Contract No.: 500-95-0047(09)
MPR Reference No.: 8756-430

MATHEMATICA
Policy Research, Inc.

Third Report to Congress on the Evaluation of the Medicare Coordinated Care Demonstration

January 3, 2008

Deborah Peikes
Randall Brown
Arnold Chen
Jennifer Schore

Submitted to:

U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
Office of Research, Development, and Information
C3-19-07 Central Bldg.
7500 Security Blvd.
Baltimore, MD 21244-1850

Project Officer:
Carol A. Magee

Submitted by:

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director:
Randall Brown

CONTENTS

TABLES

FIGURES

BRIEF OVERVIEW OF FINDINGS

Chronic illnesses, such as heart disease and diabetes, take a toll on beneficiaries' quality of life and result in significant expense for Medicare. While 44 percent of beneficiaries in 2005 were treated for one or more of nine chronic illnesses, they accounted for nearly 85 percent of all Medicare spending.[1] These beneficiaries often have multiple chronic illnesses, which compounds the cost and complexity of their care: Anderson and Horvath (2002) report that 20 percent of beneficiaries have 5 or more chronic illnesses, but these individuals account for nearly two-thirds of all Medicare expenditures.

Chronically ill patients make a myriad of decisions daily about issues that affect their health, such as diet, exercise, medication, and when to seek medical attention. Often, however, patients do not make wise decisions about these issues, either because they do not fully understand the implications of a particular decision for their health, or because they lack the support that would help them to make better choices.

Despite the costs and complexity of providing effective chronic care, studies have suggested that many acute health problems, and the resulting monetary and social costs, can be prevented if physicians provide medical care that has been proven to be effective; patients adhere to recommended diet, medication, exercise, and self-care regimens; and providers communicate more effectively with each other and their patients. A number of small pilot programs designed to improve patients' adherence to treatment regimens and physicians' adherence to professional guidelines have been found to be effective in improving patient outcomes and reducing costs. Other evidence suggests that better medication adherence leads to improved outcomes and lower costs and that patient self-management education can lower service use and costs. The potential to improve outcomes has led many managed care plans and indemnity insurers to develop their own programs or contract with disease management providers for such programs (see Villagra and Ahmed 2004 for evidence of the effectiveness of disease management for patients with diabetes in a managed care setting). The Disease Management Purchasing Consortium estimates that the revenue of disease management organizations that provide outsourced services grew from $78 million in 1997 to $1.2 billion in 2000 and is projected to grow to $1.8 billion by 2008. Evidence from large-scale studies on the effectiveness of care coordination is not yet available, however, studies of disease management interventions have shown mixed effects on health outcomes and cost.

1. The Medicare Coordinated Care Demonstration

To determine whether care coordination improves the quality of care and reduces Medicare expenditures, the Balanced Budget Act of 1997 mandated that the Secretary of Health and Human Services conduct and also independently evaluate care coordination programs in the Medicare fee-for-service setting (Section 4016 (b)(3)(B) of H.R. 2015). The legislation authorizes the Secretary to permanently implement components of the demonstration if the evaluation shows that the projects (1) reduce net Medicare expenditures, or (2) increase the

[1] These background estimates were generated by CMS using data from the chronic care warehouse, for a five percent sample of beneficiaries.

"quality of health care services and satisfaction of beneficiaries and health care providers" without increasing expenditures. CMS has extended three of the programs.

The mandated Medicare Coordinated Care Demonstration (MCCD) is among the first random assignment studies of care coordination. It tests specifically whether such interventions can lower costs and improve patient outcomes and well-being in the Medicare fee-for-service setting, for beneficiaries with chronic illnesses.

The Centers for Medicare & Medicaid Services (CMS) selected 15 demonstration programs for the MCCD in January 2001 in a competitive awards process under which each program was allowed to define, within broad boundaries, its own intervention and target population. The programs each began enrolling patients at some point between April and September of 2002, and were authorized to operate for four years. Eleven of the 15 programs later requested, and were granted, two-year extensions, and will continue to enroll beneficiaries until six months before their respective ending dates in 2008.[2] The programs run by the University of Maryland, Charlestown Retirement Community, and Quality Oncology ended as planned in 2006. Georgetown University Medical School decided to end its program five months early, on December 31, 2005. Each program was responsible for identifying beneficiaries who met its eligibility requirements and recruiting them for the study. Beneficiaries who agreed to participate in one of the programs were randomly assigned by the evaluator, Mathematica Policy Research, Inc. (MPR), to either the treatment group, which received the program's intervention, or the control group, which received usual care. Both groups continued to obtain their traditional Medicare-covered services from fee-for-service providers. CMS paid each program a negotiated monthly fee for each beneficiary in the treatment group, ranging from $50 to $444 per beneficiary, for each month the beneficiary remained eligible for and enrolled in the program.

As required in the Congressional mandate, two earlier reports to Congress described the types of programs and beneficiaries participating in the demonstrations, summarized the attitudes of patients and physicians toward the programs, and provided interim estimates of program impacts on quality of care, service use, and Medicare expenditures. Over the first 25 months of the demonstration, only one program (Mercy) had reduced hospitalizations (by 27 percent); that program was also the only one to reduce Medicare Part A and B expenditures (by 13 percent), but not by enough to cover its program fees. Some other programs had favorable treatment-control differences, but they were not statistically significant. Including program fees, the interim findings indicated that cost neutrality was *possible* for five programs, but the evidence was not conclusive. Six programs definitely increased total expenditures, and the remaining four probably increased expenditures.

This third report to Congress synthesizes findings from more than four years of the demonstration programs' operations, and provides the most comprehensive and rigorous estimates ever presented on the effectiveness of care coordination interventions in a Medicare fee-for-service setting. As required, the analysis was conducted on each program separately, given their differences in target populations and interventions. However, the report also describes the aggregate findings across all programs for some key outcomes, as a meta-analysis of the effects of the 15 programs.

[2] In late September 2007, CMS agreed to delay the termination of enrollment. The programs may now continue to enroll beneficiaries until three months before their respective ending dates in 2008.

2. Findings from the Third Report

Most of the care coordination programs tested in this third report had limited or no improvements in quality of care, few achieved cost neutrality, and none reduced total Medicare expenditures when care coordination fees were included. Still, five of the programs (Georgetown University, Health Quality Partners, Medical Care Development, QMed, and Quality Oncology) did have some modest favorable effects on quality without significantly increasing total Medicare expenditures. While these relatively more successful programs in this report differ somewhat from the ones showing the most promise in the second report, the finding of limited success is similar. Treating only statistically significant treatment-control differences as evidence of program effects, the results from this third report show:

- No clear effects on patients' adherence to medication, diet, exercise, or self-care, despite a substantial increase in the percentage of beneficiaries reporting that they received health education
- Few effects on beneficiaries' overall satisfaction with care
- Improvements in a few quality of care indicators (such as increased testing rates for cholesterol) for patients with diabetes and coronary artery disease
- A reduction in the proportion who had potentially preventable hospitalizations for congestive heart failure, but an increase in the proportion who had preventable hospitalizations for diabetes
- No effects on functioning or mortality
- A slight reduction in the number of hospitalizations across all programs combined (4.5 percent), driven by significant, sizable reductions (17 and 24 percent) in 2 of the 15 programs (Mercy Medical Center and Georgetown University) and moderate non-significant treatment-control differences of -10 to -16 percent in 3 others (Health Quality Partners, Hospice of the Valley, and University of Maryland) (Table A)
- A reduction in Medicare Part A and B expenditures, excluding program fees, for one program (Health Quality Partners)

Based on the patterns of treatment-control differences in hospitalizations, Medicare Part A and B expenditures, and total Medicare expenditures (including care coordination fees), we conclude that only 2 of the 15 programs (Health Quality Partners and Georgetown University Medical School) are clearly cost neutral. (See Table A. Section F contains more details.) Ten of the programs (Avera, Carle Foundation, CenVaNet, Charlestown Retirement Community, CorSolutions, Hospice of the Valley, Jewish Home and Hospital Lifecare System, Mercy Medical Center, University of Maryland Medical School, and Washington University School of Medicine) significantly *increased* total costs to Medicare. The remaining 3 (QMed, Quality Oncology, and Medical Care Development) may be cost neutral or may have treatment-control differences that, by chance, approximately offset the fees paid.

Among the five programs that are or may be cost neutral, only two (Health Quality Partners and QMed) appear to be sustainable, promising models at the monthly fees they received under the demonstration. While neither program *reduced* overall Medicare expenditures (including care coordination fees) during the demonstration period, they did not increase them, and both

TABLE A

TREATMENT-CONTROL DIFFERENCES IN HOSPITALIZATIONS AND TOTAL EXPENDITURES
INCLUDING PROGRAM FEES THROUGH JUNE 2006,
AMONG BENEFICIARIES ENROLLED THROUGH JUNE 2005
(Regression Adjusted)

	Annual Number of Hospitalizations	Monthly Total Medicare Expenditures, Including Program Fees[a]		
	Impact (As Percentage of Control Group Mean)	Treatment-Control Difference ($)	Impact (As Percentage of Control Group Mean)	Cost Neutral
1,100 or More Treatment Group Members Enrolled Through June 2005				
Carle	–0.2	190	26.7†	No
CorSolutions	–3.0	217	8.3†	No
Washington University	–2.2	231	12.2†	No
415 to 725 Treatment Group Members Enrolled Through June 2005				
Avera	2.3	261	19.4†	No
CenVaNet	7.2	113	13.1†	No
Charlestown	12.3	374	36.5†	No
Health Quality Partners	–13.6	2	0.3	Yes
Hospice of the Valley	–10.7	167	8.1†	No
Jewish Home and Hospital	4.4	299	17.1†	No
Medical Care Development	–3.3	36	2.6	Unlikely
Mercy	–17.0*	135	11.3†	No
QMed	–7.4	-2	–0.2	Possibly
Under 115 Treatment Group Members Enrolled Through June 2005				
Georgetown	–24.1*	-93	–3.7	Yes
Quality Oncology	–1.8	-62	–1.9	Unlikely
University of Maryland	–16.1	1,080	41.5	No
All Programs	**–4.5***	**154**	**11.3†**	**No**

Sources: Medicare Enrollment Database and Standard Analytic File.

Note: Numbers marked by a * symbol denote statistically significant treatment-control differences at the 10 percent level for hospitalizations. Differences in expenditures including care coordination fees were tested at the 20 percent level; the symbol † denotes differences that are statistically significant at the 20 percent level (see section F for rationale for this non-traditional significance level). Negative estimates imply that hospitalizations or Medicare expenditures are lower for the treatment group, a favorable outcome.

[a]Sample member observations are weighted according to the proportion of the followup period the individual met CMS's demonstration-wide eligibility criteria (alive, in fee-for-service, have both Part A and Part B coverage, and have Medicare as primary payer).

programs improved some of the quality of care indicators examined. Given that the estimated savings in Part A and B costs for QMed were not statistically significant, however, CMS may want to reduce the fee somewhat below that paid during the demonstration to increase confidence that this program would be cost neutral were it to be extended. Two of the other three potentially cost neutral programs either dropped out before the end of the initial four-year period (Georgetown University) or did not seek an extension (Quality Oncology), and neither program was able to recruit adequate numbers of beneficiaries to make them promising models for the Medicare program. The fifth potentially cost neutral program (Medical Care Development) had small, statistically insignificant differences in hospitalization and Medicare Part A and B expenditures that were not large enough to offset program fees, making cost neutrality unlikely. A sixth program, Mercy Medical Center, significantly reduced hospitalizations, but had a sizable fee that was over twice as large as the estimated savings in Part A and B expenditures. If these same reductions could be achieved with an intervention that could be afforded at a fee less than half the amount paid during the demonstration, this program could be cost neutral as well. CMS has extended three of the programs.

