Final Contract Report

Monitoring and Evaluating Medicaid Fee-for-Service Care Management Programs: A User's Guide

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-04-0001

Prepared by:
AcademyHealth, Washington, DC

Authors
Sharon B. Arnold, Ph.D.
Amanda Folsom, M.P.H.
Emily Bosk

AHRQ Publication No. 08-0012
November 2007

This User's Guide was prepared by AcademyHealth under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-04-0001). The findings and conclusions in this document are those of the authors, who are responsible for its content, and do not necessarily represent the views of AHRQ. No Statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

This User's Guide is intended to help policymakers and others understand the evidence, methodology, and issues related to evaluating the costs, quality and impact of care management programs.

Suggested Citation:
Arnold S, Folsom A, Bosk E. User's Guide for Monitoring and Evaluating Medicaid Fee-for-Service Care Management Programs. Final Contract Report (Prepared by AcademyHealth under Contract No. 290-04-0001.) AHRQ Publication No. 08-0012. Rockville, MD: Agency for Healthcare Research and Quality; November, 2007.

Foreword

Across the country, a wide range of care management (CM) initiatives are underway in Medicaid programs. These initiatives seek to reduce costs and improve the quality of care for individuals with chronic conditions. The Centers for Medicare & Medicaid Services (CMS), private vendors, States, and others are undertaking evaluations to determine whether CM programs will lead to cost savings.

With so much activity in this arena and a high degree of variability among care management programs, States and others are interested in understanding how to weigh the evidence presented by vendors and relate it to the array of program designs being offered. In particular, a clearer understanding of methodologies and issues related to evaluating the costs and quality of these programs will help States to assess the impact of their CM initiatives.

This guide walks through the steps necessary for evaluating a Medicaid CM program to assess the economic and quality impact of CM interventions. From how to get started, to thinking about the budget, to executing the evaluation itself, this user's guide is a resource through each stage of the evaluation process. Sections contain background information, State examples, checklists, and charts to help you answer your evaluation questions and make sense of the process.

What does the term "care management" mean?

In this guide, we use the term "care management" to refer to disease management programs, case management programs, and care coordination programs.

The guide moves beyond "one size fits all," providing information and perspective on the benefits of specific designs for specific situations. In this way, it can be a resource to States as they create their CM programs or negotiate evaluation methods during CM vendor contracting and reconciliation. In addition, the guide can assist States in critically reviewing other evaluation findings, including published articles and data reported by CM vendors in their proposals and marketing materials.

The guide will be especially useful for decisionmakers and others involved with designing CM programs and overseeing their evaluation. These include directors of Medicaid care management programs, quality improvement directors, contract negotiators, program analysts, Medicaid medical directors, and program evaluators. The guide includes a section that can be used to educate higher level policymakers (*e.g.*, legislators, cabinet secretaries, governors' budget directors) about the importance of evaluating CM programs and how their decisions can impact the viability of these evaluations.

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

David Atkins, MD, MPH
Center for Quality Improvement and Patient Safety
Agency for Healthcare Research and Quality

Acknowledgments

This guide could not have been produced without the assistance of a number of people. AHRQ and AcademyHealth acknowledge the Lewin Group and the AHRQ Care Management Learning Networks for ongoing collaboration and feedback. In particular, we acknowledge and appreciate the time and effort of the following individuals:

Ron Ackerman, MD, MPH
Assistant Professor of Medicine
Affiliated Scientist of the Regenstrief Institute
Indiana University School of Medicine

Melanie Bella, MBA
Senior Vice President
Center for Health Care Strategies

Randall Brown, PhD
Vice President and Director of Health Research
Mathematica Policy Research, Inc.

Don Fetterolf, MD, MBA
Corporate Vice President
Health Intelligence
Matria Healthcare, Inc.

Denise Levis Hewson, RN, BSN, MSPH
Director of Quality Improvement and Clinical Operations
Community Care of North Carolina

Thomas Inui, MD
President & CEO
Regenstreif Institute

Jeanette May, PhD
Director of Research
Disease Management Association of America

Jean Moody-Williams, RN, MPP
Director, Division of Quality, Evaluation and Health Outcomes
Centers for Medicare & Medicaid Services

Marc B. Rosenman, MD
Research Scientist
Regenstrief Institute
Assistant Professor Pediatrics
Indiana University School of Medicine

Thomas W. Wilson, PhD, DrPH
Epidemiologist
Founder and Board Chair
Population Health Impact (PHI) Institute

Contents

Appendixes

Chapter 1. Introduction to Care Management Programs

Background

Patients with chronic illnesses represent the largest consumers of health care dollars, accounting for nearly 75 percent of total health care expenditures in the United States, including 76 percent of all hospital admissions, 88 percent of all drug prescriptions, and 72 percent of all physician visits.[1,2] As medical costs continue to rise, care management (CM) programs have been touted by the medical community, Wall Street, and State legislatures as a critical tool for cost containment.

CM programs focus on patient self-care and education to increase overall health and wellness and to avoid more costly medical interventions that result from poor self-care or a lack of adequate information. Emphasizing preventive measures, CM programs aim to reduce health care costs by limiting acute episodes of chronic illnesses. They use a variety of approaches to achieve this goal. Common features of CM programs include patient education, nurse-case manager interventions, 24-hour nurse hotlines, uniform provider guidelines to standardize care, increased attention to high risk populations, integrated health care delivery systems, and evidence based treatment plans—though this by no means represents a comprehensive list.

CM programs evolved in the private sector in an era when providing more efficient care for the chronically ill became a major priority of managed care. As States increasingly implemented Medicaid managed care programs, the perceived success of CM in the commercial market led to the integration of CM within Medicaid managed care. States have realized, however, that their most costly enrollees—those with chronic illness, who make up the majority of adult enrollees (approximately 60 percent)—are often served outside managed care in either Medicaid fee-for-service (FFS) or primary care case management (PCCM) programs.[3] States have therefore begun implementing CM within their traditional Medicaid programs, outside their managed care infrastructure, to address the needs of the chronically ill. In February 2004, The Centers for Medicare & Medicaid Services (CMS) sent a letter to State Medicaid directors encouraging States to take advantage of CM opportunities in their fee-for-service (FFS) and primary care case management (PCCM) programs.[4] More than 30 States have integrated CM into their Medicaid FFS or PCCM programs.[5]

Vendor vs. In-House Programs

Throughout this guide there will be different scenarios presented for Medicaid CM programs run "in-house" and those contracted to vendors. Although key elements affecting evaluation design and implementation will, on the whole, be similar for both types of programs, this guide will highlight cases where approaches may differ.

States craft their programs either "in-house" or through a vendor and use a variety of care models in addition to FFS and PCCM, such as enhanced primary care case management (EPCCM) and Medicaid-Medicare demonstrations (Medi-Medi). While the majority of Medicaid CM programs focus on ameliorating the medical and cost effects of specific conditions, States have recognized

the need for their programs to evolve to address the significant comorbidities of many Medicaid consumers.

Similar efforts to reduce health care costs through targeted interventions to increase overall health have emerged alongside CM. For example, employer-sponsored wellness programs aim to decrease the incidence of absenteeism, presenteeism (i.e., on the job but not at full function because of a medical condition), and health care expenditures through a variety of health promotion activities. Wellness programs in the workplace range from modest programs focusing on preventive health to comprehensive behavior modification programs for a variety of conditions, including substance abuse and obesity. Cost-benefit analyses of wellness programs have produced different results, highlighting once again the need for standardized evaluation and further study.[6]

Promises of Care Management

Medicaid programs may seek a variety of outcomes when launching CM programs. Some potential goals are:

Goals related to quality:

- Reduced health complications of chronic diseases.
- Increased use of evidence-based practice guidelines by providers.
- Increased adherence to treatment regimens by patients.
- Increased coordination of services across providers.
- Increased school or job attendance.
- Improved indicators of high quality care delivery, e.g., improved Health Plan Employer Data and Information Set (HEDIS) scores.
- Improved health and functional status of patients
- Increased ability to perform Activities of Daily Living (ADL) and improved quality of life
- Better management of psychosocial effects of disease
- Improved quality of life

Goals related to cost savings:

- Reduced medical complications from chronic disease
- Reduced lab tests for acute episodes
- More cost-effective treatment choices
- Fewer unscheduled office visits
- Fewer ER visits
- Fewer hospitalizations
- Shorter hospital stays
- Fewer hospital readmissions

Importance of Evaluating Care Management

Expenditures for CM programs have exceeded $1 billion. Despite that level of investment, or perhaps because of it, there continues to be debate about whether CM programs actually save money.[7,8,9] Costs for CM programs can include:

- Staff time (nurses, nurse-case managers, hotlines, administrative staff, contractors, actuaries, etc.).
- Costs associated with contracts.
- Data collection, reporting, monitoring, and evaluation.
- Shifts in utilization (e.g. increased use of office visits).
- Increased use of prescription drugs.
- Increased uses of lab tests.
- General program costs (including administrative costs).

Because there are numerous potential choices for the design and delivery of programs to enhance the health and health care of people with chronic illness, it is imperative that States strive to understand if current models of CM are of greater value than available alternatives. Poorly conducted evaluations provide varying data and ultimately little evidence to assess States' return on investment. Carefully conducted evaluations are critically important to determining the impact of CM on economic and quality outcomes for patients, States, and the health care industry.

Evaluation can achieve three objectives: (1) determining if program goals are being met, (2) identifying whether performance improvement is possible, and (3) determining whether similar effects can be achieved more efficiently. Evaluation provides the information necessary to make the most informed decisions about the program's future. Strong evaluations can provide an answer to the key evaluation question: "What would have happened in the absence of the CM program?"

Chapter 2. Getting Started

Completing a timely and detailed Medicaid program evaluation can be a daunting and expensive task.[10] Nevertheless, its importance is well worth the effort and investment.

It is never too early to start thinking about evaluation. Indeed, it is important to consider the evaluation even as you begin to design your CM program, as design considerations have significant impact on the outcome of your evaluation. Program design impacts everything from the data that will be available to conduct an assessment to the design of the evaluation itself. Planning early enables you to carefully consider how to maximize resources, and it potentially saves money and time. However, evaluation shouldn't be thought of as a one-time-only activity. Ongoing monitoring and assessment activities are also important, and data needs for these activities should be considered early on.

