parties. It will also require funding, which is not as readily available in the public sector today as it may have been in the past.

In the future, an approach might emerge that we may refer to as post-modern DM that surpasses our current narrow view, a multifaceted, intelligent, patient-focused approach with richer interventions.

B. What are the Implications for a Research Agenda?

To get to interventions that are both effective and efficient, more research is needed that is rigorous yet feasible. Although randomized control trials may not be the solution, we can

continue to learn by doing more observational research, more post hoc analyses, and more data integration, by marrying claims data with care management and patient self-reported data in order to get a richer picture of what might work for which population at what cost. The Expert Panel pointed to three areas that could be targeted with a research agenda.

Measuring value in chronic care

Public and private purchasers increasingly acknowledge that a broader conception of value is necessary, one that moves beyond the focus on cost and ROI. But to re-orient innovations toward creation of value and away from mere cost containment, the complex construct of value must be measured and tracked. First and foremost, a debate is needed to identify the relevant domains of value, acknowledging that those might differ with the particular context of an intervention. Many domains that are commonly seen as relevant still lack universally accepted and scientifically sound measures. For example, the area of quality of care has seen convergence of key stakeholders on core sets of process measures for the most common chronic diseases, like diabetes and heart failure. However, much less progress has been made for other chronic diseases, like stroke and arthritis; for less well defined conditions, like frailty and cognitive impairment; and for areas where medical and social needs interact. More complex constructs, like care coordination also show substantial gaps in the measurement science. More and better measures for such domains of value are a precondition for research in general and for research that integrates knowledge across programs and settings in particular.

Benchmarking evaluation methods

A wealth of evaluation methods has been developed and tested that bridge the former dichotomy of randomized controlled trials and simple pre-post comparisons; these two types of trials form the two extremes in terms of the inevitable tradeoff between scientific rigor and feasibility. Those methods, which are often referred to as observational or naturalistic designs, have the great advantage of not requiring random assignment to treatment and control groups and would therefore allow for prospective and also retrospective analyses within and across programs. A limitation is that, while those methods are widely used, there is little knowledge of their comparative performance. More research into comparing and contrasting results that were derived with different types of these methods would not only help to identify the most suitable models but would also produce insight into how closely observational designs track estimates that are derived from randomized trials and how much bias simplified methods would introduce. This insight would be of tremendous value, particularly for the selection of evaluation approaches for operational purposes.

Cross program evaluation

With the growing realization that simplistic interventions will not suffice to address the problem of suboptimal chronic care, research needs to combine the experiences of various interventions to provide insight into the big picture. Lessons can also be learned from the experiences with chronic care innovations in other countries. We have not done enough of this kind of integrative research thus far. Given the complexity of needs of patients in public sector programs, best practices and success factors are likely to be context-specific so that integrative research could help to develop more targeted and tailored interventions.

In summary, there is much more that can be done to explore and explain the promise of DM and other approaches that attempt to improve chronic care for public sector programs. Pursuing a research agenda to answer these questions will require a multi-stakeholder effort because data and knowledge will have to be pooled and shared. This effort will also require funding, which is likely to become the key constraint in light of shrinking research budgets, particularly in the public sector.