While terminating all of the demonstration programs would ensure that no new cost increases will be created for Medicare from care coordination fees, failing to pursue effective programs such as Health Quality Partners or QMed (and possibly Mercy Medical Center) may mean a missed opportunity to substantially improve the quality of care for chronically ill beneficiaries at no increase in cost to Medicare. If the interventions can maintain their effectiveness at lower cost, it may even be possible to generate net savings for Medicare. Furthermore, the benefit of identifying successful interventions could be great for Medicaid as well, as nearly all states are investing in disease management programs, typically with little or no evidence that the programs will generate the savings that commercial programs promise. Thus, honing in on a detailed, concrete description of successful interventions for those with chronic illnesses and testing the replicability of these interventions seems to warrant serious consideration.

THIRD REPORT TO CONGRESS IN THE EVALUATION OF THE MEDICARE COORDINATED CARE DEMONSTRATION

This third report to Congress on the Medicare Coordinated Care Demonstration synthesizes findings from more than four years of the demonstration programs' operations, and provides the most comprehensive and rigorous estimates to date of the effectiveness of care coordination interventions in a Medicare fee-for-service setting. The report summarizes the findings from the second report to Congress on patient and physician satisfaction and patients' knowledge, behavior, adherence, unmet needs, and well-being. In addition, new analyses presented here update the earlier reports on the characteristics of enrollees, and evaluate extended followup data on the claims-based measures of patients' Medicare service use and expenditures, and the quality of care received. A synthesis of the findings attempts to draw inferences about "what works" and "for whom" by exploring associations between program impacts and characteristics of the interventions and the target populations.

The report begins with a brief description of the potential of care coordination to improve quality and reduce costs (Section A). Sections B and C describe the program interventions and the beneficiaries served by the programs. Section D summarizes how well beneficiaries and physicians liked the program. Effects on beneficiary knowledge, adherence, unmet needs, well-being and quality of care are described in Section E. Section F presents results on Medicare-covered hospital use and expenditures, and Section G synthesizes the results and draws conclusions.

A. THE RATIONALE FOR CARE COORDINATION

Chronic illnesses, such as heart disease and diabetes, take a toll on beneficiaries' quality of life and result in significant expense for Medicare. While 44 percent of all beneficiaries in 2005 were treated for one or more of nine chronic illnesses, they accounted for 85 percent of all Medicare spending (see Brief Overview of Findings for source of these estimates). These beneficiaries often have multiple chronic illnesses, which compounds the cost and complexity of their care: Anderson and Horvath (2002) report that 20 percent of beneficiaries have 5 or more chronic illnesses, but these individuals account for nearly two-thirds of all Medicare expenditures. In addition, the care Medicare beneficiaries receive for chronic illnesses in the fee-for-service program is often uneven and of poor quality (Leatherman and McCarthy 2005; Baicker and Chandra 2004; Jencks et al. 2003).

Chronically ill patients make a myriad of decisions daily about issues that affect their health, such as diet, exercise, medication, and when to seek medical attention. Often, however, patients do not make wise decisions about these issues, either because they do not fully understand the implications of a particular decision for their health, or because they lack the support that would help them to make better choices.

Despite the costs and complexity of providing effective chronic care, studies have suggested that many acute health problems, and the resulting monetary and social costs, can be prevented if physicians provide medical care that has been proven to be effective; patients adhere to recommended diet, medication, exercise, and self-care regimens; and providers communicate more effectively with each other and their patients. A number of small pilot programs designed

to improve patients' adherence to treatment regimens and physicians' adherence to professional guidelines have been found to be effective in improving patient outcomes and reducing costs (see reviews by Chen et al. 2000; Wagner et al. 2001). Other evidence suggests that better medication adherence leads to improved outcomes and lower costs (Bagchi et al. 2007; Sokol et al. 2005) and that patient self-management education can lower service use and costs (Bodenheimer et al. 2002; Lorig et al. 2001; Lorig et al. 1999). The potential to improve outcomes has led many managed care plans and indemnity insurers to develop their own programs or contract with disease management providers for such programs (see Villagra and Ahmed 2004 for evidence of the effectiveness of disease management for patients with diabetes in a managed care setting). The Disease Management Purchasing Consortium estimates that the revenue of disease management organizations that provide outsourced services grew from $78 million in 1997 to $1.2 billion in 2000 and is projected to grow to $1.8 billion by 2008 (Matheson and Psacharopoulos 2006). Evidence from large-scale studies on the effectiveness of care coordination is not yet available ; however, studies of disease management interventions have shown mixed effects on health outcomes and cost (Gravelle et al. 2007; Smith et al. 2005; Goetzel et al. 2005; Kilpatrick et al. 2005; Galbreath et al. 2004; Ofman et al. 2004).

To determine whether care coordination improves the quality of care and reduces Medicare expenditures, the Balanced Budget Act of 1997 mandated that the Secretary of Health and Human Services conduct and also independently evaluate care coordination programs in the Medicare fee-for-service setting (Section 4016 (b)(3)(B) of H.R. 2015). The legislation authorizes the Secretary to permanently implement components of the demonstration if the evaluation shows that the projects (1) reduce net Medicare expenditures, or (2) increase the "quality of health care services and satisfaction of beneficiaries and health care providers" without increasing expenditures. CMS has extended three of the programs.

The mandated Medicare Coordinated Care Demonstration (MCCD) is among the first random assignment studies of care coordination. It tests specifically whether such interventions can lower costs and improve patient outcomes and well-being in the Medicare fee-for-service setting, for beneficiaries with chronic illnesses.

B. THE MEDICARE COORDINATED CARE INTERVENTIONS

The evaluation of the MCCD had the goal not only of estimating the impacts of the demonstration on patients' quality of care, health, and Medicare service use and expenditures, but also of assessing which intervention features contributed to success and which were barriers to improving these outcomes. To lay the groundwork for this assessment and to provide context for the evaluation's impact estimates—the focus of this report—this section provides an overview of the 15 demonstration program interventions. It begins with the evaluation's working definitions of care coordination and disease management.

Over the past two decades, disease management and care coordination have garnered much attention because of their potential to rein in the costs of treating patients with chronic illnesses, and because of industry claims of sizable cost savings. Such interventions are based on the hypothesis that some combination of the following will address the barriers to improving patient health and thereby reducing costs (see Table 1): (1) improved patient adherence to treatment

TABLE 1

BARRIERS TO IMPROVED CHRONIC CARE

Patient Behaviors
Factors hindering *adherence* (to prescribed medications, diet, exercise, self-care, and medical diagnostic and treatment services)[a]
Lack of knowledge and understanding of the importance of adherence
Inadequate skills to perform self-care (such as blood sugar testing or daily foot inspection)
Reluctance or ambivalence towards accepting chronic illness and changing long-standing habits
Lack of self-efficacy to adhere or perform self-care
Depression, fear, anxiety
Tobacco or alcohol dependencies
Poor assertiveness or communication skills with family members or health care providers
Cognitive deficits
Sensory deficits (vision, hearing)
Mobility impairments
Inadequate access to transportation
Geographic or physical isolation
Poverty/inadequate insurance coverage
Caregiving responsibilities (e.g., ill family member)
Factors hindering *appropriate response* to disease complication or exacerbation: early recognition of warning signs and symptoms, appropriate self-treatment, appropriate seeking of urgent medical care
Lack of knowledge and understanding of the importance of early detection and management of deterioration
Inadequate skills (to recognize warning signs, to self-manage, or to be assertive in getting through to the doctor)
Lack of self-efficacy to recognize problems and respond appropriately
Lack of self-efficacy and skills to manage transitions between care settings (hospital to SNF to home health and outpatient care—new self-care instructions, follow-up appointments, changes in medications)
Depression
Cognitive deficits
Poverty/inadequate insurance coverage
Transportation difficulties
Physician Behaviors
Factors hindering delivery of high-quality chronic illness care: assessment, monitoring, care planning, evidence-based care, patient education, and prompt responses to changes in patient status[b]
Barriers to Evidence-Based Care
Inadequate time
Underdeveloped patient communication and counseling skills
Lack of self-efficacy to counsel on lifestyle and adherence
Inadequate office systems to support adherence to recommended guidelines for diagnosis and treatment
Lack of reminder systems and patient registries
Acute care focus during office visits
Lack of contact with patient between visits
Lack of incentives in reimbursement system
Barriers to Communication with Patients
Inadequate time
Inadequate office communication and triage systems
Acute care focus during office visits
Underdeveloped patient communication and counseling skills

TABLE 1 *(continued)*

Physician Behaviors *(continued)*
Lack of self-efficacy to counsel on lifestyle and adherence
Lack of contact with patient between visits
Lack of awareness of patients' specific barriers to adherence and self-care
Lack of incentives in reimbursement system
Barriers to Communication Across Providers and Management of Transitions Between Care Settings
Inadequate time
Lack of contact with patient between visits
Lack of awareness of other providers' treatments
Lack of knowledge of transitions between care settings (hospital to SNF to home health and outpatient care)
Lack of incentives in reimbursement system
Appropriate Drug Therapy and Avoidance of Polypharmacy
Inadequate time
Lack of contact with patient between visits
Lack of awareness of other providers' treatments (or adverse reactions to those treatments)
Lack of knowledge of medication changes between care settings (hospital to SNF to home health and outpatient care)
Lack of incentives in reimbursement system
Acute care focus during office visits
Inadequate office systems to support adherence to recommended guidelines for diagnosis and treatment
Lack of reminder systems and patient registries

[a]Examples of self-care include weighing oneself daily for CHF or checking blood sugar for diabetes. Examples of adhering to medical diagnostic and treatment services include keeping appointments for visits to specialists, physical therapy, or special diagnostic or imaging tests.

[b]Assessment refers to a thorough, in-depth examination that would uncover the patient-related barriers listed above.

regimens, (2) increased physician use of evidence-based guidelines for medications and other treatments, (3) improved communication between patients and providers and across providers, (4) better management of transitions between care settings, (5) careful monitoring of symptoms and of problems related to polypharmacy, and (6) improved access to health-related services. Figure 1 presents a model of how the MCCD programs might affect such change and an overview of the measures the evaluation used to assess change.

The terms care coordination and disease management are often used interchangeably and these two types of interventions may, in fact, share some features. However, they typically differ on a few dimensions. Care coordination usually refers to an array of services for patients with multiple medical or behavioral health conditions or who are medically complex. It often involves assigning patients to a single staff member or staff team to (1) monitor clinical care and support services; (2) improve the flow of information across providers and, in particular, assist with transitions between care settings; and (3) assist in accessing needed health and support services. (This service is sometimes also referred to as case management or care management.) Disease management typically includes services that (1) teach members about their disease and how to adhere to prescribed diet, medication, exercise, and self-care regimens; (2) monitor members' clinical status and adherence to treatment recommendations; and (3) monitor provider adherence to evidence-based practice guidelines. Furthermore, disease management is usually targeted to individuals with specific chronic diseases, such as heart failure or diabetes, although many patients with these conditions also have other chronic diseases. The target diseases often have complex treatment regimens, and keeping them under control requires the sustained efforts of patients and physicians.