As you begin planning for the evaluation of your Medicaid CM program, there are a number of important factors to consider as you balance a rigorous methodology with a feasible evaluation plan. This chapter is organized around several key steps for getting started.

Getting Started Action Steps

- ❒ 1. Identify core program goals and evaluation questions.
- ❒ 2. Identify which data are needed and which are available.
- ❒ 3. Establish an evaluation timeframe.
- ❒ 4. Determine how much funding is available.
- ❒ 5. Select evaluators.

Action Step 1: Identify Core Program Goals and Evaluation Questions

Goals

The first critical step in launching CM evaluation efforts is to identify your core program goals and what you hope to accomplish through performance monitoring and evaluation efforts.

Although they operate CM programs in very different contexts, States generally want to demonstrate value—that is, they want to know that their CM program is improving the health of enrollees while yielding an economic return. Value is subjective—there is no single measure that assesses value.[11] For the purposes of evaluation, you will want to identify both a group of measures and a credible methodology that allows you to assess whether your CM program is demonstrating "value" as defined in your State's context. In completing this step, it is critical to maintain realistic expectations for what your CM program can accomplish.

States invest in evaluation efforts for many different reasons, including:

- **Determining whether the CM program is successful.** States want to know whether their CM programs are successful in meeting their core goals. In making this determination, they want to understand what would have happened had they not implemented their CM program.
- **Program management and quality improvement efforts.** States may rely on their evaluation as a tool for managing their programs, making programmatic adjustments, and improving quality.
- **Budget and cost-containment efforts.** Some States implement CM programs with explicit savings targets and outcomes expectations and need to evaluate whether those expectations have been met. For example, in Washington[12] and Texas,[13] State officials were given cost savings goals through legislation. As a result, one important aim of their evaluation efforts is an assessment of whether specific cost savings have been achieved.
- **Funding and reauthorization.** Some States invest in evaluation efforts so that they can make a case for ongoing or increased funding or program reauthorization by CMS or State legislation. Evaluations may also respond to specific waiver requirements or legislative inquiries.
- **Expansions.** Some States aim to expand their CM programs to new populations or new areas of the State by demonstrating quality improvements and cost savings.
- **CM vendor procurements and contract renewals.** States may seek to make adjustments in their CM delivery systems or make contracting decisions based on evaluation findings.

A thorough and balanced evaluation should have explicit goals. New evaluation questions may arise over time as discoveries are made and concerns are voiced.[14] Flexibility in defining the scope of the evaluation is also important. You will find that conducting a comprehensive evaluation is often an iterative process.

Evaluation Questions

States generally want their evaluations to answer multiple questions such as:

- Have health outcomes and quality of care improved?
- Have we achieved gross and net cost savings to Medicaid?
- Have we achieved a positive return on investment?
- Was the chosen CM approach (as implemented) less costly to Medicaid compared with the available alternatives (e.g. the prior model for CM; alternative CM approaches)?
- Compared with the available alternatives (e.g. the prior model for CM; alternative CM approaches), did the chosen CM approach (as implemented) result in higher immediate costs to Medicaid but with the return of significantly higher quality that is likely to reduce future expenditures considerably?
- Have enrollees demonstrated better self-management?
- Have enrollees and providers expressed satisfaction with the program?
- Have enrollees used more preventive and primary care services and fewer acute care services, such as emergency department visits and inpatient services?
- Have we reduced health disparities?

- Have providers adhered to evidence-based practice guidelines?
- What was the impact of factors other than the CM program (e.g., provider incentives) that may cause the same outcomes targeted by the CM program (and, thus, may explain some of the "impact" being attributed to the program)?
- Has the program reached the intended population?

Factors that Make Evaluation Difficult

Evaluating a CM program can be a challenging task. Multiple factors make evaluation of State Medicaid CM programs difficult. It is important to keep mitigating factors in mind as you plan, conduct, and present your evaluation. Some of these factors are:

- Limited resources (staff time, money), combined with pressure to collect findings quickly, create constraints.

- Decisions by State legislatures guide evaluation design as officials dictate implementation requirements. For example, some States may not have the opportunity to do pilot programs.

- The timeline used for evaluating the impact of CM programs affects the results obtained. Some interventions will take longer than 1 year to see significant results, while other savings seen immediately may decrease over time. This occurs because of the complex interplay of program implementation factors (e.g. the natural phase of program "ripening" or "ramp-up" until full effectiveness is achieved), disease factors (e.g. time and effort are invested now to prevent disease-related complications in the future), and timing (e.g. an influenza vaccination campaign will have large early effects, particularly if a program is launched in late summer, it is an epidemic influenza year, and efficacious vaccine is available for delivery to clients).

- Outcomes can vary by population or disease group. For example, a Medicaid CM program for patients with asthma could result in cost savings at the same time that a program for patients with diabetes in the same State could experience net losses. These differences in impact could occur for a variety of reasons, including the population targeted and the mechanism by which impacts are achieved. Some programs target groups that have been underserved and may uncover significant unmet needs that can drive costs up in the short term.

- The particular characteristics of the Medicaid population can further complicate the evaluation process. With frequent "churning" in Medicaid enrollment, evaluations need to distinguish between populations that are CM eligible versus CM enrolled versus CM engaged. An "intention to treat" analysis, which is discussed in Chapter 3, is one way States can address these differences.

- A number of methodological issues can arise, including the use of a control group, selection bias, and "regression to the mean," all of which are discussed in Chapter 3. While control groups ideally should be used in all evaluations in order to isolate program effects, appropriate control groups may be challenging to identify and evaluate.

- If clear and specific standards for data collection and sharing have not been incorporated into a vendor-run program, both acquiring and compiling data can be difficult.

Considerations

A comprehensive evaluation may attempt to address all of the above questions and more. You will need to carefully manage expectations, and you may find that you need to narrow the scope of your evaluation questions in order to respond to specific and time-sensitive program management, budgetary, or legislative pressures. In addition, you may need to focus your list of evaluation questions in light of other key considerations like data availability, timeframe, and funding as discussed below. With limited resources, it will be advisable to develop a hierarchy of evaluation questions, distinguishing between process and outcome questions, and organizing them according to other key considerations such as feasibility, relevance, and cost.

Some questions may be relevant within a few months of program implementation; others may not be answerable early on. It will be important to distinguish these questions and ask different questions at different points along the pathway. For example, for some conditions cost savings may take time to develop, so you may need to identify intermediate outcomes that may be suggestive of future cost savings. In addition, different diseases have different natural histories and can be affected by seasonal changes, and therefore, they will have different timeframes for showing clinical and financial results. States should seek clinical guidance on when evaluation questions are relevant for particular conditions.

Action Step 2: Identify Which Data You Need and Which Data Are Available

States generally want to access a range of data sources to look at how care is provided, health outcomes, satisfaction, and financial measures. For successful evaluation efforts, the goal is to efficiently use a portfolio of measures. For example, you may find that combining certain qualitative and quantitative data sources, rather than a broad array of sources, is an efficient way to provide a comprehensive picture of your CM program.

Table 1 summarizes which types of data are generally used with which types of measures and identifies a few advantages and challenges for each data source. You will need to carefully specify your data needs based on the specific intervention you are measuring. Keep in mind that there are tradeoffs associated with different data choices.

You may not need data on every enrollee to estimate the effects of your CM program. Carefully selected random samples of adequate size may be sufficient for many measures. Drawing adequate samples of Medicaid enrollees, however, can sometimes be challenging. The pool from which you want to draw your sample may be too small as a result of a variety of factors, i.e., attrition, loss to followup, and difficulty contacting enrollees.

Table 1: Potential data sources and types of measures

Possible data sources	Advantages	Challenges	Measures
Administrative claims data	Data available for full population, standard reporting formats used	Coding practices and robustness of data can change over time (*e.g.*, data may become more complete and reliable as claims submissions improve). In addition, certain clinical conditions or events may be underreported or underdiagnosed. Inconsistencies in reliability and validity, missing data.	Quality: Process Quality: Outcome Utilization Cost
Medical records	Rich clinical data that are not available through administrative data; electronic medical records offer an opportunity to collect more clinical data at low cost	May be expensive and labor-intensive to collect, reporting formats may vary, care must be taken to remove patient identifiers	Quality: Process Quality: Outcome
CM administrative records	Data on CM interventions, flexibility to modify to meet reporting and evaluation needs	Limited to interventions for which care manager has records	Quality: Process Utilization
Surveys	Primary means of capturing satisfaction data, validated tools such as Consumer Assessment of Healthcare Providers and Systems (CAHPS)[15] available, can do random sampling	Must ensure sample and response rate are sufficient. Potential for response bias.	Satisfaction

Data Availability

An extremely important consideration in the early stages of designing an evaluation is which types of data sources are available. Medicaid programs generally rely on administrative claims data for fee-for-service program monitoring and evaluation. Although claims data are generally the easiest data to access, there can be significant variability and time lags before such data are available. In addition, some aspects of care may not be easily captured in administrative records. For example, some services like immunizations may not generate a claim.

If it is necessary to collect new data for the evaluation, planning for that should begin early. You may find that other sources of data greatly enhance your evaluation, such as medical records,

other administrative records (e.g., client self-reported data, clinical case management reports, and financial reports, etc.), and satisfaction surveys.

If you have a vendor-run CM program, it is especially important to negotiate access to data for your evaluation, as well as the details of the measures and methods involved in the evaluation. You will need to incorporate details about how and with what frequency you exchange data within your vendor contract, and you may want to consider including financial penalties for noncompliance. You should also arrange for independent validation and verification of any vendor-reported results.