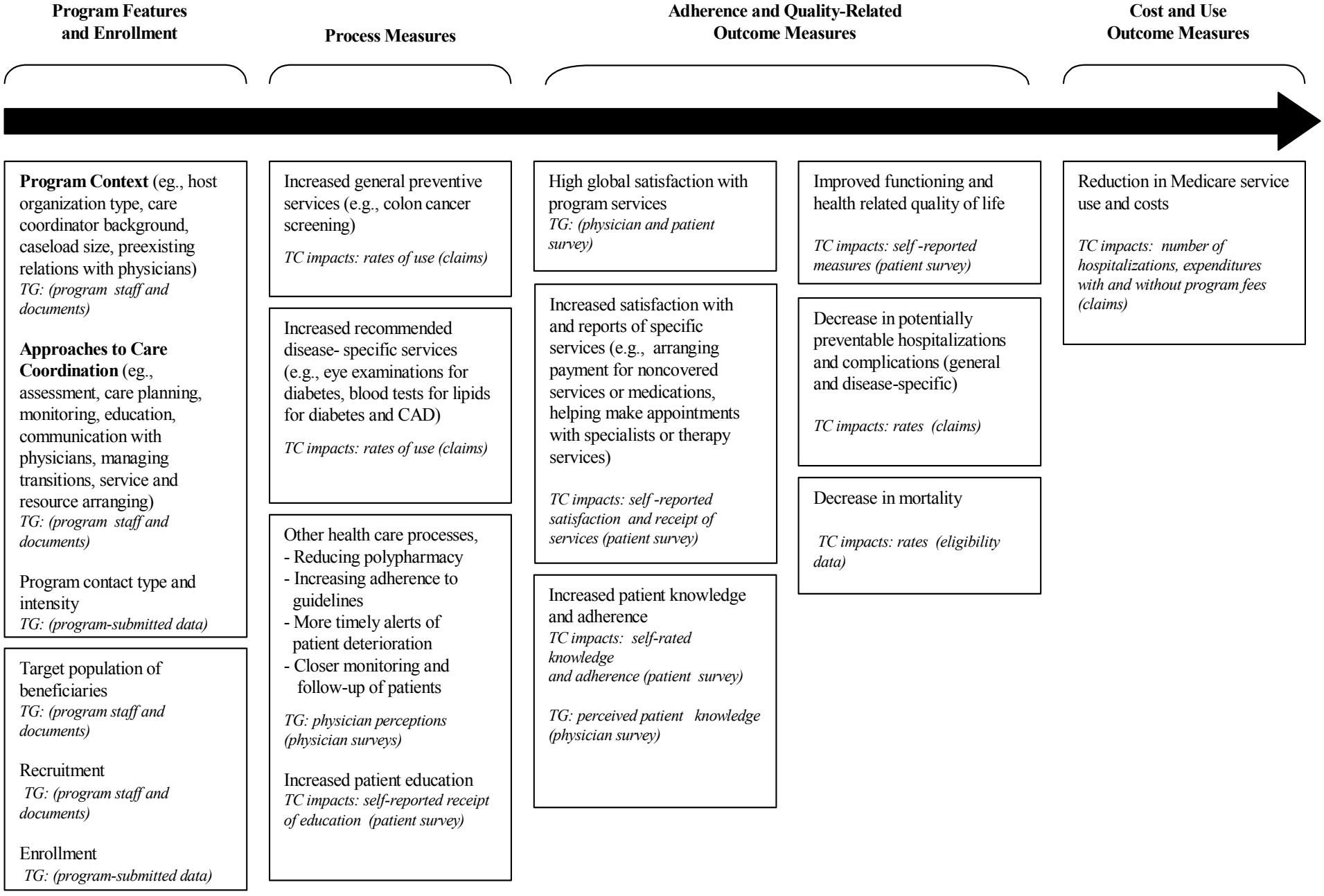
The MCCD programs were largely free to design their interventions, and most included aspects of both care coordination and disease management, with the focus of each program depending on the experience or usual business of its host organization. For simplicity (and consistency with the title of the demonstration) the rest of this report refers to all programs' interventions as care coordination and the programs' patient care staff as care coordinators. (See Brown et al. 2007 for a more detailed description of the programs. Updated profiles of the programs are available on request. These documents are the sources of information presented in this section.)[3]

Program Host Types, Service Areas, and Target Populations. The program hosts (that is, the demonstration grantees) included commercial disease management providers, academic medical centers, and hospitals, among others (Table 2). The programs operated in 16 states (mostly in the Northeast and Midwest) and the District of Columbia; five served beneficiaries living in sparsely populated rural areas. They varied in the numbers and types of chronic conditions they targeted, with 6 focusing on a single specific condition (most frequently heart failure), and the others serving patients with a variety of diagnoses. Negotiated program fees ranged from $50 to $444 (each program's fee was required to be less than 20 percent of the average expected Medicare expenditure for its target population, because literature suggested this

[3] The information reported in Appendix A and presented in the remainder of this section was obtained during the evaluator's site visits and telephone calls to the programs and reflects what program staff said.

FIGURE 1

LOGIC MODEL FOR DEMONSTRATION OUTCOMES



TG = measures that were available for treatment group members only.

TC impacts = measures available for both treatment and control group members and comparisons of which thus represent demonstration impacts. Text in parenthesis describes the source of data for the measure.

TABLE 2

CARE COORDINATION PROGRAMS PARTICIPATING IN THE EVALUATION

Host Organization (Monthly Program Fee)[a]	Host Organization Type	Program Service Area	Program Target Diagnoses
Avera Research Institute/Avera McKennan Hospital and University Health Center ($316)	Hospital	Rural counties in Iowa, Minnesota, Nebraska, and South Dakota	CHF
Carle Foundation ($159)	Integrated delivery system	Rural counties in east central Illinois and west central Indiana	Heart conditions, diabetes, chronic lung disease
CenVaNet ($80)	Care coordination provider	Richmond, Virginia	Heart conditions, diabetes, chronic lung disease, cerebrovascular disease
Charlestown Retirement Community ($244)	Retirement community	3 retirement communities in the Baltimore area	Heart conditions, diabetes, COPD
CorSolutions ($444)	Disease management provider	Harris County (Houston), Texas	CHF
Georgetown University Medical School ($320)	Academic institution	Washington, DC, and parts of Maryland and Virginia	CHF
Health Quality Partners ($108)	Quality improvement services provider	Eastern Pennsylvania (rural)	Heart conditions, diabetes, asthma, moderate to severe hyperlipidemia or hypertension
Hospice of the Valley ($224)	Hospice	Maricopa County, Arizona (greater Phoenix)	CHF, COPD, cancer, neurological conditions
Jewish Home and Hospital Lifecare System ($317)	Long-term care provider	Manhattan, New York City	Heart conditions, diabetes, chronic lung disease, cancer, liver disease, stroke, or other cerebrovascular disease, psychotic disorder, major depressive or anxiety disorder, Alzheimer's disease, or other cognitive impairment
Medical Care Development ($297)	Hospital consortium	Rural areas of Maine	Heart conditions
Mercy Medical Center/North Iowa ($257)	Hospital	Rural areas of Iowa	CHF, chronic lung disease, liver disease, stroke, vascular disease, renal failure
QMed ($96)	Disease management provider	2 counties in northern California	CAD
Quality Oncology ($140)	Disease management provider	Broward County, Florida (Miami)	Cancer
University of Maryland Medical School ($350)	Academic institution	Baltimore	CHF
Washington University School of Medicine ($173)	Academic institution with disease management provider[b]	St. Louis, Missouri	High-risk patients who are clinically unstable

CAD = coronary artery disease; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease.

[a]The program fees for CorSolutions, Health Quality Partners, and Jewish Home and Hospital are blended averages of their payments, which differed depending on whether, for CorSolutions, patients were offered prescription drug coverage, or for the other two programs, on patient severity levels. The other 12 programs' fees are the programs' approved first-year fees for patients following their first month in the programs.

[b]The collaboration with the disease management provider was dissolved in January 2006.

was likely to be the maximum percentage savings the programs would be able to achieve (Chen et al. 2000).

Host Relationships with Physicians. Programs may find it easier to build relationships between physicians and care coordinators if program administrators or care coordinators and physicians have preexisting organizational links, such as the program's host also employing the physicians, or if the physicians are familiar with program staff from some previous joint endeavor. Even for programs that intend to minimize their demands on physicians, organizational links may at least provide name recognition for the program among its patients' physicians when care coordinators contact them. For most of the programs (9 of 15), at least some physicians were employed by the program host or an organization affiliated with the program. Physicians of patients in 12 of the 15 programs had worked previously with program leadership or care coordinators. In only 2 programs were physicians neither affiliated with program hosts nor familiar with staff.

Care Coordinator Education and Caseload Size. All but 1 program required that their care coordinators be registered nurses; however, only 4 required that they be at least baccalaureate trained. (Table 3 presents selected program characteristics.) Thirteen programs also required that the coordinators have specific experience with cardiac, geriatric, medical-surgical, or community nursing. Caseload size varied widely, from a low of 36 patients per coordinator to a high of 155. Care coordinators across the 15 programs contacted patients from just over once a month to twice a week, on average.

Assessment and Care Planning. All programs began care coordination by assessing patients' needs and conditions, after which they developed patient care plans. Of the 15 programs, 10 conducted at least part of the assessment in person. The assessments culminated in care plans designed to fill the gaps in the patients' knowledge and treatment. These plans were developed collaboratively with patients and, when appropriate, their families. Six programs *routinely* included physicians in care planning, either by eliciting their input to the care plans or by asking them to review plans. Two programs consulted physicians on a case-by-case basis, while the others did not involve physicians in care planning.

Monitoring. All 15 programs routinely monitored patients, primarily by telephone. Many (11) also monitored patients in person, but most of those did so infrequently. Routine monitoring included discussion of symptoms and other health issues and the provision of emotional support. Three programs provided all their patients with home telemonitoring devices. The devices transmitted patients' weights, other clinical indicators, and symptom reports to their care coordinators every day. Program staff telephoned patients when devices transmitted out-of-range readings, or if a scheduled transmission was not received. One program provided periodic ambulatory ischemia monitoring.

Education. Education was the cornerstone of most of the programs' interventions. All programs provided patient education, except one whose intervention focused on telemonitoring for quick identification of worsening CHF symptoms. Almost all programs that provided education used standard curricula based on nationally published guidelines. Educational materials were part of some programs' electronic case management systems. Most education was provided telephonically by program care coordinators. Seven programs provided additional

TABLE 3

SELECTED PROGRAM CHARACTERISTICS

	CC Must Be BSN or MSN Prepared	Typical CC Caseload	Mean Number of Contacts per Month	Percentage of Contacts In-Person	Initial Assessment Routinely in Person	Home Telemonitor Used[a]	Education Based on Behavior Change Model[b]	Physicians Expected to Contact CCs	Physicians Routinely Expected to Participate in Care Planning	Program Payment to Physicians
Avera		86	8.2	1.6	✓	✓		✓	✓	$30 pppm
Carle		155	1.4	31.4			✓	✓	✓	For meetings with CCs
CenVaNet		75	1.4	18.1	✓		✓			
Charlestown		60	2.3	31.9					✓	$26 pppm
CorSolutions		145	2.6	3.7	✓[c]		✓			For phone conferences with CCs
Georgetown University	✓	36	5.9	14.1	✓	✓			✓	For in-person conferences with CCs
Health Quality Partners		106	2.2	41.6	✓[d]		✓			
Hospice of the Valley		40	2.5	37.1	✓		✓			
Jewish Home and Hospital	✓	66	2.5	40.2	✓			✓		$28 pppm
Medical Care Development		70	1.5	29.4	✓		✓			$20 pppm
Mercy	✓	50	1.4	69.2	✓				✓	
QMed		150	1.2	7.6				✓		For review of program reports
Quality Oncology		40	n.a.[e]	0.0			n.a.[f]			For provision of medical records
University of Maryland	✓	71	3.9	6.5	✓	✓	n.a.[g]			$100 pppm
Washington University		70	1.2	4.7			✓		✓	

Source: Data on number of contacts and whether contacts were in person were prepared by the programs; statistics refer to the first year after enrollment for patients enrolling during the programs' first years of operations. Other information comes from telephone and in-person interviews and other communications with program staff over the four-year demonstration.

[a]QMed periodically tested its patients with an ambulatory ischemia monitor. CenVaNet, Jewish Home and Hospital, and Mercy used home tele-monitors, but only for a minority of patients for less than four years.

[b]Behavior change and readiness to change models became more popular during the later years of the demonstration. Many of the programs noted did not initially include patient educator training in these methods, but introduced it later.

[c]CorSolutions initially contracted with local home health agencies to conduct part of the initial assessment, but discontinued this practice later in the demonstration.

[d]Health Quality Partners routinely assessed only its high-risk patients in person.

[e]Quality Oncology noted that its care coordinators were not recording all their patient contacts; therefore, this figure is not presented.

[f]Quality Oncology targeted cancer patients. Their education is shorter term and focuses on recognition of adverse treatment effects. Thus, behavior change is not relevant to program teaching.

[g]University of Maryland did not provide patient education; its intervention was the provision of home telemonitoring for patients with CHF.

CC =care coordinator; BSN=baccalaureate degree in nursing; MSN =masters degree in nursing; pppm = per patient per month; n.a. = not available.

training to nurses about how to use behavior change or learning theories. Programs assessed the effectiveness of their education by reviewing clinical indicators or home monitoring data for evidence of improved health, by relying on patients' self-reported changes in behavior or their responses to questions about their knowledge, or by a combination of the two.

Communication and Coordination. Most programs sought to improve communication between patients and physicians by training patients to communicate better or by sending physicians regular written reports. Most program staff believed that training patients or their caregivers to communicate better, rather than primarily coordinating care on behalf of patients, would help prepare the patients to act as their own coordinators once the demonstration was over. To support patients in this activity, some programs taught patients to take prepared lists of questions to their office visits, while others gave patients schedules of tests they should receive.

Most programs tried to minimize their demands on patients' physicians and their office staff, recognizing that the physicians and staff were already busy providing patient care and responding to insurers. Although care coordinators for all the programs contacted patients' physicians directly by telephone to report serious health problems, routine communication for many was primarily via written reports documenting patient health and related issues between office visits. One program (Carle), however, held regular formal conferences with physicians to discuss individual patient health and treatment adherence. Another (QMed) had its quality manager visit physicians to discuss significant deviations from evidence-based practice, using data obtained from ambulatory ischemia monitoring and physicians' medical records as a basis. Five programs had care coordinators practice in the same location as physicians (for example, clinics or hospitals operated by the program host) to increase the possibility of face-to-face communication; another routinely sent its care coordinators along on some patient office visits. Ten of the programs paid physicians either a monthly stipend per patient (typically $20 to $30, although one paid $100) or a fee for participating in meetings or for sharing medical records.

Timely notification of adverse events, such as emergency room visits or hospital admissions, is crucial to a care coordinator's ability to assist the patient effectively. This information enables care coordinators to ensure that the patient understands discharge instructions, to review new medications the patient may have been given by hospital staff and to investigate the potential for polypharmacy problems, and to facilitate the receipt of and smooth the transition to post-discharge home care or other services. Nine programs had procedures to learn about hospitalizations quickly either by having hospitals notify program staff when they admitted its patients, by having program staff review hospital and emergency room admission lists, or by following up on missing telemonitoring reports. The other six programs likely had less timely notification since they relied on patients or their caregivers to call the program about admissions or only learned about admissions during routine monitoring contacts.

Most programs also sought to improve patient health by increasing access to health-related goods or support services, although improving access to such care was not a focus and few patients received this help. Most typically programs referred patients to or arranged for: home care, home-delivered meals, transportation, or low-tech monitoring devices such as bathroom scales. This arranging was usually done by referring patients to service providers, but eight programs paid for goods or services for the small number of patients who needed such assistance.

Electronic Data Systems. All programs used electronic case management systems, although the sophistication of their systems varied widely. Most of the programs had no direct access to patients' medical records or clinical data, as is frequently the case in the fee-for-service sector. Thus, they relied on patient self-reports for clinical data they felt they needed for care management, as well as the data the care coordinators themselves generated from patient assessments and ongoing monitoring.

Six programs purchased commercial case management software, while the 4 whose hosts (or in the case of Washington University, the host's collaborating partner) were disease management providers used the systems developed for commercial clients. Four used systems they developed themselves. The remaining program, whose intervention focused on home telemonitoring, relied on data from that device. Care coordinators from 12 programs regularly used their data systems to support their work with patients. (The other 3 programs used their systems only for operational oversight.) Eleven of the 15 programs generated reports from their data systems reminding coordinators about when to contact patients, and 13 of 15 generated reports on patients' clinical indicators and outcomes.

Program Intervention Summary Index. The evaluation included the development of a classification system, based in part on Chen et al. (2000), to quantify the intensity and comprehensiveness of program efforts in each of ten care coordination domains. Using site visit interview notes and program documents, two raters blinded to the results of the impact analysis independently completed structured forms to assign numeric indices (on a 1 to 100 scale) to each of the programs for the following domains:

Program staffing	Improving provider practice
Initial assessment	Service and resource arrangement
Problem identification and care planning	Information technology and electronic records
Patient education	Ongoing monitoring
Improving communication and coordination	Quality management and outcome measurement

The consistency of the two raters in assigning numeric indices across the ten domains was generally moderate to excellent, as assessed by commonly used methods for measuring inter-rater reliability. The raters' indices for each domain were then averaged, and within each domain the 15 programs were rank ordered by their average index values and divided into quintiles (that is, the first quintile consisted of the 3 programs with the highest index values for that domain, then the second quintile comprised the next 3 programs, and so on). This classification system provided the evaluation with a summary measure of the strength or focus of each program in each domain, relative to the other 14 programs. Appendix A provides the ratings of individual programs on each domain. (See Brown et al. 2007 for a more detailed description of the classification methodology.) Section G below synthesizes the findings from the classification system with program impact estimates in order to identify program components that might merit replication, as directed by the demonstration's mandating legislation.

C. BENEFICIARIES SERVED

Collectively, the programs enrolled a total of 18,309 beneficiaries in the research sample through June 30, 2005, the cutoff date used for inclusion in this report (Table 4). Despite earlier difficulties recruiting patients (Brown et al. 2007), 12 of the 15 programs redoubled their recruitment efforts and enrolled substantially more than the target of 686 beneficiaries that MPR set as the minimum necessary for the evaluation to detect policy relevant effects. Having close relationships with physicians before the demonstration began and having access to databases, such as clinic or hospital records, that could be used to identify potentially eligible beneficiaries distinguished the programs with the most success in recruiting patients. The other 3 programs each enrolled under 250 beneficiaries, which limits the likelihood that the evaluation can observe statistically significant treatment-control differences, even if the programs actually had impacts, unless the true impacts were quite large.

The programs enrolled a diverse population of generally high-cost beneficiaries with chronic illnesses.

1. Overall, the most commonly occurring chronic conditions among enrollees during the two years prior to enrollment were coronary artery disease (63 percent), congestive heart failure (53 percent), diabetes (40 percent), and chronic obstructive pulmonary disease (33 percent). Other conditions experienced by sizable proportions of the sample include cancer (25 percent), stroke (25 percent), depression (18 percent), asthma (17 percent), peripheral vascular disease (15 percent), and dementia (9 percent).

2. Three programs (Charlestown, Hospice of the Valley, and Jewish Home and Hospital) drew one-fourth or more of their enrollees from the "oldest old" (beneficiaries who were age 85 or older), and three (CorSolutions, University of Maryland, and Washington University) targeted and enrolled a high proportion (14 to 27 percent) of younger beneficiaries with disabilities.

3. Compared with all Medicare beneficiaries, patients in the programs were more educated and had higher incomes.

4. Overall (with a few notable exceptions), the programs enrolled few black or Hispanic beneficiaries, few patients under age 65, and few who were also enrolled in Medicaid.

Most of the programs succeeded in enrolling patients with serious chronic illnesses, but a few enrolled relatively healthy patients. Monthly Medicare expenditures averaged more than $2,000 during the two years preceding enrollment in 4 programs, between $1,000 and $2,000 for 6 programs, between $600 and $999 for 2, and less than $600 for 3 others (average Medicare expenditures for beneficiaries nationally were $566 per month in 2003 (calculations based on Table 15, Health Care Financing Review 2006). In 9 of the 15 programs, enrolled patients had an average of one or more hospitalizations per year during the two years prior to enrollment.