Action Step 3: Establish an Evaluation Timeframe

Timeframe Covered by the Evaluation

You will need to address the question of what time period to cover in your evaluation. It can be challenging to determine whether your program is sufficiently mature for evaluation and whether the outcomes you are looking for can be measured with reliable data.[16] States should be aware that evaluations based on short followup periods (1 to 2 years) are potentially misleading and may overstate costs or miss long-term benefits.[17] At a minimum, you should allow at least 3 years of program experience. The most robust evaluations are often ongoing, iterative, and capture data at various points in time, taking into account factors other than CM that may be responsible for the outcomes.[18]

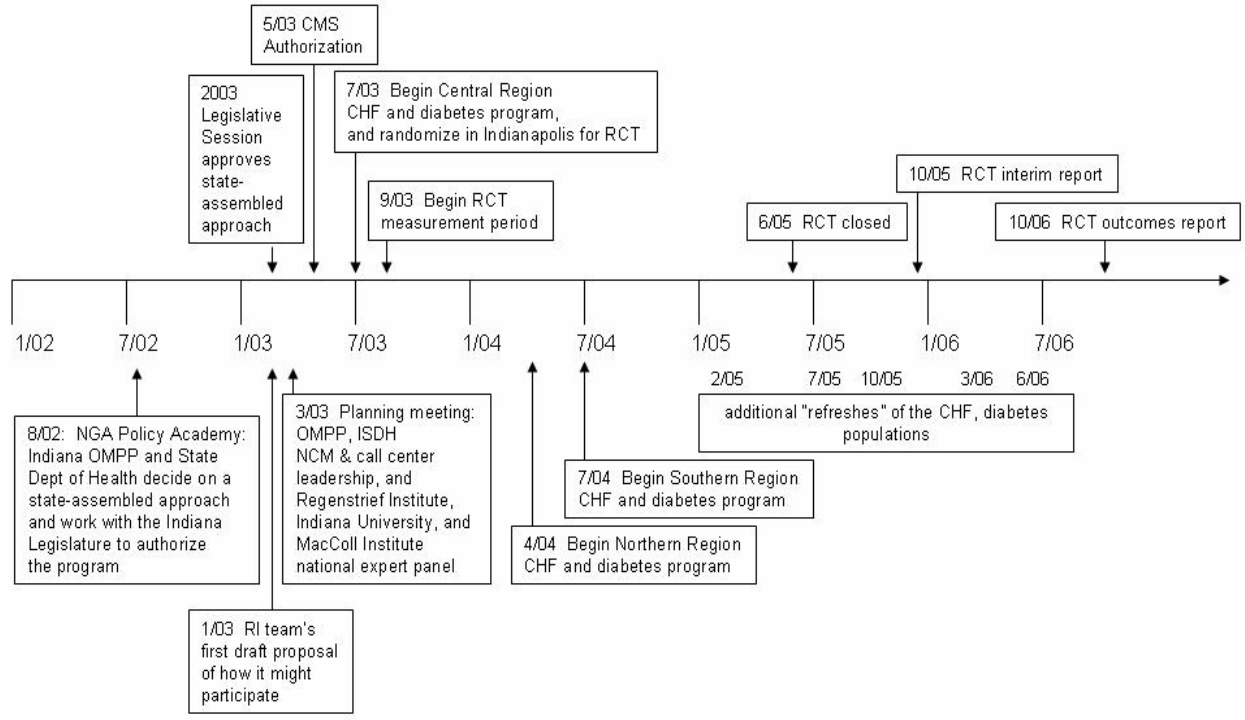
CM program design (e.g., pilots, phased-in enrollment, or voluntary enrollment [opt-in/opt-out]) influences the evaluation methods you use and the time period that should be covered by the evaluation. For example, if there was a start-up period during which individuals gradually enrolled in your CM program, the evaluation timeframe will need to take into account any lag times that may have occurred during enrollments and adjust for them appropriately. In addition, if you are doing a pre-post comparison, any policy and program changes over time will influence how you define your baseline period. Also, if the program is initially piloted in one region or enrollment is phased-in, identifying a comparable reference group of eligible people who did not enroll may make evaluation easier. This is discussed in more detail in Chapter 3.

Lessons from the Field

The University of Washington conducted an early clinical evaluation of Washington's CM program, which was phased-in between April 2002 and July 2002. The evaluation period was defined as July 2002 through October 2003. The evaluators acknowledged that a longer evaluation period could have been beneficial and that the full effects of CM may not have been evident in the 16-month timeframe.[12,19] Washington's actuarial evaluation of cost savings was conducted after the first year of CM, and it has been conducted in each subsequent year (allowing for 6 months of data run-out). The actuarial study compares a 1-year pre-implementation baseline to a 1-year post-implementation period.

Because of resource availability, new Medicaid programs and policies are often rolled out in stages by aid category or geographic region. Indiana took advantage of such a natural opportunity and used a randomized control trial (RCT) in two large, urban group practices and an observational analysis of staggered implementation with repeated measures in the central, northern, and southern regions of the State. Figure 1 illustrates how Indiana rolled out its program in relation to the RCT.

Figure 1. Indiana's evaluation timeline



Source: Rosenman M. Regenstrief Institute, Inc., Indianapolis, IN; November 2006. Used with permission.

In another example, Virginia began operating a voluntary (opt-in) CM pilot program in June 2004 and used it as a model for their new, voluntary (opt-in) CM program that was implemented in January 2006. In order to use a pre-CM reference group to evaluate their new program, Virginia is considering using a pre-June 2004 pilot baseline. In addition, the State may convert their new CM program to an enrollment opt-out arrangement in the future, which will influence their decisions about an evaluation design.[20]

Timeline for Conducting the Evaluation

It is very important that you try to anticipate potential future evaluation needs early on and develop a realistic evaluation timeline. If you are operating your CM program under a waiver, CMS has certain evaluation requirements. Programs authorized under 1915(b) waivers must be

independently assessed for the first two waiver periods (2 years each). If your program is vendor-run, you should identify clear timeframes for your vendor's major evaluation efforts, while also building in some flexibility in your vendor contracts to respond to unanticipated evaluation needs.

Your timeline will need to account for data completion time: many States need to allow at least 6 to 8 months for claims data run-out. Texas, for example, set deadlines for claims data extractions that were 7 months following the end of the period being evaluated. You may want to use the time during which you are waiting for data completion to set up your evaluation and test-run your measures.

Considerations: Ongoing Program Monitoring

It is also important to plan for ongoing program monitoring to identify opportunities for continuous quality improvement. You will want to regularly ask whether your CM program is accomplishing what you had hoped. While it makes sense to conduct an independent and thorough program evaluation every few years, you should also plan to continually monitor your program by evaluating performance based on a subset of your measures on a monthly, quarterly, semi-annual, or annual basis. Measures that you might use for ongoing program monitoring include, for example, physician visits, pharmacy utilization, hospital admissions, emergency department (ED) use, and HbA1c and retinal screens for patients with diabetes. You will want to consider how and when you report out the findings of your ongoing program monitoring efforts. For example, North Carolina provides quarterly feedback to providers based on their analysis of administrative data and annual feedback based on chart reviews.[21]

Action Step 4: Determine How Much Funding Is Available

Funding plays a critical role in defining how much a State can invest in its evaluation efforts. Some States elect not to budget a specific amount for CM evaluation activities. Rather, their evaluation activities may be subsumed within a variety of contracts and general administration budget line items or may not be budgeted at all. Other States build evaluation costs into their vendor contracts. Regardless of the model, a comprehensive evaluation will require resources, both human and financial, and should be planned for accordingly.

Human Resources

First, you should ensure that you have qualified staff available to focus on the evaluation. The number and type of staff members needed on an evaluation team will vary depending on the scope of the evaluation and whether you decide to conduct parts of the evaluation in-house, contract it out, build components into your CM vendor contract, or use some mixture of these arrangements. In fact, many States do include a mix of in-house and outsourced components of their evaluation.

In-House Resources

At minimum, you are likely to need a senior-level, in-house staffer who can devote a portion of his or her time to oversight of the evaluation project. Additional staff resources are generally preferable. Senior staff and program managers should play an active role in identifying the evaluation questions, determining which data and methods to use, and interpreting the findings, regardless of the evaluation arrangement you choose. Also, as a general rule, program managers should plan to delegate the data analysis and the oversight of many of the analytic details to qualified internal or external resources.

If any of the evaluation activities are conducted in-house, you will probably also need a program analyst or policy analyst to manage day-to-day evaluation tasks, a clinician to assist with the design of measures and interpretation of findings, and at least one data programmer/statistician to perform data analyses.

All members of the CM program staff, whether directly involved in the evaluation or not, should expect to devote some portion of their time to the evaluation and performance monitoring efforts. For example, the staff person who oversees the CM hotline may not be directly involved in the program evaluation but will likely contribute in some manner, e.g., by pulling hotline call records and/or being available for a key informant interview.

External Resources

If you use an outside evaluator (e.g. an actuarial firm, research group, or local university), it is best to involve the evaluator as early as possible in the program and evaluation design. As an example, Texas, which uses an actuarial firm for most of its CM evaluation activities, found that involving the actuary at the point of program design and implementation was critical.

States that have savings or performance guarantees may find that their evaluations are subjected to a high level of scrutiny, and outside help may be necessary to conduct a rigorous evaluation. Outside review of the proposed evaluation approach can be helpful even if parts of the evaluation are to be conducted in-house. In addition, a small amount of technical assistance from an expert can significantly enhance staff capabilities.

Suggestion for Vendor-Run Programs

If you are contracting with a CM vendor, it is advisable to clarify up front what your expectations of the vendor are with regard to ongoing performance monitoring and evaluation activities. Which portion of the evaluation is each party responsible for conducting? Remember to also build in some flexibility with your vendor for ad hoc data requests to meet any unanticipated evaluation needs.

You will also need to determine how vendor-reported data and evaluation findings will be validated—including verification of the credibility of the methods used to generate findings—and how much the validation activity will cost. Addressing these questions early will help both you and the vendor allocate the appropriate resources to meet your evaluation goals.

Lessons from the Field

In Pennsylvania, the State contracts out portions of its evaluation while also retaining some evaluation responsibilities in-house. The State contracts with an actuarial firm to conduct a cost-effectiveness analysis. Pennsylvania also uses a health policy analyst to do in-house quality and financial performance evaluation. In its CM vendor contract, the State set forth specific evaluation expectations and an evaluation timeframe. When the State prepared an independent program assessment for its 1915(b) waiver renewal, it used evaluation expertise from the State's Comptroller's office.[22]

In Vermont, the legislature passed the "Blueprint for Health" Chronic Care Initiative in 2005. This initiative includes Medicaid and all other Vermont health care payers. The "Blueprint for Health" is the State's plan for implementing a chronic care infrastructure that includes the prevention of chronic conditions and chronic CM programs. "Blueprint for Health" applies an integrated approach to patient self-management, community development, health care system and professional practice change, and information technology initiatives.[23]

Vermont issued separate requests for proposals (RFPs) for the evaluation, monitoring, and provision of services for its CM program. This approach divides up these functions to ensure that the specific goals of each part of CM are met, underscoring the importance of evaluation and performance monitoring. The RFPs also emphasize transparency and collaboration as core components of CM evaluation and implementation.

Other Resources

In addition to human resource costs, you will also incur costs associated with the execution of the evaluation and the dissemination of findings. These costs will be a function of the data sources and research methods you use. Costs will also depend on how much you choose to invest in engaging stakeholders and disseminating results.

Your investment in data resources will depend on the availability and quality of your data and the complexity of your methodologies. For example, if you already have high quality claims data available, your costs may be limited to the expenses of scrubbing and formatting the data. If your data are less reliable, you may need to make greater investments in your data systems to support the analyses. You may also find that acquiring supplemental data through medical record reviews or primary data collection (e.g., surveys, or focus groups) can be extremely informative but is relatively costly and labor intensive. However, carefully selecting random samples of adequate size may help you maximize your investment.

Innovative Data Collection Strategies

Several States have expanded their portfolio of measures through innovative data collection strategies and strategic partnerships. In North Carolina, the Medicaid program contracted with Area Health Education Centers (AHECs) to collect medical record data in a way that is less expensive. AHECs are located in nine regions of North Carolina and provide educational programs and information services to the health care workforce.[24] A team of six auditors on staff with the AHECs conducts independent medical record reviews on random, practice-specific samples. The Medicaid program reimburses the AHECs on a per-chart basis ($20 per chart plus a 5 percent administrative fee). In 2006, the State spent almost $400,000 to review approximately 9,000 charts for patients with asthma and 8,500 charts for patients with diabetes.

In Indiana, the State partnered with the Regenstrief Institute to access electronic medical records (EMRs) to enhance its capabilities for clinical measurement.[25] In an urban, multi-site group practice that serves a large number of patients with Medicaid coverage, the presence of an EMR system gave the team the opportunity to use both clinical data and Medicaid claims data for program support and evaluation. EMR data can help in analyses of baseline characteristics of potential CM participants and in sample size calculations for a controlled trial. Clinical data can also help validate population selection algorithms that are based on Medicaid administrative data, and they can provide information on some variables (such as HbA1C levels) that are not recorded in Medicaid claims.

Washington Medicaid uses a client services database maintained by the Research and Data Analysis Division at the Washington Department of Social and Health Services to do ad-hoc analyses and drill down on particular evaluation questions. The database compiles Medicaid eligibility and claims data with mental health, long-term care, and developmental disabilities data to give a more complete picture of the enrollee experience. Washington, like a number of other States, chose to modify the Consumer Assessment of Healthcare Providers and Systems (CAHPS™), rather than creating a new survey, to collect client satisfaction data for their CM population.

Some States have partnered with their External Quality Review Organization (EQRO) or Quality Improvement Organization (QIO) to collect hospital records. Virginia is considering ways to offer incentives to providers to encourage their involvement in clinical data collection.[a]

Lessons from the Field

The experience of Texas Medicaid in refining their data system exemplifies the need to factor in expenses that can result from data collection and evaluation. Texas staff spent hours creating a system to generate data and exchange information between the State's system and the contractor's system. Because Texas Medicaid was involved in major information technology (IT) system upgrades at the same time they were also designing their CM program, staff found that they needed a senior-level program champion on their operations/IT staff to ensure their specific CM data and evaluation needs were prioritized. CM staff met with the IT team (including the champion and at least one IT programming specialist) weekly as they were developing their data system. They found that long-term continuity in core staff involvement was critical to the success of their program.

Action Step 5: Select Evaluators

A good program evaluator should have strong critical thinking and be skilled in analysis. The evaluator should have a thorough understanding of how your CM program operates and should be familiar with the attributes and limitations of your data. The evaluator(s) for your CM program may be independent of your program or come from within your organization. Some States use a mix of outside and in-house evaluators.

Deciding who will conduct your evaluation depends on several factors, including whether you have a vendor-run CM program, what your evaluation questions are, and what types of analytic methods and data sources you use. For example, if your evaluation questions are related to the economic impact of your CM program, you may need a different type of evaluator than if your questions focus primarily on clinical impact.

States that have invested in building up in-house research and evaluation units may opt to conduct some portion of their evaluations in-house. Those that do not have staff focused on evaluation will typically look outside for evaluation expertise and use in-house staff primarily for ongoing program monitoring. States that have a data warehousing arrangement with a local university or other research partner may also opt to outsource portions of their evaluation. As a general rule, actuarial evaluations require States to solicit outside expertise, which likely will not come from a university.

[a] Note, in any initiatives using medical record data for evaluation purposes, extreme care must be taken to remove individual patient identifiers.

You should carefully consider your organizational capacity when making decisions about who will conduct your CM evaluation. Resource availability will certainly influence that decision. In-house evaluation efforts, like outsourced efforts, have strengths and weaknesses. For example:

- In-house evaluation activities provide an opportunity for building in-house expertise and capacity. However, in-house evaluation efforts, particularly if conducted by your CM vendor, may be criticized for a lack of objectivity. If you conduct portions of your evaluation in-house, you may want to consider having an external advisory committee made up of recognized program and evaluation experts to provide input on evaluation questions, research design, and interpretation of findings.

- An outsourced evaluation is generally viewed as more objective than an evaluation conducted in-house. However, independent evaluators may not be entirely free of conflicts of interest. No matter how much detail your contractor shares with you, there may still be a "black-box" element, which can be significant. Plan to conduct some form of independent validation and verification of your contractor's findings, which can be difficult unless there is transparency about the methods used.

It is also important to note that university partners or other collaborating organizations may not be perceived as completely independent if they are affiliated with your Medicaid program. If your local university or other organizational affiliate has been involved at any stage of the CM program development and management, there will be a natural allegiance and a desire to see the CM program proven effective. This may diminish the perceived objectivity of your evaluation. Addressing this perceived conflict upfront will promote both transparency and assist others in clear interpretations of the findings.[26]

Lessons from the Field

As Washington's CM program evolved over time, the State used four different evaluators: in-house staff, CM vendors, an external clinical evaluator, and an actuarial firm. Washington has continually used its in-house staff to do some performance monitoring and evaluation activities and its CM vendors to provide client-reported health status, access, and satisfaction data. The State used the University of Washington to conduct a clinical evaluation of the first year of the CM program and contracted with their EQRO to do the clinical evaluation thereafter. The State uses an actuarial company to examine CM cost savings. In the future, Washington's EQRO will synthesize findings from all the evaluators.

The involvement of several evaluators has allowed Washington Medicaid to take advantage of in-house evaluation capacity while also leveraging specific outside expertise from several contractors. Each evaluator provided a different perspective on the effectiveness of Washington's CM program, which in some cases had the benefit of providing an opportunity for comparisons and validation. While the use of multiple evaluators has led to variations in methodologies, definitions, and measures, the State has found that the greatest challenge of having multiple evaluators has been presenting the findings.[27] In particular, juxtaposing the vendor-reported results with findings from other evaluators has not always provided a clear story.

Stakeholder Involvement

States should also include key stakeholders (such as consumers, advocates, providers, legislators, vendors, and other interested parties) throughout the evaluation process, from the design stages to the interpretation of the findings. In particular, it will be important to gain stakeholder involvement in determining the appropriate outcomes to measure. If you design and execute your evaluations through a collaborative and transparent process, you will find that the evaluation gains credibility. In addition, States may find that they can use the evaluation as a tool for building stakeholder involvement in the CM program. Increased involvement, particularly within the provider community, may strengthen the overall CM program. It is important to note, however, that actively involving stakeholders may require a significant investment of time and energy on the part of program staff.

Lessons from the Field

North Carolina involved its provider community in every aspect of its CM program. The State found that the time spent creating a culture of cooperation has had a tremendous impact as the State builds the program. Providers, in consultation with the State, choose which performance measures are most appropriate for each condition. Program administrators find that their coordination with providers has made the program and its evaluation efforts more sustainable over time because providers are invested in the goals of the program and the outcomes of the evaluation. They also recognize that once a structure has been put in place to communicate with providers, the program is able to more easily troubleshoot programmatic and evaluation issues.

In Indiana, a rural health center emerged as a “pathfinder” or vanguard practice. Through close collaboration with this health center, the State and its evaluator, the Regenstrief Institute, identified certain characteristics that varied between practices. The partnership has helped the State drill-down in specific practice areas and test evaluation strategies.

Some States with vendor-run programs have used their vendors to establish provider advisory boards. Texas, for example, contractually requires its CM vendor to engage providers on an ongoing basis through quarterly meetings and regular communications. The provider advisory board has been extremely active throughout the history of the Texas CM program. The board plays an ongoing role in helping the State identify performance measures and recommend program modifications or policy changes.

Virginia has designed a new CM program that is very patient-centered. The State, with their CM vendor, has carefully created a process for engaging enrollees, family members and personal representatives. Consumer engagement will continue to be a priority as Virginia develops its evaluation plan.

Chapter 3. Evaluation

The choice of evaluation design has implications for many aspects of the evaluation and should be considered carefully.[28,29,30] The following paragraphs discuss the basic components of the evaluation.

Evaluation Action Steps

- ❑ 1. Select a reference group or groups.
- ❑ 2. Structure the evaluation.
- ❑ 3. Select analytic methods.
- ❑ 4. Identify and address potential confounding factors.
- ❑ 5. Select measures.
- ❑ 6. Identify and address data issues.
- ❑ 7. Consider sample size.

Action Step 1: Select a Reference Group or Groups

A reference (or control) group is an equivalent comparison group that was not subject to CM (the intervention). A reference group provides a basis for comparison between it and the intervention group (people receiving CM) and allows for an assessment of what the effect of a program has been. The use of an appropriate reference group is an essential part of a credible evaluation. Table 2 provides a description of three different types of reference groups.

How a reference group is selected will depend on the design of the CM program. For example, it is much harder to select a reference group when CM is applied to a select group of individuals who volunteer for CM (e.g., in an opt-in program). This is true because people who volunteer may be different in important ways from people who do not volunteer.

Another factor that should be considered has to do with the availability for analysis of certain types of data, which may be influenced by the type of reference group selected.