TABLE 4

BASELINE CHARACTERISTICS OF RESEARCH SAMPLE RANDOMIZED THROUGH JUNE 2005
(Percentages Unless Otherwise Noted)

Program	Number in Research Sample Randomized Through June 2005	Diagnosis[a]											Age			Medicare Use During Two Years Prior to Randomization	
		CAD	CHF	Diabetes	COPD	Cancer	Stroke	Depression	Asthma	PVD	Dementia	ESRD	Average Age	≤64	≥85	Average Annualized Number of Hospitalizations	Average Monthly Expenditure ($)
Avera	856	76.8	98.8	40.4	43.9	23.7	22.0	15.4	16.6	15.2	4.1	0.7	78.5	0.0	20.0	1.6	1,314
Carle	2,650	49.0	26.7	40.2	23.3	21.0	13.5	13.3	15.8	8.5	5.2	0.2	75.5	2.0	11.3	0.6	529
CenVaNet	1,441	73.9	48.1	51.1	28.0	27.3	26.4	10.8	15.5	14.5	4.8	0.1	77.2	0.0	11.9	0.8	784
Charlestown	830	65.7	44.7	25.9	38.9	32.7	32.2	18.4	17.1	52.0	9.4	0.4	83.8	0.0	43.5	0.8	959
CorSolutions	2,624	84.3	98.2	55.5	50.2	17.2	40.8	22.2	22.6	18.2	12.4	0.3	74.0	14.7	12.5	2.1	2,291
Georgetown	228	82.5	99.1	54.8	40.8	23.2	28.5	13.6	24.1	18.0	13.6	0.0	76.8	1.8	15.4	2.5	2,396
Health Quality Partners	1,464	35.2	11.3	24.5	13.1	22.0	14.4	8.2	10.9	6.2	1.8	0.0	75.0	0.0	7.0	0.4	451
Hospice of the Valley	1,039	62.5	55.0	31.0	50.8	31.3	36.2	24.0	24.4	20.2	25.3	0.5	80.1	0.0	26.8	1.3	1,622
Jewish Home and Hospital	868	50.4	38.0	34.0	22.7	29.6	29.1	34.0	16.3	22.0	36.8	0.5	82.4	0.1	38.4	0.9	1,438
Medical Care Development	1,313	90.3	63.4	43.3	36.2	19.0	18.9	19.0	14.3	12.3	2.9	0.4	75.0	6.1	11.3	1.5	1,492
Mercy	930	65.2	61.7	33.3	53.0	23.7	26.8	24.5	16.7	15.9	6.9	3.9	77.6	4.2	17.2	1.1	1,064
QMed	1,404	49.7	40.0	26.3	14.2	20.1	14.0	9.4	9.5	6.9	1.6	0.1	73.7	8.5	5.1	0.4	517
Quality Oncology	210	46.9	19.1	25.4	35.4	99.0	15.8	11.5	14.8	8.6	5.7	0.0	75.3	7.2	12.9	1.0	2,337
University of Maryland	178	78.1	94.9	45.5	40.4	12.4	25.8	12.4	13.5	19.1	5.6	0.0	72.8	13.5	5.1	2.0	2,594
Washington University	2,274	55.8	42.2	42.6	32.2	35.7	24.1	24.3	17.3	13.6	11.6	8.1	70.0	26.7	9.9	1.6	1,899
All Programs	**18,309**	**63.2**	**52.5**	**39.7**	**33.3**	**25.3**	**24.5**	**17.9**	**16.7**	**15.2**	**9.3**	**1.4**	**75.7**	**7.2**	**14.9**	**1.1**	**1,284**
Medicare Total	**42,320,000**	**40.2[b]**	**40.2[b]**	**12.0**	**20.7**	**16.9[c]**	**n.a**	**n.a**	**15.2**	**n.a.**	**5.0[d]**	**n.a.**	**n.a.**	**14.5**	**11.7**	**0.4**	**566**

Sources: Medicare National Claims History File, Standard Analytic File, and Enrollment Databases for estimates of demonstration enrollees; Medicare Current Beneficiary Survey (MCBS) 2003, Health Care Financing Review (2006), and 100 percent Denominator File 2004 for "Medicare Total" estimates.

[a]Medical conditions treated during the two years before randomization, as reported in Medicare claims data.

TABLE 4 *(continued)*

[b]Data available only for Medicare beneficiaries living in the community, with "heart disease," which includes both CAD and CHF; included for comparison purposes only.

[c]Excludes skin cancer.

[d]Includes only beneficiaries with Alzheimer's disease.

CAD = coronary artery disease; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; ESRD = end-stage renal disease, PVD = peripheral vascular disease; n.a. = not available.

Three of these programs averaged two or more hospitalizations per patient per year. As expected under random assignment, the treatment and control groups had similar preenrollment characteristics (not shown, available on request).

D. PATIENTS' AND PHYSICIANS' SELF-REPORTED EXPERIENCES WITH THE PROGRAMS

Survey data collected on patients in the 12 programs with over 300 enrollees by the end of their first year and on the patients' physicians in all 15 programs suggest that the programs are popular with both patients and physicians. Enrollees were interviewed between 7 and 12 months after enrollment.

About two-thirds of treatment group patients received care coordination. Two-thirds of patients across all programs combined reported that "someone like a nurse, social worker, or geriatric nurse" had helped arrange or coordinate their health care. There was wide variation across programs, from only 30 percent in QMed saying they had received such help to 81 percent in Mercy. Fifteen percent of control group members reported having received assistance of this type, with the incidence ranging from 3 to 28 percent across programs. With one in three treatment group members stating that they had received no care coordination, and one in seven control group members stating that they had, programs faced challenges in demonstrating intervention impacts. Among those reporting having received care coordination, however, treatment group members were more satisfied than control group members with the help received.

Treatment group members rated the programs and care coordinators highly in two of four areas. Patients were asked to rate programs on four dimensions—(1) support and monitoring, (2) health education, (3) service arrangement, and (4) adherence assistance. Patients liked the support and monitoring they received. In the four measures for this dimension, 60 percent of programs' patients, on average, gave excellent ratings for coordinators' demonstrating a "caring attitude", and over 50 percent gave excellent ratings for coordinators' "ability to stay in touch."[4] The average percentages of excellent ratings for the other two measures—including patients and their families in decisions, and helping patients cope with illness and avoid complications—were lower at around 45 percent.

In the health education dimension, patients thought their nurse case managers were knowledgeable, one of the four measures in this dimension. The average proportion giving excellent ratings for care coordinators' knowledge was over 50 percent, and in only two programs did less than 43 percent of patients give excellent ratings. The average proportions giving excellent ratings for the other three measures—ability to get answers from the physician, ability to explain medical terms, and ability to explain warning signs—were around 40 to 43 percent. Jewish Home and Hospital (with 20 to 25 percent) and QMed (with 23 to

[4]These average proportions were calculated at the *program* level. Each of the 12 programs counted equally regardless of size; in other words, a program with relatively few enrollees counted the same as one with many enrollees.

27 percent) were in the lower range for these three measures. Avera (44 to 62 percent), Carle (45 to 57 percent), and Health Quality Partners (49 to 67 percent) were in the higher range.

Reflecting less program focus on service arrangement, programs received lower ratings for the measures in the service arrangement dimension (helping to make appointments, helping to arrange payment for noncovered services, and recommending community services), with percentages with fair/poor ratings ranging from 10 to 24 percent, and percentages with excellent ratings ranging from 35 to 40 percent.

Finally, the average proportions of excellent ratings for care coordinators' help in adhering to exercise were 35 percent, diet 38 percent, and medication 43 percent. These results suggest that care coordinators devoted slightly more effort to counseling about medication adherence than about other behaviors. Few patients rated the programs as fair or poor in these three dimensions, though.

Patient survey results were generally in accord with results from the evaluation's program classification algorithm. On several dimensions, Health Quality Partners received consistently higher ratings from their patients than the other programs, in agreement with its highest ranking on the scoring on patient education and its high ranking on patient monitoring. Avera's high satisfaction rating on explaining early warning signs is consistent with its third-highest ranking in the classification algorithm on patient monitoring (which reflected, in part, its use of home telemonitoring). Carle's high ratings from patients on getting answers from physicians matches its top ranking on improving communications and coordination among providers and its closer integration with patients' physicians. Carle's high ratings on service arrangement also agree with its placement by the classification algorithm as among the top three programs in that area.

Physicians of treatment group members were generally happy with the programs, but some physicians disliked certain aspects. Physicians treating treatment group patients in all 15 programs were interviewed about several program dimensions, including effects on physicians' practices (clinical care, time and paperwork burden, and financial impact), patients' knowledge and behavior, service arrangements for patients, care coordination, physicians' relationship with patients, and patient outcomes. Physicians widely agreed that the programs made things easier for the physicians' office staff and did a good job of monitoring and followup. Sixty-seven percent of physicians, on average, felt that the program increased patients' overall quality of care, and 80 percent said they would recommend the program to patients and colleagues (60 percent "definitely" and 20 percent "probably").

There were large differences of opinion across the programs about some program features, though. The proportion of physicians rating program reports on patients as "very useful" ranged from 0 to 91 percent (mean 42 percent); polypharmacy problems as "better" from 11 to 81 percent (mean 56 percent); physician and staff telephone time as "better," from 4 to 88 percent (mean 55 percent), and overall work of caring for patients as "easier," from 33 to 100 percent (mean 75 percent). There was similarly wide variation across the 13 other specific dimensions (Brown et al. 2007).

Physicians' opinions also differed across programs. Only 11 percent of physicians said that Quality Oncology improved patients' quality of care and that they would definitely recommend the program to others. In contrast, 95 percent of physicians felt the Charlestown program improved patients' quality of care and would definitely recommend the program to others.

Charlestown consistently received higher physician ratings than did other programs, while three programs (CenVaNet, QMed, and Quality Oncology) consistently received lower ones.

Ratings of physician satisfaction generally corresponded with the results of the program classification algorithm. For example, Carle and Charlestown, whose physicians' ratings for "effects on their practices" were consistently higher than those of the other programs, were also in the top quintile for "improving provider practice," with Carle ranking the highest. Mercy's program received higher physician ratings than the other programs on perceived service arrangement and care coordination effects, consistent with its ranking in the top two quintiles for the categories of service and resource arranging, and for improving communication and coordination. At the other end of the spectrum, QMed and Quality Oncology had, across all categories, overall physician satisfaction ratings consistently lower than the cross-program average by more than one standard deviation, which coincides with their placement in the bottom two quintiles by the program classification algorithm.

There were no treatment-control impacts on patients' satisfaction with health care. Although both treatment group members and their physicians generally rated the programs favorably, there were no consistent treatment-control differences in patients' satisfaction with various health care experiences. The evaluation interviewed both treatment and control group members on their perceptions of their freedom of choice in treatment; providers' keeping in touch with each other; explanations from specialists; discussions of treatments, side effects, and tests; and the timeliness of receiving test results. In only 1 of the 12 programs in the patient survey (Avera) were treatment group members significantly more likely than controls to report feeling that they had choice in the treatment of their condition. Differences favoring the treatment group occurred most often for providers' keeping in touch with each other (5 of the 12 programs: Avera, Carle, Charlestown, Health Quality Partners, and Mercy). Treatment group members in 4 of the programs (Hospice of the Valley, Health Quality Partners, Mercy, and QMed) also gave more favorable ratings for explanations of treatments. Satisfaction with explanations of side effects and explanations from specialists were significantly greater for the treatment than the control group for only 3 (Carle, Hospice of the Valley, and Mercy) and 2 (Avera and Medical Care Development) of the programs, respectively, and with explanation of tests for only 1 (Avera). None of the programs led to treatment-control differences on how promptly patients received test results.

A few programs appeared to have greater impact overall than others on patients' satisfaction with care. Avera's and Mercy's treatment groups each gave significantly higher ratings than their control groups on three of the six measures. Three programs (Carle, Health Quality Partners, and Hospice of the Valley) had significant differences on two of the six, and two (Charlestown and QMed) had significant effects on one of the measures. The four other programs in the survey had no discernible effects on patients' satisfaction.

E. PROGRAM EFFECTS ON QUALITY OF CARE

As shown in Figure 1, the programs were expected to improve patients' self-care behavior and adherence to treatment recommendations, and to coordinate care to meet patients' service needs, thereby increasing patients' health, functioning, and well-being. The evaluation compared the treatment and control groups' survey responses on receipt of health education, self-care knowledge and behavior, unmet needs, and health status and well-being to determine whether the programs had the intended effects. Table 5 summarizes the results. The results in the first five columns were reported in the second report to Congress (Brown et al. 2007); the last two columns combine findings previously reported in the second report to Congress (from the patient survey) and more recent analyses of Medicare claims data.

Overall, the programs appeared to have no consistent discernible effect on numerous measures of behaviors and outcomes except receipt of health education.[5] While there were isolated treatment-control differences for a few outcomes for a few programs, there was no pattern suggesting that the programs, as a group or individually, had sizeable or pervasive effects in any area besides receipt of health education. Across the 51 other measures, 6 of the programs had a few more significant favorable estimates than the 2 or 3 that would be expected by chance (Carle, CenVaNet, CorSolutions, Hospice of the Valley, Health Quality Partners, and QMed).

The large effects on health education did not lead to effects on self-reported knowledge, adherence, or health-related behaviors. Despite more of the treatment group members reporting they had received health education, there were no effects for any of the 12 programs on patients' self-reported adherence to diet, exercise, or taking medications. Only scattered favorable effects were observed on self-reported understanding of healthy behaviors, but these were too sporadic to suggest meaningful effects for all but 1 or 2 programs. Across the measures, 2 programs (CenVaNet and Health Quality Partners) had somewhat more favorable treatment-control differences than the other programs.

Only two programs (CorSolutions and Hospice of the Valley) had effects on multiple measures of patient well-being, and these were limited. The treatment groups in those two programs were significantly more likely to report feeling their condition placed less of a burden on family than were the control groups (both programs), feeling calm and peaceful (in CorSolutions only), and having less pain (in Hospice of the Valley only). However, even these two programs had a favorable effect on only two or three of the eight measures of well-being that were examined. (Six programs each had significant effects on one of the eight measures.) In addition, only three programs (Medical Care Development, Mercy, and QMed) had a favorable

[5]See Appendix Table B.1 for a list of the survey measures included in each category in Table 5.

TABLE 5

SUMMARY OF TREATMENT-CONTROL DIFFERENCES IN QUALITY OF CARE

Category of Outcomes (Number of Measures)	Health Education (5)	Knowledge and Behavior (8)	Service Arrangement and Unmet Needs (7)	Functional Status (9)	Health Status and Well-being (8)	General and Disease-Specific Preventive Care (11)	General and Disease-Specific Potentially Preventable Hospitalizations[a] (8)
Avera	●●	○		○	●	○	●
Carle	●●●●			○○○		●●●●●●●	
CenVaNet	●●●●●	●			●	●●●●	●○
Charlestown			●	○	○	●●	●○
CorSolutions	●●		●●	○	●●●	●●	●
Georgetown University						●[b]	●
Hospice of the Valley	●		●		●●	○	●●●
Health Quality Partners	●●●●●	●●	●		●	●●○	
Jewish Home and Hospital				○	●	●	○○
Medical Care Development	●●			●	○		○
Mercy	●●●●●		○	●	●		
QMed	●	●○	○	●	●	●	●○
University of Maryland						[b]	●
Washington University	●○			○			○

Source: Treatment-control differences from patient survey and Medicare claims data (over a two-year follow-up period for enrollees randomized through June 2004). The Georgetown University and University of Maryland programs did not have sufficient numbers of enrollees to be included in the patient survey, and so the survey-based measures are shaded for these two programs. The measures summarized in this table were not appropriate for the Quality Oncology program, which focused on cancer patients, so no survey or claims-based analyses of quality were conducted for that program. The claims-based indicators were calculated over a two-year followup period for beneficiaries enrolled through June 2004, with observations for which no event was observed weighted to reflect the number of months of exposure.

Note: ● = Treatment-control difference favoring the treatment group, significant at the 10-percent level.
○ = Treatment-control difference favoring the control group, significant at the 10-percent level.
For example, the summary results of ●●○ for the Health Quality Partners program for the General and Disease-Specific Preventive Care domain, which has 11 separate measures, indicates that at p=0.10 level, the program had 2 measures favoring the treatment group, 1 favoring the control group, and the remaining 9 not statistically significant.

[a]Potentially preventable hospitalizations are those that should be avoidable with close patient monitoring, early detection of illness, and prompt, appropriate outpatient treatment. Those that are "general" are ones that all patients might be prone to (such as hospitalizations for CHF, dehydration, or urinary tract infections). Those that are disease-specific are analyzed among patients with specific diagnoses, for example, hospitalizations for diabetes complications among patients with diabetes.

[b]Of the 11 general and disease-specific preventive care measures, 3 are from patient survey data and 8 from Medicare claims data, so only the 8 claims data-based measures were analyzed for the Georgetown and University of Maryland programs.

treatment-control difference on any of the nine survey-based measures of functioning (for example, ability to eat independently), and for six programs (Avera, Carle, Charlestown,

CorSolutions, Jewish Home and Hospital, and Washington University), the treatment group reported significantly worse health status on one or more measures. However, it is difficult to conceive of a mechanism by which the demonstration programs would adversely affect patients' functioning. Furthermore, one should expect about half the sites to have one significant negative estimate of the nine measures used just by chance. Finally, there is no evidence of adverse effects on other health outcomes. Thus, these scattered treatment-control differences showing worse functioning for the treatment group than the control group are interpreted as chance differences, rather than as evidence that six of the programs have caused patients' functioning to decline.

There were favorable impacts on preventive care. Favorable program impacts for all programs combined were observed on several of the 11 preventive care measures. Impacts for the combined samples were favorable and statistically significant for four of the six disease-specific measures (urine tests for protein and hemoglobin A1c tests for patients with diabetes, and blood lipid tests both for patients with CAD and those with diabetes), with a few positive effects of modest size in general preventive care (such as flu shots and mammography). The Carle program had seven significant differences favoring the treatment group, one for general care and six for disease-specific care. The CenVaNet program had effects on four measures, two for general care and two for disease-specific care. There were a few other positive differences for both general care and disease-specific measures scattered across the other programs.

There was evidence that some types of potentially preventable hospitalizations were reduced. The results suggest that the programs may have led to decreases in some types of hospitalizations that should be avoidable with close monitoring of patient symptoms, early detection of illness, and prompt, appropriate outpatient treatment (Table 5). Across all programs there were more statistically significant differences favoring the treatment group for CHF hospitalizations among patients with CHF, and for CHF hospitalizations among patients with diabetes (a common problem among patients with diabetes) than would have been expected by chance alone, with the differences spread across eight programs.[6] Treatment-control comparisons in which sample members with CHF from all programs were pooled together also showed a small (3 to 6 percentage points, or about 17 percent of the control group mean) but statistically significant difference favoring the treatment group in the proportion with a preventable hospitalization for CHF. The treatment-control differences suggest favorable program impacts on the proportion with potentially preventable CHF hospitalizations for Avera, CorSolutions, Georgetown, Hospice of the Valley, and QMed.

One program may have reduced mortality. Mortality rates for the first two years after enrollment ranged from 6 percent to 46 percent across the 15 programs (not shown), but the treatment-control difference was statistically significant for only two of the programs. Georgetown's treatment group had a mortality rate significantly lower than that of its control

[6] The multiple comparisons of the treatment and control groups across 14 programs and 3 different cohorts of patients with varying lengths of follow-up meant that some differences might have appeared statistically significant by chance alone.

group, but for Jewish Home and Hospital, the treatment group had a significantly higher mortality rate than the control group. Furthermore, neither of these differences was statistically significant when examined over the first year after followup. Thus, the observed differences are likely to be due to chance rather than true program effects. When all programs are combined, the difference in mortality rates was small (less than one percentage point) and not statistically significant.

F. PROGRAM EFFECTS ON MEDICARE HOSPITAL USE AND COST

By improving patient adherence, the timeliness of response to worsening symptoms, and other aspects of the quality of care, care coordination programs are expected to reduce hospitalizations, the key factor in cutting Medicare expenditures for beneficiaries with chronic illnesses. To measure these effects, the evaluation compared the post-enrollment Medicare hospital use and expenditures of the treatment and control groups in each program, and among all 15 programs combined. Effects are estimated separately for each program because the demonstration tested 15 very different program models. Combining the programs and estimating effects for the overall demonstration provides a meta-analysis of the 15 programs, and increases the evaluation's ability to detect small effects. This is important because it makes it more likely that the evaluation will detect moderate-sized effects that would be too small to detect in individual programs due to small samples and large variation in the outcomes.

Three outcomes were examined: (1) the average number of hospitalizations (which account for the largest share of costs); (2) the average Medicare expenditures per month in fee-for-service, without the care coordination fee; and (3) the average Medicare expenditures per month in fee-for-service, including the care coordination fee.

Impacts were examined using data from Medicare claims with dates of services through June 2006, for beneficiaries who enrolled between the program's start date in 2002 and June 30, 2005. This period spans a maximum of between 46 and 51 months of operations for each program, depending on when the program began enrollment. Beneficiaries are observed over the total number of months from the first full month after they enrolled through June 2006. Reflecting the staggered enrollment and, for some beneficiaries, disenrollment or ineligibility, the average number of months of followup per beneficiary is 29 across the programs, and ranges from 19 to 36.

Results are regression adjusted to account for any chance preenrollment treatment-control differences in age, gender, race, reason for entitlement, presence of 10 chronic conditions, prior expenditures, prior use of home health care, skilled nursing facility, or hospital services. Because the outcomes, especially Medicare expenditures, are highly variable, there can be sizable treatment-control differences due to chance. Statistical tests of significance that take into account the size of the estimate, the sample size, and the degree of variation in the values are used to determine whether a difference is likely due to the program. The program results are arrayed by sample size, to make clear for which programs the estimates are the most precise and reliable.

1. Overall, There Were Modest Effects on Hospitalizations, Driven by Sizable Differences in Five Programs.

Across all the programs combined, the treatment group had slightly fewer hospitalizations per year than the control group. The programs combined reduced hospitalizations by 4.5 percent per year (p=0.02, Table 6).[7]

Most of the programs had no effects on hospitalizations. The exceptions were Mercy and Georgetown University: Mercy reduced hospitalizations by 17 percent, and Georgetown University by 24 percent (p=0.02 and 0.06, respectively).

The treatment groups in three other programs (Hospice of the Valley, Health Quality Partners, and University of Maryland) had 10 to 15 percent fewer hospitalizations than the control groups; but because the differences were not statistically significant, they may have been due to chance. The differences were 11 percent (p= 0.12), 14 percent (p= 0.13), and 16 percent (p = 0.35), respectively.

2. Effects on Medicare Part A and B Expenditures Without Fees Were Limited.

Only one program reduced Part A and B expenditures. When observations from all the programs are combined, there was no significant difference between the treatment and control groups in average monthly Medicare Part A and B expenditures (Table 6). Like results for hospitalizations, most (11) of the programs had no effects on Medicare expenditures (without program fees), but there were a few exceptions. Health Quality Partners is the only program that clearly reduced expenditures. The treatment group's expenditures were 14 percent lower than those of the control group, and the results are statistically significant (p=0.07). This is consistent with its treatment group having had 14 percent fewer hospitalizations than the control group (p=0.12) and suggests that the program did reduce fee-for-service expenditures.

Despite significant reductions in hospitalizations for Mercy and Georgetown University, the evidence for a reduction in Medicare expenditures is weak for both programs. This may be due to the large variance in Medicare expenditures and the smaller proportional treatment-control difference in costs than in hospitalizations. Mercy's treatment group had 9 percent lower expenditures than the control group (p=0.13), whereas the reduction in hospitalizations was

[7] Appendix Tables B.2–B.4 contain more detailed tables on the hospitalization and Medicare expenditure outcomes.