Reference groups can be selected purposefully before the CM program is implemented (prospective controls), or the evaluation may have to look back and try to find an appropriate reference group after the program is implemented (retrospective controls). The reference group needs to be very similar or equivalent to the intervention group. This reference group can either be pulled from the potential CM population, or a separate group can be identified that is as similar to the CM group as possible. In the absence of an independent reference group, a pre/post analysis can be used that compares the impact of CM on a group with the time prior to CM, if the before period, when projected forward, is an accurate prediction of what would have happened in the absence of CM.[31]

Table 2. Types of reference groups

Reference groups	Description	Advantages	Challenges
Randomized control group	Participants randomly assigned to treatment and reference groups	Gold standard; most rigorous	May pose ethical and/or political concerns; does not protect against the impact of changes in provider practice
Staged implementation	Program rolled out in certain areas before others	May be more feasible than a randomized control group; some differences can be controlled for through statistical analyses	Must ensure the population groups in the roll-out areas are similar to the groups in other areas, and that certain policies do not affect one area differently from another
Matched control	Reference group selected that is as similar as possible to treatment group	May be more feasible than a randomized control group; some differences can be controlled for through statistical analyses	Must consider all factors that may affect the outcome independent of the intervention

Although a reference group is essential for evaluating the impact of the CM program, it may not be necessary for program monitoring. Because of the expense and difficulty of maintaining a reference group for long periods of time, before/after comparisons may be better suited to assess process measures related directly to the intervention as part of program monitoring; however these comparisons may potentially be misleading due to regression to the mean, selection bias, and other factors discussed below.[32]

Randomized Controlled Trials

The randomized controlled trial (RCT) is considered the gold standard of evaluation designs, and it provides the most definitive results. However, an RCT may not be feasible or practical in all instances.

In an RCT, participants are randomly assigned to treatment and reference or control groups. The random assignment helps decrease the possibility that the treatment and reference groups are systematically different or nonequivalent. Since the treatment and reference groups are pulled at random from the same population, any differences in outcomes between the groups are assumed to be due to differences in the receipt of the intervention (in this case, CM). In addition, since the treatment and reference groups are operating concurrently, the design protects against external factors or evolution of treatments that translate into differences in treatment patterns over time (i.e., these "co-interventions" become available to both treatment and reference groups equally).

A randomized control group can be used in voluntary CM programs, if people are randomized to CM and reference groups after they volunteer. This eliminates any selection bias caused by differences between volunteers and those who don't volunteer for CM, but this method is not without political risk.

For programs that are designed to change physician behavior, RCTs do not automatically protect against the influence on care that may occur as physicians and other providers learn from the intervention and begin to offer a different (higher) level of care, not only to participants in the treatment group, but also to control patients. This problem can be reduced by assigning participants to the treatment and reference groups in groups, based on the identity of their primary care physician. Subtleties in the design and evaluation of RCTs are typically unfamiliar to Medicaid personnel, so consultation by regional (often university-based) experts is important to consider prior to deciding on a particular study design.

As previously noted, many Medicaid programs have concerns that the RCT may pose ethical or political concerns resulting from some of the potentially eligible population being allowed to receive the treatment sooner than others. (The reference group may have volunteered, but will not receive the CM services.) For many Medicaid programs, this will require a CMS waiver of State-wideness that will allow them to partially implement a program to a subset of eligible people. It will also require the political will to ask for participation in a program where only half of the subjects will receive the intervention—although it is important to remember that the effectiveness of CM is still unproven, and the potential ethical concerns are not the same as withholding a service with established efficacy.

Conversely, a decision to use a less rigorous evaluation approach could be considered more of an ethical dilemma if it leads to inaccurate conclusions that a program is saving Medicaid dollars when, in fact, it increased spending and had only modest effects on quality of care and the health of members. If such a finding diminishes enthusiasm to look for opportunities to improve CM further, it could be a detriment to the future health of the State. Moreover, controls in an RCT are typically offered the current standard of care delivery (i.e., no ongoing care is withheld) until some future point in time after which the State will plan to offer CM to all eligible and interested members, should it prove effective in enhancing care. On a more practical level, some programs have managed to solve this problem by explaining to potential participants that there is insufficient capacity to take all interested parties at once, and the program will be phased-in over time.

Lessons from the Field

Indiana Medicaid chose to use an RCT in conjunction with an observational analysis of staggered implementation with repeated measures. Most CM evaluations have used observational designs in which the evaluators have no control over who receives the CM intervention and who does not. An observational design was used in the central, northern, and southern regions of the State. The RCT was used in two large urban group practices where enrollees' start dates were randomly staggered according to clinic sites. The RCT was intended to help identify and measure potential biases that might have impacted results obtained from the observational design.

Alternatives to RCTs

Even when an RCT is not practical, there are other effective means (e.g., staged implementation and matched controls) of including reference groups that, if appropriately designed, can strengthen your evaluation and be more feasible to implement. Although these alternatives are more subject to selection bias and other limitations than RCTs, they still can be helpful in identifying and isolating program effects and offer advantages over actuarial adjustment alone.

In a staged implementation, as in Indiana, where a program rolls out in some communities before it goes Statewide, it may be possible to use individuals as a reference group who may be eligible for the CM program but are located in an area where it is not yet available. Ideally, there will be no important differences in this reference population and the population receiving CM services other than the geographic area. However, it will be necessary to look at prior year trends to make sure the two groups were similar before the CM program began. You also will need to carefully consider whether there are any in effect during the implementation period that could impact costs and quality differently for the reference group. Some differences can be controlled for via statistical analysis.

In other cases, it may be necessary to identify and select a separate reference group. A matched control is selected to be similar to the group receiving CM services in as many ways as possible. Matched controls may be selected to be similar on the basis of demographic characteristics (age, sex, socioeconomic status) and disease state. In particular, factors such as use of health care services in the prior year and health habits are important to consider. In general, all factors that may affect outcomes independent of CM should be considered in selecting the reference group. As in a staged implementation, if the sample cannot be matched on all characteristics, some differences can be controlled for in a statistical analysis. However, the less reliance there is on statistical modeling of this type, the more compelling and robust your estimates will be.

Lessons from the Field

North Carolina Medicaid, with the Cecil G. Sheps Center for Health Services Research at the University of North Carolina, conducted an evaluation using a matched control. The study compared the costs and utilization of Medicaid enrollees with asthma and diabetes in the CM program to enrollees with the same conditions in ACCESS, the State's traditional PCCM program. Because there were significant differences in the ages of enrollees with asthma between the two programs, the evaluators used age-adjustment throughout.[33]

Action Step 2: Structure the Evaluation

There are two main structures you may consider when designing your evaluation: cross-sectional evaluations and longitudinal evaluations.

Cross-Sectional Evaluation

A cross-sectional evaluation is done at a single point in time, presumably after the CM program has been implemented. In order to make the case that the CM program had an impact, it is necessary to compare those that received CM with a group of people that did not receive CM—i.e., the reference group. Presumably, the reference group will be as similar as possible to the CM (or intervention) group. It is often a challenge in cross-sectional evaluation designs to make sure that the reference group is comparable, and analyses often need to statistically control for potential sources of differences.

Your evaluator may look to how programs were implemented—if implemented partially (in only one region of the State) or in a staged manner, the evaluator may have the convenience of a ready made reference group, provided that beneficiaries in the program and reference group had similar prior utilization and cost patterns. In fact, whenever claims data are available for the period of time prior to the implementation of CM, it is advisable to test out the analytic approach and model using these data. You may want to raise this idea with evaluators to determine if it is feasible, as it will strengthen confidence in the final analytic results.

Another challenge in cross-sectional evaluation designs is to make sure that the data to be used in the evaluation are comparable between the CM and reference groups. Since program data are often used as part of the evaluation, obtaining comparable data on the reference group may be a challenge. Remember to take care to ensure that differences in findings are not due to data differences. This is a critical, and difficult to address, issue.

Lessons from the Field

The North Carolina evaluation conducted by the Sheps Center at the University of North Carolina used a cross-sectional evaluation design. The evaluation compared enrollees with asthma and diabetes receiving CM to a similar group of enrollees in the State's PCCM program who were not receiving CM. The two groups were compared during a single point in time (2000-2002). The evaluators underscored the importance of adjusting for differences in the enrolled populations.

Regression to the Mean

The phenomenon known as "regression to the mean" is a particular challenge to comparability in pre-post designs (before/after comparisons) when the criterion for eligibility for CM is high medical costs. Since a group of people who all have high medical costs in one year will tend to have average costs that are considerably lower in the following year, it is often a challenge to separate out differences that may have been caused by the CM program from other factors—such as high quality health care in the community, the natural history of the disease, or random fluctuation between one year and another— that together are often called "regression to the mean." The use of an equivalent reference group can help you separate regression to the mean from true program effects.

Longitudinal Evaluations

Longitudinal evaluations evaluate differences before and after implementation of a CM program. The evaluation may include data

from only one point in time before the implementation and one point after implementation, or it may follow participants through several stages post-implementation.

Longitudinal evaluations also look to compare the CM group with a reference group. In a simple pre-post design, the sample is used as its own reference group, and characteristics of the sample before and after are compared. While a pre-post design ensures similarity between the treatment and reference groups, there are potential confounding factors that must be considered. For example, if the standards of care have changed since the CM program was implemented, it will be important to separate out changes that may be due to CM from changes that may be due to changing standards in care (such as new guidelines or the introduction of new drugs or treatments), or simply the aging of the population.

Lessons from the Field

The Disease Management Association of America (DMAA) evaluation guidelines reinforce the importance of transparency in evaluation methodologies. CM program evaluators should be able to clearly explain not only the methods used but also the impact these methods have on the interpretation of results. For example, including a discussion of the limitations of a pre-post design will help others to interpret the findings with regard to both the strengths and weaknesses of the study. The DMAA acknowledges the challenge of striking a balance between rigorous methods and a feasible evaluation design. Acknowledging how you have dealt with this challenge will help in understanding the evaluation results.[34]

Washington and Pennsylvania used a longitudinal design (pre-post analysis) to evaluate cost savings in their CM programs. They compared the costs associated with the population targeted for CM during the measurement year to a baseline reference group of individuals who met the criteria for CM in the year prior to implementation.[12,22]

In a cohort design, both reference and intervention groups are followed over time. Any differences between the two groups are presumed to be a result of CM, since both groups may be subject to the same environmental pressures, such as changes in the standards of care over time and the possibility of regression to the mean. Also, inherent differences between the two groups that persist can be eliminated by comparing changes rather than absolute levels. Table 3 provides a comparison of different evaluation design options.