TABLE 6

TREATMENT-CONTROL DIFFERENCES IN HOSPITALIZATIONS AND MEDICARE EXPENDITURES, WITH AND WITHOUT PROGRAM FEES THROUGH JUNE 2006 AMONG BENEFICIARIES ENROLLED THROUGH JUNE 2005

	Sample Size Through June 2005				Impact (As Percentage of Control Group Mean)			
						Monthly Medicare Expenditures[b]		
	Treatment Group	Control Group	Average Number of Follow-up Months Through June 2006	Average Fee Received Per Member Per Month[a]	Annual Number of Hospitalizations	Excluding Care Coordination Fees	Including Care Coordination Fees	Cost Neutral
1,100 or More Treatment Group Members Enrolled Through June 2005								
Carle	1,335	1,315	35.9	$148	–0.2	5.8	26.7†	No
CorSolutions	1,496	1,128	24.1	$192	–3.0	1.0	8.3†	No
Washington University	1,142	1,132	28.3	$160	–2.2	3.7	12.2†	No
415 to 725 Treatment Group Members Enrolled Through June 2005								
Avera	430	426	24.2	$273	2.3	–0.9	19.4†	No
CenVaNet	721	720	34.1	$70	7.2	4.9	13.1†	No
Charlestown	413	417	29.2	$218	12.3	15.2*	36.5†	No
Health Quality Partners	739	725	29.1	$102	–13.6	–13.9*	0.3	Yes
Hospice of the Valley	526	513	19.2	$179	–10.7	–0.6	8.1†	No
Jewish Home and Hospital	433	435	29.7	$221	4.4	4.4	17.1†	No
Medical Care Development	661	652	25.3	$126	–3.3	–6.5	2.6	Unlikely
Mercy	463	467	31.4	$248	–17.0*	–9.4	11.3†	No
QMed	706	698	36.7	$81	–7.4	–10.5	–0.2	Possibly
Fewer than 115 Treatment Group Members Enrolled Through June 2005								
Georgetown University	114	114	26.7	$242	–24.1*	–13.3	–3.7	Yes
Quality Oncology	106	104	17.1	$64	–1.8	–3.9	–1.9	Unlikely
University of Maryland	92	86	22.6	$277	–16.1	30.8	41.5	No
Overall	**9,377**	**8,932**	**29.1**	**$155**	**–4.5***	**–0.1**	**11.3†**	**No**

Note: Numbers marked by a * symbol denote statistically significant treatment-control differences at the 10 percent level for hospitalizations and Medicare expenditures without fees. Differences in expenditures including care coordination fees were tested at the 20 percent level; the symbol † denotes differences that are statistically significant at the 20 percent level. Negative estimates imply that hospitalizations or Medicare expenditures (with or without the fee included) are lower for the treatment group, a favorable outcome.

TABLE 6 *(continued)*

[a] The average care coordination fee paid per month is the total paid to the program for enrolled beneficiaries over the followup period, divided by the number of beneficiary months included in the study (that is, months during which the beneficiaries were alive and met CMS's demonstration-wide eligibility criteria). It differs from the negotiated fee because sample members who disenrolled from the program were retained in the study as long as they remained eligible.

[b] Sample member observations are weighted according to the proportion of the followup period the individual met CMS's demonstration-wide eligibility criteria (alive, in fee-for-service, have both Part A and Part B coverage, and have Medicare as primary payer).

17 percent. For Georgetown, the difference in Medicare expenditures was 13 percent and not statistically significant (p=0.34), compared with a 24 percent reduction in hospitalizations.

While the treatment-control differences in the number of hospitalizations was sizeable (though not statistically significant) for Hospice of the Valley and University of Maryland, the differences in Medicare expenditures were essentially zero for Hospice and favored the control group for Maryland. Thus, we conclude that neither of these programs had favorable effects on costs or hospitalizations.

It appears that one program *increased* Medicare Part A and B expenditures. For Charlestown, expenditures per year for the treatment group were a sizable 15 percent higher than for the control group (p=0.08).

3. The Effects on Cost Neutrality Were Mixed: 10 Programs *Increased* Total Costs Including Program Fees, and 5 May Have Been Cost Neutral.

One of CMS's goals is to determine whether the programs were cost neutral, that is, whether the programs generated reductions in traditional Medicare Part A and B expenditures large enough to offset the costs of providing care coordination. One might think that only a program that was shown to have significantly reduced Part A and B costs could possibly be cost neutral. However, interpreting the estimates is more complicated. Because the sample sizes are relatively small and the variance of expenditures is large, the estimates of program effects on Part A and B expenditures have a relatively large confidence interval around them. In addition, some programs have a fee that is a relatively small percent of the control group mean. Thus, the savings in Part A and B would not have to be very large, in percentage terms, to be sufficient to cover the fee. For some programs, these combined factors make it difficult to conclude with much statistical precision whether the programs actually reduced Part A and B expenditures, and if so, whether it was by enough to offset the fee. The 90 percent confidence interval around the estimated treatment-control difference in Part A and B expenditures includes zero for all but one program (Health Quality Partners), which would lead us to conclude that these programs did not generate a statistically significant reduction in Part A and B expenditures. However, the confidence interval for six programs (discussed below) also includes savings amounts that would cover the program fee, so statistical tests do not enable us to conclude with confidence that the program increased net costs to CMS. That is, the program may have been cost neutral, even though we cannot be confident that it really did reduce Part A and B expenditures.

This ambiguity, while confusing, properly represents the uncertainty about the results. The most conservative approach is to conclude that, unless the treatment-control difference in traditional Medicare expenditures was significantly different from zero, there would be no need to conduct the additional test to determine whether it was significantly larger than or equal to the fee. However, that approach could lead to an erroneous conclusion, especially in those programs for which the fee received is small in comparison to the control group mean for Medicare expenditures.

The evaluation addresses this dilemma in two ways. First, it tests whether the total Medicare cost difference between the treatment and control groups, including the care coordination fees, is significantly different from zero at a 20 percent significance level instead of the traditional 10 percent level. This approach boosts the statistical power of the analysis: it increases the likelihood of correctly concluding the program increased net costs when it really did so. However,

it also increases the likelihood of incorrectly concluding that net costs were increased when they actually were not. In practice, the only effect of using a statistical significance level of 20 percent instead of the traditional 10 percent is to shift our classification of one program—Hospice of the Valley—from "possibly cost neutral" to "not cost neutral." Second, given the potential uncertainty about the estimates, to determine whether program effects appear to be large enough to cover the cost of the fees, the evaluation relies on a combination of findings about Part A and B expenditures, total expenditures (including program fees), and effects on hospitalizations. While sizeable effects on hospitalizations are not necessary for generating the small impacts needed to cover the cost of the fees for most of the programs, reductions in hospital use would be by far the most likely way in which programs could generate Medicare savings.

Overall, the demonstration increased total costs by 11 percent. When all the programs are considered together, the demonstration was not cost neutral. Total expenditures for the treatment group, including the program fees, are actually *higher* than expenditures for the control group by 11 percent ($p<0.001$) (Table 6)—approximately, the amount of the fees paid.

While 10 of the 15 programs increased total Medicare expenditures, 2 generated enough savings in traditional Medicare expenditures to offset the program fees, and 3 others may have been cost neutral. Ten programs clearly *increased* total Medicare expenditures (Table 6). The treatment groups in 9 programs had statistically higher expenditures than the respective control groups: Carle, CorSolutions, Washington University, Avera, CenVaNet, Charlestown, Hospice of the Valley, Jewish Home and Hospital, and Mercy (despite the strong evidence that Mercy reduced hospitalizations and perhaps Medicare expenditures excluding program fees). The treatment group's total expenditures (including the care coordination fees) for the University of Maryland's program were much (42 percent) larger than the control group's (Table 6), and there was no effect on hospitalizations, so we conclude that this program was not cost neutral. We cannot formally reject the hypothesis that the expenditures for the University of Maryland's treatment and control were equivalent, but this is likely to be due to the small sample size for this program.

There is strong evidence that two of the other five programs were cost neutral—Health Quality Partners and Georgetown University. Health Quality Partners' treatment group had fewer hospitalizations and lower Medicare Part A and B expenditures than the control group. The difference in hospitalizations—14 percent—was just above conventional levels of statistical significance ($p=0.12$), the reduction in Part A and B expenditures of 14 percent was statistically significant ($p=0.07$), and there was no difference in total expenditures including program fees (0.3 percent; $p=0.96$). Georgetown University's treatment and control group costs, including the program fees, were statistically comparable (3.7 percent lower for the treatment group; $p=0.79$), and the evidence of a large reduction in hospital use of 24 percent was strong ($p=0.06$). While the estimated treatment-control difference in Medicare Part A and B expenditures was not statistically significant, it was large enough, at $335 per month, to more than offset the monthly program fee paid of $242.

The remaining three programs—QMed, Quality Oncology, and Medical Care Development—had statistically comparable treatment and control expenditures including program fees. The very small positive or negative net treatment-control differences (-1.9 to +2.6 percent) *could* indicate that they are cost neutral, or they could simply be random treatment-control differences that exceed the low program fees. To draw conclusions for these three programs, the evaluation considered effects on Part A and B costs, hospitalizations, emergency room use, and the size of the fee that needed to be covered. The evidence is strongest for QMed—although the differences

were not statistically significant, the treatment group had substantially (7 percent) fewer hospitalizations and 11 percent lower Medicare Part A and B expenditures than the control group, and the sample size is large. In contrast, the analogous differences were quite small and not significant for Medical Care Development and Quality Oncology, making it unlikely that they were cost neutral. To ensure that the findings were not sensitive to outliers, the extreme values of Medicare expenditures in each program (those above the 98th percentile for that program) were recoded to the value of the 98th percentile. These estimates yield the same conclusions drawn above.

G. SYNTHESIZING THE FINDINGS: WHAT WORKS AND WHAT DOESN'T

The results presented here from the first large-scale, rigorous study of care coordination programs illustrate the difficulties of reducing the need for expensive medical care among Medicare beneficiaries with chronic illnesses. While patients liked the programs, and 6 of the 15 programs had positive impacts on some quality-of-care indicators, nearly all were unsuccessful in changing patients' behavior. Taken collectively, the programs did lower hospitalizations by 4.5 percent, but this modest reduction did not translate into Medicare Part A and B savings. As a result, when the fees the programs charged for care coordination are included, the 15 programs combined *increased* total Medicare expenditures by 11 percent.

None of the individual programs *reduced* total Medicare program expenditures over the period examined. However, the authorizing legislation for the demonstration allows the Secretary of Health and Human Services to implement on a permanent basis any program components that are "beneficial" to the Medicare program, defined as ones that either reduce expenditures or improve quality and are cost neutral. Ten of the 15 programs increased total expenditures. One of these, Mercy Medical Center, significantly reduced hospitalizations, but had a sizeable fee that was over twice as large as the estimated savings in Part A and B expenditures. If these same reductions could be achieved with an intervention that could be afforded at a substantially lower fee than paid during this demonstration, this program could be cost neutral. Only 2 of the programs were clearly cost neutral (Georgetown University and Health Quality Partners). Due to the large variation in Medicare costs, we cannot definitively state whether the other 3 programs that may have been cost neutral (QMed, Quality Oncology, and Medical Care Development) actually were cost neutral. However, based on the pattern of findings, only QMed seems likely to have achieved cost neutrality. In addition, Georgetown, Health Quality Partners, and QMed each improved quality of care as measured by some of the indicators examined and did not increase total expenditures. However, there are too many program features and combinations of program features to determine definitively, without more detailed qualitative assessment, which ones are unique to these three programs and responsible for their success, and whether they would be replicable. CMS has extended three of the programs.