Table 3. Comparing evaluation designs

Design	Reference group	Bias	Confounding	Validity	Ability to generalize
RCT	Randomly selected "eligibles"	Very low	Very low	Very high	Low to high
Quasi-RCT	Nonrandomly selected "eligibles"	Low	Low	High	Low to high
Cohort	Naturally excluded "chronics"	Moderate	Moderate	Moderate	High
	Naturally excluded "nonchronics"	High	High	Low to moderate	High
Pre-Post	Intervention group in an earlier time period	Low	Moderate	Moderate	Moderate
Actuarial	Predicted cost trends of "nonchronics"	Very high	Very high	Low	High

Source: Adapted from a presentation by Ackerman RT, to the AHRQ Learning Workshop, Nov 2, 2006.

Action Step 3: Select Analytic Methods

The analytic methods used to conduct an evaluation are also important and should be considered up front. Most economic evaluations use statistical methods to assess differences between CM and reference groups on measures of interest. These statistical methods estimate the costs of individuals receiving CM compared with individuals in the reference group, controlling for other factors that are considered important to the outcome. An alternative is to use actuarial methods, which project expenditures for groups of individuals adjusting for factors that might impact cost trends.

The evaluation strategy should specify whether total costs or disease-specific costs are being addressed. Because of the challenges associated with parsing out health care costs associated with a particular disease, and because of a high prevalence of comorbidities in the CM population, most evaluations examine total costs.

When estimating effects on costs, the unit of analysis should be the individual's cost per month observed throughout the year ("per member per month" analysis). The cost per month should be weighted by the proportion of the year they are observed (i.e., enrolled in Medicaid). This will allow the evaluator to adjust for differences in the CM and reference groups in the proportion of people who leave the program during the year, including those who die or are otherwise lost to observation. This is especially important in a Medicaid population, where people move in and out of Medicaid eligibility. If there appear to be differences in mortality between the CM and reference groups, this may suggest that the two groups are not very comparable. It is highly unlikely that CM would have a significant effect on mortality within a year or two. In addition,

since people who die are much more expensive during their last year of life, a difference in mortality between the two groups could lead to significant differences in costs.

Another issue to consider is how outliers will be handled. Outliers are extremely expensive cases that have the potential to skew the results; in some cases it may be reasonable to truncate the expenditures for outliers. However, outliers do impact the total costs to Medicaid, and it is possible that a CM program could impact the number of outliers. For example, if CM is expected to reduce the number of outlier cases by rationalizing care, then truncating outliers may mask the potential impact of CM.

Action Step 4: Identify and Address Potential Confounding

In spite of the best efforts to select reference groups that are similar, sometimes there are confounding factors that may impact the outcomes. Confounding factors can be differences between the CM group and the reference group, environmental factors, or other obstacles. It is important to select a reference group that minimizes the potential for confounding and to identify any potential confounding factors and control for them in statistical analyses whenever possible. Your statistical analyses could include risk stratification, where the CM and reference groups are further divided according to disease status, and separate comparisons are made for people in different disease states. Other multivariate statistical methods, such as multivariate regression analysis, could also be used to control for the effect of differences in population characteristics or to estimate the impact of environmental factors that could affect outcomes. Remember, most efforts to control for potential confounders rely on the availability of data about these other factors. A careful evaluation should consider its limitations, which will include the extent to which possible unmeasured confounders may have impacted the results.

Action Step 5: Select Measures

The mechanism by which CM achieves its effects is thought to be two-fold: either by improving or rationalizing the use of health care services or by reducing the likelihood of adverse events or preventing further decline in health (and thus reducing the need for additional health care services). Your evaluation should consider both mechanisms, and measures should be chosen that link the goals and objectives of CM to potential outcomes. For example, quality measures should align with quality objectives and include both intermediate and long term impacts and financial measures with fiscal objectives.

Measures should also have the potential for change in the timeframe selected for the evaluation. Another important consideration in selecting measures is the availability of data to support the measures. The choice of evaluation design may influence the availability of data. (See Action Step 6.) Finally, feasibility is an important consideration. You should select measures that can be calculated with existing data and have demonstrated reliability and validity. In addition, measures that have been used in other studies and other populations, for which benchmarks exist, can add credibility to evaluation findings. Appendix 1 summarizes examples of different types of measures that you may want to consider.

Quality Measures

Quality measures can include measures of access, outcomes, patient experience (satisfaction), processes of care, and/or the structure of the care environment. In evaluating CM programs, it is best to include a mix of measures as there may be many factors other than CM that impact outcomes. In terms of timeline, access, structure, and process measures may be easier to detect in the shorter term and may be easier to link back to the CM program.

AHRQ maintains a Web site with links to a large number of quality measures, many of which are appropriate for assessing the impact of CM (visit AHRQ's Quality Measures Clearinghouse at http://www.qualitymeasures.ahrq.gov/). In addition, CMS has recently released *The Guide to Quality Measures: A Compendium*.[35] Some of these measures are readily available from administrative data. However, many involve new data collection or review of medical records, which can be expensive. It may be possible, however, to collect these data on a random subsample and obtain results that are likely to be very similar to those for the full population, resulting in substantial savings. Since new data collection is so expensive, it is unlikely that baseline or prior history will be affordable or feasible.

Financial and Administrative Measures

Another set of goals for CM is the rationalization or reduction of expensive services and the reduction in health care spending. Appropriate measures include those of use (numbers of hospitalizations and lengths of stay, numbers of emergency room visits, numbers of physician visits) and expenditures for care. Many of these measures are readily available from claims data and other administrative databases. However, administrative and other claims data, like any data, are subject to issues regarding data reliability, quality, and completeness. Administrative data may not reliably capture the information of interest or may capture that information only for subsets of the population. In addition, data may not be available on a timely basis. Claims data in particular are subject to time lags, and data for more complicated care are often subject to a greater time lag in availability.

In assessing the financial impact of CM, it is important to identify financial expenditures for one-time program start-up costs, ongoing administrative costs, and medical costs. All of these costs are legitimate financial expenditures associated with CM. However, the decision to include or exclude some or all of these costs may vary depending on the questions being addressed. An alternative way to consider costs is to separate out fixed costs from variable costs. While fixed costs must be allocated across program participants, it may also be possible to allocate them across the expected life of the program (rather than in a single year) to more realistically distribute these costs.

In Indiana's evaluation, the State distinguished between one-time start-up costs (e.g., office equipment) and ongoing operational costs. The State divided ongoing operational costs into those that were affected by patient volume (variable costs like nurses' salaries and benefits) and those that do not change with patient volume (fixed costs like insurance). This allowed the State to frame results for policymakers that included traditional estimates based on total expenditures, as well as estimates that excluded one-time start-up costs, in the event that the State might

perceive past one-time expenditures as "sunk" and, therefore, less germane to a decision about continuing to fund the program. Moreover, categorizing ongoing costs as fixed or variable allowed the State to project the impact of hypothetical changes in member volume (i.e. expanded reach of the program beyond the ramp-up period) on future cost-effectiveness estimates.

You also may identify measures to evaluate the process of implementing and operating CM programs. These program process measures may include number of clients per care manager or number of contacts per patient.

Lessons from the Field

States often begin their selection of measures by looking to nationally accepted measures and metrics that other States have used. Texas identified a group of core measures and associated performance corridors. In particular, they identified a group of measures from which they felt they could measure cost-savings. They also developed a supplemental list of measures they believed would be good for additional benchmarking and program monitoring.

North Carolina partnered with a mini-collaborative of clinicians to select their measures. They began by reviewing national clinical practice guidelines, particularly the National Institutes of Health (NIH) asthma guidelines[36] and the American Diabetes Association (ADA) clinical practice recommendations.[37] Once they had reviewed clinical guidelines and national measures, North Carolina used several important criteria to guide their measure selection process, namely:

- Identify measures associated with evidence-based best practices.
- Measure interventions that have a clinical impact.
- Choose measures for which data are available.
- Ensure measures are appropriate for the population (e.g., consider tailoring measures for a pediatric population, identify which continuous eligibility criteria are appropriate).
- Coordinate measure selection with measurement by other purchasers in the market.

Pennsylvania operates ACCESS Plus, an enhanced PCCM program with a CM component, in rural regions of the State, and HealthChoices, a mandatory managed program, in urban areas of the State. Pennsylvania officials chose a group of Health Plan Employer Data and Information Set (HEDIS) measures that they were already using in their HealthChoices program so that they could draw comparisons between the two programs.

Action Step 6: Identify and Address Data Issues

Identifying and obtaining reliable and valid data may be one of the most challenging aspects of evaluation and one of the reasons why it is so important to plan the evaluation early—while it is possible to identify and tailor data for the analyses. While administrative data sources are often preferred because of their low cost and availability, there still are challenges to consider.

Data Reliability

- Ensuring data reliability is important to the validity of the evaluation results. Once you have taken inventory of available data, you will need to examine whether the data are reliable. In particular, you will need to consider the following questions with respect to the data:
- Are the data complete?
- How much data run-out time is needed to ensure you have received a complete data set?
- How much "data scrubbing" is needed to clean up the data file(s)?
- Have the data been validated and/or reconciled?
- Do you need to merge certain data sets to give a complete picture?
- Are there artifacts in the data related to program modifications, policy changes, or data reporting anomalies?

Baseline data. In cases where CM evaluations are designed after the intervention has been implemented, obtaining adequate and comparable data for the time period before the intervention can be a challenge. In many cases, the intervention results in the capture of new data. However, the lack of these data in the pre-implementation phase can be problematic, since it will not be possible to determine whether changes occurred as a result of CM, unless an RCT design is being used. (In that case, any difference in outcomes between the two groups is assumed to be due to the intervention.)

Comparable data. Obtaining adequate and comparable data for the reference group can also be a challenge. For example, States that compare their CM population to a reference group in another delivery system (e.g., managed care) must ensure that coding practices in their fee-for-service claims data are comparable to coding in their managed care encounter data. Comparable data are critical; without comparability it will not be possible to attribute observed differences to the CM program.

> **Intention to treat analysis.** If you use a controlled design, it will be necessary to use the same measures to compare the reference group to the intervention group. For this reason, an intention to treat analysis should also be used when defining the comparison groups. Intention to treat means all target enrollees in the intervention period are included regardless of whether or not they received the complete intervention. The same criteria are also applied to the reference group. By using this analysis, you can also account for the extent to which the CM program was successful in attracting and retaining participants.

Action Step 7: Consider Sample Size

It is important to conduct an evaluation that has the potential to provide credible evidence of result. An important factor in assuring credibility is having an adequate sample size with the power to detect statistically valid differences that result from CM. The sample size needed in an evaluation is directly related to a number of the design factors that have been discussed earlier. The outcome measures chosen and the expected differences both within and between the CM and

reference groups will in large part determine the sample size needed for the evaluation; a power analysis will provide estimates of the sample sizes needed to obtain results.[b]

The choice of research design and statistical methods used will also determine sample size. Longitudinal designs will frequently require larger sample sizes, since some of the sample will be lost to attrition over time. Complex statistical methods, such as risk stratification, will also require larger sample sizes.

Drawing adequate samples of Medicaid enrollees can be challenging. You may find that the pool from which you want to draw your sample may be too small as a result of a variety of factors, i.e., attrition, loss to follow-up and difficulty contacting enrollees.

[b] A power analysis is used to identify the necessary sample size to see a statistically significant result, given estimates about differences between the experimental and control groups and the variance in responses.

Chapter 4. Presenting Your Findings

Presenting Your Findings Action Steps

- ❒ 1. Analyze your audience and your objective.
- ❒ 2. Develop a dissemination strategy.
- ❒ 3. Make the case.
- ❒ 4. Translate your data.

Action Step 1: Analyze Your Audience and Your Objective

When developing a plan to present evaluation findings, consider several points that will shape both the content and format of your results. Assess the nature of your audience (see Table 4) and the intended outcome(s) from dissemination of the evaluation data (e.g. to secure funding for the program, increase public knowledge of its cost savings, recruit participants, satisfy mandated reporting requirements, etc.).

Managing expectations. It is critical to manage your audience's expectations of your program. Releasing preliminary evaluation results help create an early sense of what your CM program can and will achieve within a certain time frame. It is not uncommon for legislatures to expect to see total cost savings within a period of 12 months. These false expectations can have serious ramifications if a program fails to meet these goals.

Before deciding on the format for the presentation or report, you should think through the following questions:

- Who is the audience for this report/presentation?
- What does the audience want to do with the findings?
- What do you want your audience to do with the findings? (Do you need to inform or persuade?)
- Is there a gap between what the audience wants to do with the data and what you would like them to do with the results? If so, how will you reconcile this difference?
- What level of depth is appropriate for the audience's technical knowledge?
- What should your presentation strategy include in order to satisfy your audience and reach your objectives?

Table 4. Potential audiences for evaluation results

Audience	Their needs & objectives	Technical knowledge	Potential audience size	Frequency	Useful media
Program managers	• Be fully informed of program's costs/benefits • Consider necessary changes or improvements to program	High: requires little additional explanation	Limited	Annually, semi-annually, or quarterly, depending on contract requirements or legislative mandate	PDF booklet or printed report
Legislatures/ State officials	• Be fully informed of program's costs/benefits • Consider necessary policy or budget changes to program	Low to medium: Will require background information	Limited	Annually, semi-annually, or quarterly-depending on contract requirements or legislative mandate	PDF booklet, printed report, Presentations, or briefings
Other States	Share information to assist with program design, implementation, and evaluation of other States' CM programs	• High: If among those that already have a CM program • Low if designing or contemplating CM program	Limited	Depends on requests for information	Booklet, presentation, static or dynamic Web site
Media	Inform the public	Low	Large	Depends on local market	Press releases, one-on-one briefings, reports
Potential enrollees	To gain knowledge of program's benefits	Low	Large	Depends on program's goals	Booklet, presentation, static or dynamic Web site
General public	Be informed of program's costs/benefits	Low	Large	Depends on program's goals	Press releases, simple & friendly Web site
Participating or potential participating providers	Be informed of the program's costs/benefits related to patient care	Depends on provider involvement	Large	Depends on program's goals	Booklet, presentation, static or dynamic Web site
Research institutions	Use evaluation information for larger studies and meta analysis	High	Limited	Investigator Initiated	Raw data, dynamic Web site, and PDF report

Note: Potential audience size varies. In this table, "limited" is approximately 10-100, and "large" may be measured in thousands or more.

Action Step 2: Develop a Dissemination Strategy

To begin, ask "What options do you have to share this information?" Table 5 lists several formats to consider when presenting evaluation findings. The advantages and disadvantages of each format are outlined below.

Table 5. Possible dissemination formats

Format	Advantages	Disadvantages
Printed report	Low cost	Accessibility is limited to people you send it to, or who know you
PDF booklet	Low cost, zero marginal cost, can be posted on Web sites and distributed electronically	Dense, not interactive
Presentations	Gets the point to the right people, meets reporting requirements, cost per presentation is low	Multiple presentations are cost prohibitive, information is one-time only (not continually accessible)
Static Web site	Accessible to a wider audience including unintended audiences, able to present information in varying levels of depth	Costly, hard to update
Dynamic (database-based) Web site	Allows researchers to focus on/ compile portions of data for their own research, possible to re-use project data store, releasing only unidentifiable data	Costly, needs knowledgeable data analysts, some data may be proprietary
Press releases and briefings	Reaches a wide audience	Only summarizes, no guarantee that the media will disseminate information
Downloadable raw data (for SPSS, SAS, etc.)	Low cost	State has little say in how the data are used and interpreted, confidential and/or identifiable data cannot be published

Action Step 3: Make the Case

The dissemination method you choose and the type of information you can convey are intertwined. Once you determine your audience (that is, to whom you intend to present the evaluation findings) and how to present the findings, the next step is to determine what information to include.

For example, if you intend to use the results of the evaluation to support continued funding, you will likely highlight pieces of the evaluation that demonstrate areas where your program is meeting your measures for success (such as high cost savings or quality improvements). To strengthen the overall presentation, consider including a discussion of how you plan to address any weaknesses the evaluation revealed. This can help diffuse any criticism.

Action Step 4: Translate the Data

Remember that your audience may not be familiar with CM so it is important to establish the context for the evaluation before you give results. A short history of the program and its goals will allow your audience to make sense of the evaluation. Key aspects of translating research data for a general audience are:

- Identify key messages and frame your translation around these points.
- Use layman's terms and avoid technical jargon.
- Define key terms.
- Use clear and simple explanations.
- Use easy to read charts and graphs.
- Give a simple explanation of the methodology.
- Include an executive summary with written documents.

When writing for senior-level decisionmakers, policy officials, or a large general audience it may be tempting to use complicated language to explain your points. This is a common mistake that should be avoided. People often confuse technical language or complicated sentence structure with being well informed. You will be more effective if you stick to clear explanations that anyone could understand.

Chapter 5. Conclusions and Future Implications

Evaluation can help policymakers determine whether their CM initiatives are achieving the intended goals, whether performance improvements are possible, and whether there are more efficient ways to achieve similar effects. This User's Guide provides practical guidance to State decisionmakers charged with developing, implementing, and evaluating CM programs. Because each State will be implementing its CM program in a unique environment with different goals and objectives, we have attempted to outline the steps necessary and tradeoffs to consider when designing and implementing an evaluation rather than recommending a specific evaluation design.

How States Have Used Their Evaluations

Evaluation results from States have had an enormous impact on changing the course of the program. We hope that this guide will enable you to use evaluation to maximize your investment in CM. For example, North Carolina's evaluation efforts demonstrated significant cost savings from their CM program. After the findings were presented to the legislature, the State doubled program enrollment in 2005, and in September 2006, the program became Statewide. North Carolina's CM program is now considered the State's principal vehicle for managing care.

Washington State also shared its evaluation findings with the legislature. However, officials in Washington have found that their evaluation efforts have had the greatest impact on ongoing program management—identifying problem areas and opportunities for improvement. In particular, Washington has used its evaluation findings to improve CM vendor contracts and contract management.

As many States gain experience with designing, implementing, and evaluating Medicaid fee-for-service CM programs, policymakers and researchers will gain a greater depth of understanding of whether, and how, CM works.

Continued Evaluation Is Necessary for the Evolution of CM Programs

Like most of health care, CM programs are continuing to evolve to maximize benefits. Although CM programs were initially single-disease based, these programs are evolving away from disease-specific management to address the multiple and complex needs of patients. However, even as States refine and integrate their CM programs or shift their focus to other strategies for chronic care management, many of the principles of CM evaluation that are outlined in this guide will still be relevant.

References

[1] Suh R. Presentation of evaluation findings (draft). Rockville, MD: Agency for Healthcare Research and Quality; 2005.
[2] Johnson A. Measuring DM's net effect is harder than you might think. Manag Care 2003 June;12(6):28-32.
[3] Williams C. Medicaid disease management: issues and promises. Kaiser Commission on Medicaid and the Uninsured. Washington, DC: Kaiser Family Foundation; September 2004.
[4] Letter to State Medicaid Directors, Baltimore, MD: Centers for Medicare & Medicaid Services, 2004 Feb 25.
[5] Health Strategies Consultancy. Disease management in Medicaid. Oakland, CA: California HealthCare Foundation, 2004.
[6] Leatherman S, Berwick D, Iles D, et.al. The business case for quality: case studies and an analysis. Health Aff 2003 March-April; 22(2):17-30.
[7] Holtz-Eakin D. An analysis of the literature on disease management programs. Washington, DC: Congressional Budget Office; 2004.
[8] Goetzel RZ, Ozminkowski RJ, Villagra VG, et al. Return on investment in disease management: a review. Health Care Financ Rev 2005;26:1-19.
[9] Glabman M. 'Take my word for it': the enduring dispute over measuring dm's economic value. Manag Care Mag Apr 2006. Available at: http://www.managedcaremag.com/archives/0604/0604.dmvalue.html. Accessed October 19, 2007.
[10] Chang D, Burton A, O'Brien J, et al. Honesty as good policy: evaluating maryland's medicaid managed care program. Milbank Q 2003;81(3):389-414.
[11] Wallace P. Is there long-term value in programs? Reflections on the 2004 CBO report. The Disease Management Colloquium. Philadelphia: Jefferson Medical College; 2005 June 22.
[12] Lind A. Department of Social and Health Services, Washington State. Personal communication; March 17, 2006.
[13] Vann C. Texas Health and Human Services Commission. Personal communication; March 26, 2006.
[14] Suchman E. Evaluation research: principles and practice in public service and social action programs. New York: Russell Sage College; 1967.
[15] Consumer Assessment of Healthcare Providers and Systems.™ Rockville, MD: Agency for Healthcare Research and Quality. Available at http://www.cahps.ahrq.gov/. Accessed October 19, 2007.
[16] Rosenman M, Holmes A, Ackermann R, et al. The Indiana chronic disease management program. Milbank Q 2006;84(1):135-63.
[17] Congressional Budget Office. An analysis of the literature on disease management programs. Washington, DC: Government Printing Office; 2004.
[18] Fitzner K, Sidorov J, Fetterolf D, et al. Principles for assessing disease management outcomes. Dis Mang 2004;7(3):191-201.
[19] Christakis D, Connell F, Richardson A, et al. Report of disease management evaluation. Seattle, WA: University of Washington; 2004.
[20] Lawson K. Virginia Department of Medical Assistance Services. Personal communication; March 29, 2006.
[21] Hewson D. Community Care of North Carolina. Personal communication; October 12 2006 and March 16, 2007.
[22] Morgan J. Pennsylvania Office of Medical Assistance Programs. Personal communication; March 6, 2006.
[23] Blueprint for Health. Burlington, VT: Vermont Department of Health; 2005. Available at http://healthvermont.gov/blueprint.aspx. Accessed October 23, 2007.
[24] North Carolina Area Health Education Centers. Chapel Hill, NC: North Carolina AHEC Program. Available at http://www.ncahec.net . Accessed October 23, 2007.
[25] Inui T. Regenstief Institute, Indianapolis, IN. Personal communication; March 8, 2006.
[26] Rosenman M, Holmes A, Ackermann R, et al. The Indiana Chronic Disease Management Program. Milbank Q 2006;84(1):135-63.
[27] Lind A, Kaplan L, Berg G. Evaluation of an asthma disease management program in a Medicaid population. Dis Manag 2006;14(3):151-61.
[28] Gold MR, Siegel JE, Russell, et al, eds. Cost-effectiveness in health and medicine. Chicago, IL: Health Administration Press; 1996.
[29] Rossi PH, Lipsey MW, Freeman HE. Evaluation: a systematic approach. Thousand Oaks, CA: Sage Publications, 2003.

[30] Veney JE, Kaluzny, AD. Evaluation and decision making for health services programs. Englewood Cliffs, NJ: Prentice Hall; 1998.
[31] Wilson T. Evaluating ROI in State disease management programs. Issue Brief, State Coverage Initiatives. Washington, DC: AcademyHealth; 2003.
[32] Fetterolf D, Wennberg D, DeVries A. Estimating the return on investment in disease management programs using a pre-post analysis. Dis Manag 2004 Spring;7(1):5-23.
[33] Evaluation of Community Care of North Carolina asthma and diabetes management initiatives: January 2000-December 2002. Raleigh, NC: NC Foundation for Advanced Health Programs, Inc; 2004. Available at: http://www.communitycarenc.com/PDFDocs/Sheps%20Eval.pdf. Accessed October 23, 2007.
[34] Linden A, Adams JL, Roberts N. Evaluation methods in disease management: determining program effectiveness. Washington, DC: Disease Management Association of America; 2006. Available at http://www.dmaa.org/pdf/Evaluation_Methods_in_DM.pdf. Accessed October 23, 2007.
[35] Family and Children's Health Programs Group, Center for Medicaid and State Operations. The guide to quality measures: a compendium. Medicaid and SCHIP quality improvement, Volume 1.0. Baltimore, MD: Centers for Medicare & Medicaid Services; Sept 2006. Available at http://www.cms.hhs.gov/MedicaidSCHIPQualPrac/Downloads/pmfinalaugust06.pdf. Accessed October 24, 2007.
[36] National Heart, Lung, and Blood Institute. Guidelines for the diagnosis and management of asthma (EPR-3). Rockville, MD: National Institutes of Health; July 2007. Available at http://www.nhlbi.nih.gov/guidelines/asthma. Accessed October 24, 2007.
[37] American Diabetes Association clinical practice recommendations, 2007. Diabetes Care 2007;1(Suppl 30). Available at http://www.diabetes.org/for-health-professionals-and-scientists/cpr.jsp . Accessed October 24, 2007.

Appendix 1. Example Measures for Evaluating the Impact of Medicaid Care Management

This table provides examples of different measures that could be used by States to monitor and evaluate their CM programs. This is not an exhaustive list of measures. We have divided the measures into six types: intervention-process, quality-process, quality-outcome, satisfaction, utilization, and cost.

Appendix table. Examples of evaluation measures

Type	Measure	Description	Data source
Intervention – process	Eligible for CM	Percent of beneficiaries who are eligible for the CM program, among all beneficiaries	Administrative data (and/or health risk appraisals)
	Enrolled in CM	Percent of beneficiaries who are enrolled in CM, among all eligible	Administrative data
	Engaged in CM	Percent of beneficiaries who are engaged in the CM program, among all enrolled	Administrative data
Quality - process	Smoking cessation	Percent of enrollees who received advice to quit smoking among all smokers enrolled in CM	Medical record
	Blood pressure	Percent of patients with BP measurement recorded among those enrolled in CM	Medical record
	LDL cholesterol screen	Percent of patients who have documentation in the medical record of cholesterol screening within the last year among those enrolled in CM	Medical record
	Lipid profile	Percent of patients receiving at least one LDL-C screen among those enrolled in CM	Administrative: lab claims
	ACE inhibitor	Percent of patients with CAD who have diabetes and were prescribed ACE inhibitor	Administrative: pharmacy claims

Appendix table. Examples of evaluation measures (continued)

	HbA1c screen	Percent of patients with diabetes receiving HbA1c screen within the past year	Medical record
	Retinal exam	Percent of patients with diabetes who have had a retinal exam in the past year	Administrative: claims
Quality - outcome	Blood pressure management	Percent of patients with most recent BP <140/90 mmHg among those enrolled in CM	Medical record
	LDL cholesterol control	Percent of patients with diabetes with most recent LDL-C <130 mg/dL	Medical record
Satisfaction	CAHPS™	Patient experience of care	Survey
Utilization	Hospital admissions	Number of hospital admissions in last year among those enrolled in CM	Administrative: claims
	ED use	Number of emergency department visits in last year among those enrolled in CM	Administrative: claims
	Physician visits	Number of physician visits in last year among those enrolled in CM	Administrative: claims
	Pharmacy	Number of 30-day prescriptions filled in last year among those enrolled in CM	Administrative: pharmacy claims
Cost	Hospital claims	Total $ of hospital claims paid	Administrative: hospital claims
	Pharmacy claims	Total $ in pharmacy claims paid	Administrative: pharmacy claims
	Total expenditures	Total $ in claims paid	Administrative: claims

Appendix 2. Glossary of Evaluation Terms

Baseline data: Starting data that can be used for comparing data collected after the intervention begins.

Comparable data: Data from another population or location that contain similar characteristics to those of the group undergoing the intervention. These data are used when it is not possible to select a reference group from within the study.

Cohort evaluation: Follows the reference and intervention group over time.

Control group: Also known as a reference group. A group not receiving the care management intervention or services. This population is used as a comparison group with those receiving the intervention. Selection of the control group depends on the design of the care management program.

Confounding variable: A factor that cannot be controlled or measured that may influence the outcome of the evaluation.

Cross-sectional evaluation: Conducted at a single point in time; after the program has been implemented.

Longitudinal evaluation: Examines the differences in intervention outcomes at several points in time both before and after the implementation of a care management program.

Matched control: A population that is similar to the intervention group chosen to serve as a reference group.

Pre/post design: The population receiving the intervention is used as its own reference group. Characteristics of the group before and after the intervention are compared.

Randomized control trial: Participants are arbitrarily assigned to the reference or treatment group. Randomized control trials analyze program effectiveness and are considered the scientific gold standard because they control for any potential biases in the groups that may unfairly influence the evaluation.

Risk stratification: A statistical process in which the care management and reference groups may be divided according to disease status (or case mix) that allows for separate comparisons to be made for people in different disease states.

Regression to the mean: A phenomenon that is similar to the "law of averages." In care management, for example, a group of people who have high medical costs in one year will tend to have average costs that are considerably lower in the following year.

Staged implementation: When a care management program is rolled out in several phases. Staged implementation can be selected to designate a reference group by using those who have not yet received services as the reference (or control) population.

Appendix 3. Additional Resources

Congressional Budget Office. An analysis of the literature on disease management programs. Washington, DC: Government Printing Office; 2004.

Crippen DL. Disease management in Medicare: Data analysis and benefit design issues. Congressional Budget Office Testimony; Statement before the Special Committee on Aging; United States Senate; Sept 19, 2002.

Outcomes guideline report. Washington, DC: Disease Management Association of America; 2006 Dec.

Duncan I. Accuracy in the assessment of return on investment of defined population interventions. Chicago, IL: Jt Comm J Qual Patient Saf 2005;31(6):357.

Duncan I. Evaluating Disease management savings outcomes. Schaumburg, IL: Society of Actuaries; Oct 2005.

Linden A, Roberts N. A users guide to the disease management literature: Recommendations for reporting and assessing program outcomes. Am J Manag Care 2005;11(2):81-90.

MacDowell M, Wilson T. Framework for assessing causality in disease management programs. Washington, DC: Disease Management Association of America; 2003.

Rosenman M, Holmes A, Ackermann R, et al. The Indiana Chronic Disease Management Program. Milbank Q 2006;84(1):135-63.

Volpel A, O'Brien J, Weiner J. Strategies for assessing health plan performance on chronic diseases: Selecting performance indicators and applying health-based risk adjustment. Hamilton, NJ: Center for Health Care Strategies; March 2005.

Williams C. Medicaid disease management: issues and promises. Kaiser Commission on Medicaid and the Uninsured. Washington, DC: Kaiser Family Foundation, Sept 2004.

Wilson T, Thar W, Gruen J, et al. Assessing ROI of defined-population disease management interventions. Jt Comm J Qual Patient Saf 2004;30(11):614-21.

Wilson, T.W. Evaluating ROI in State disease management programs. State Coverage Initiatives Issue Brief. Washington, DC: AcademyHealth, Nov 2003.

Wilson T, Thar W, Gruen J. Reply. Jt Comm J Qual Patient Saf 2005;31(6):358.