Currently, CMS is negotiating extensions with some of the programs.

1. Inferences About the Promise of Coordinated Care

The programs tested in this demonstration do not offer solutions to the mushrooming costs of treating beneficiaries with chronic illness. The Medicare trustees currently project that the program will be bankrupt in 2019, which makes this a critical time to find ways to improve care for people with chronic illness. Yet the findings on the lack of effectiveness of care coordination for most of the programs argue strongly against the wholesale adoption of care coordination programs in the Medicare fee-for-service setting without further testing. These results indicate that care coordination as provided by these programs would not help control expenditures and, for most of the programs, would paradoxically accelerate the time until the Medicare program loses solvency. Moreover, the findings contradict the oft-heard claims of disease management companies that they not only are cost neutral but substantially reduce total costs; in this demonstration, 10 of the 15 programs increased total Medicare expenditures when the costs of the intervention are included, and none of the other 5 actually reduced expenditures.

The programs improved quality of care on some dimensions , but did not reduce mortality or improve patient functioning. Only a few programs improved patients' quality of care, and even for these programs the effects were limited. Overall, the programs as a whole showed small but significant improvements in some testing indicators for people with diabetes and coronary artery disease, and reduced the proportion of CHF sample members who had preventable hospitalizations for CHF by about 17 percent, regardless of whether this is examined over a one-, two-, or three-year period after enrollment. Three programs (Georgetown, QMed, and Hospice of the Valley) had consistently sizable and statistically significant effects on these hospitalizations, and two others (CorSolutions and Avera) also had sizable differences favoring the treatment group. However, preventable hospitalizations for patients with diabetes *increased* overall, suggesting that either the programs identified some diabetic patients who needed inpatient care but would not otherwise have received it in that year, or that some unnecessary hospitalizations were induced by the demonstration.

The programs were no more cost-effective for particular diagnostic groups than others. The findings do not support the conventional wisdom in the field that programs that target people with CHF are more likely to generate savings. In fact, two of the three programs that exclusively targeted beneficiaries with CHF (CorSolutions and Avera) had no effect on the total number of hospitalizations or Medicare Part A and B expenditures, and therefore increased total expenditures sizably.[8] Across all programs combined, neither the number of hospitalizations nor expenditures for treatment group patients with CHF were different from those for the corresponding control group members (estimates available on request). Turning to the program-specific estimates, the magnitude of impacts on hospitalizations for the subgroup of beneficiaries with CHF was similar to the magnitude for the overall sample. For Medicare Part A and B expenditures, there were no statistically significant effects for beneficiaries with CHF. Apparently, the significant reductions in the proportion who had preventable hospitalizations for CHF were not large enough to generate savings for this group of patients.

[8] The absence of effects on number of hospitalizations among CHF patients in these two programs contradicts the findings reported above that, for both programs, their CHF patients had a markedly lower probability of having a preventable hospital admission for CHF than did their corresponding control group members. Apparently, either CHF patients in the treatment group members who were admitted had slightly more admissions than controls for CHF reasons, or they were more likely to be admitted for other diagnoses (for example, diabetes, which is a common co-morbidity for people with CHF).

Impacts did not grow as the programs learned more about delivering the intervention, or as beneficiaries had more time in the programs. For nearly all programs, estimated effects for enrollees who enrolled in the second or third years of the demonstration were similar to those for enrollees enrolling in the first year of operations (measured over the first 12 or 24 months after the patient's enrollment). When comparing across patient follow-up periods (first, second, and third years after enrollment), only one program (Hospice of the Valley) showed a pattern that suggests program effects may have increased with patients' length of time enrolled. Such a pattern might have been expected, especially for patients with some conditions, such as diabetes, who may take longer to show improved outcomes than those with other conditions.

2. What Worked Best and Why

The logic model described earlier suggests a number of potential pathways to a successful intervention. The evaluation attempted to find evidence of relationships between program features and favorable impacts on cost and quality of care outcomes. Summary process measures of the quality and intensity of each program's intervention on various dimensions, structural features of the programs, and characteristics of the target population were each examined for evidence of associations with favorable cost and quality impacts.

No particular program types or target populations were consistently associated with favorable cost and quality outcomes. Table 7 examines the features of the six potentially effective programs: two that were cost neutral (Georgetown University and Health Quality Partners), three that may have been (QMed, Quality Oncology, and Medical Care Development), and one that significantly reduced hospitalizations but increased total expenditures (Mercy). This analysis indicates that the six potentially effective programs are extremely diverse along each dimension, including organization type, size, program fee, diagnostic mix, and severity of illness of their enrollees. The last column of the table displays the average of the program features across all programs, to illustrate how the potentially effective programs are distinguished from the others.

Programs with the most in-person contacts were generally more successful. One program intervention feature does stand out. Five of the 15 programs in the demonstration had more than 0.8 in-person contacts per patient per month on average, and three of these programs (Georgetown, Health Quality Partners, and Mercy) were among the six potentially effective programs (Table 7). This finding suggests that in-person contacts, whether in the patient's home or at a physician's office or clinic, help establish the trust and rapport needed for the patient to be responsive to the care coordinator's advice. Program staff have also noted that seeing the patient and his or her living environment (including aspects that the patient might not tell staff about on the telephone) greatly enhances their ability to understand the patient's situation and tailor the intervention to it. However, in-person visits are more expensive. Furthermore, two of the six promising programs (QMed and Quality Oncology) did *not* see patients in person, suggesting that while relatively frequent in- person visits may be beneficial, they do not appear to be essential.

TABLE 7

CHARACTERISTICS OF PROGRAMS THAT HAD SOME EVIDENCE OF EFFECTIVENESS

	Cost Neutral		Possibly Cost Neutral			Reduced Hospitalizations but Increased Total Costs	
	Georgetown University	Health Quality Partners	QMed	Quality Oncology	Medical Care Development	Mercy	Average Among All 15 Programs
Size and Organizational Characteristics							
Number of Treatment Group Members Enrolled Through June 2005	114	739	698	104	652	467	595
Organization Type	Academic	Quality improvement provider	Disease management provider	Disease management provider	Hospital	Hospital	
Characteristics of Enrollees Through June 2005							
Proportion with CHF (%)	99	11	40	19	63	62	53
Proportion with CAD (%)	83	35	50	47	90	65	63
Average Annualized Number of Hospitalizations in Two Years Prior to Enrollment of Control Group	2.5	0.4	0.4	1.0	1.5	1.1	1.1
Average Monthly Medicare Expenditures in Two Years Prior to Enrollment of Control Group	$2,396	$451	$517	$2,337	$1,492	$1,064	$1,284
Intensity and Location of Contacts For First Year Enrollees During the Year After Random Assignment							
Mean Number of Monthly Contacts	5.9	2.2	1.2	NA	1.5	1.4	1.9
Average Caseload per Staff	36	90	200	40	70	50	104
Mean Number of Contacts per Month That Were in Person	0.83	0.92	0.09	0.0	0.44	0.97	0.48
Program Fees Received Through June 2006 for Enrollees Through June 2005							
Approved Monthly Fee for Active Patients	$320	$108	$96	$140	$297	$257	$235
Average Program Fee Received per Month in Evaluation Sample	$242	$102	$81	$64	$126	$248	$155
Estimated Effects on Medicare Expenditures							
Without Care Coordination Fees	-$335	-$100	-$83	-$126	-$90	-$113	-$1
With Care Coordination Fees	-$93	$2	-$2	-$62	$36	$135	$154

Source: Estimates are drawn from earlier analysis presented in this report.

TABLE 8

CLASSIFICATION INDICES OF PROGRAMS THAT SHOWED SOME EVIDENCE OF EFFECTIVENESS, IN QUINTILES

	Cost Neutral		Reduced Hospitalizations but Increased Total Costs	Possibly Cost Neutral		
	Georgetown University	Health Quality Partners	Mercy	QMed	Quality Oncology	Medical Care Development
Program Staffing	2	2	1	5	1	4
Initial Assessment	1	1	3	5	3	3
Improving Communication and Coordination	2	2	1	4	5	3
Ongoing Monitoring	4	2	4	2	1	5
Problem Identification and Care Planning	1	3	1	4	5	3
Patient Education	3	1	1	4	4	4
Improving Provider Practice	2	4	5	1	3	2
Service and Resource Arranging	3	4	2	4	5	2
Information Technology and Electronic Records	3	5	4	4	1	3
Quality Management and Outcome Measurement	2	3	4	2	1	4

Source: Means of two independent rating scores of each program by evaluators. Raters consulted program documents, telephone and site visit interview notes, evaluation case studies, and evaluation first-year reports to complete structured assessment forms. The forms asked a series of questions on the 10 domains listed in the row headings. A ranking in quintile 1 means that the program had one of the three highest values among the 15 programs; a ranking in quintile 5, the three lowest. See Brown et al. (2007) for details of the classification algorithm and the component items from which rankings in each domain were derived.

The results indicate that focusing on particular features is neither necessary nor sufficient for success. QMed, for example, one of the three possibly budget neutral programs still operating, scored in the fourth or fifth quintile on 7 of the 10 measures but was unique among the potentially effective programs in focusing on improving provider practice. Carle Clinic, on the other hand (not shown here), scored in the first or second quintile on 9 of the 10 dimensions, but did not reduce hospitalizations or costs.

Despite the lack of clear evidence that having a strong intervention in certain domains is essential for success, some features were weakly associated with better outcomes. First, the programs with some evidence of effectiveness were more likely to be relatively stronger on program staffing (for example, requiring care coordinators to be more educated or to have more training); four of the six were classified in quintile 1 or 2. Second, the potentially effective programs also were rated as having stronger initial assessments (for example, conducting more structured, comprehensive assessments); the two programs that were clearly cost neutral were in quintile 1. Finally, the potentially effective programs tended to be rated highly on improving communication and coordination (for example, care coordinators knowing about all of a patient's physicians or medications); three of the six promising programs, including both those that were cost neutral, scored in quintile 1 or 2. The program features that did not appear to matter include

information technology and electronic records and service arranging. Note that while most of the six potentially effective programs did not have high relative ratings on patient education, one of the cost effective programs and the only currently operating program that significantly reduced hospitalizations were both ranked in the highest quintile.

3. Reasons Why There Were Few Program Successes

The finding that two-thirds of the programs show no evidence of favorable impacts on Medicare Part A and B expenditures or hospitalizations, and that none show statistically significant savings in total expenditures, likely reflects a number of factors described below.

Most programs lacked extensive care coordination experience, and many lacked experience working with fee-for-service Medicare beneficiaries. Although CMS selected programs that had demonstrated some prior success with their proposed demonstration interventions, for most, the weakness of the evidence for such success suggests that they were not programs with *proven* track records as CMS had hoped. Moreover, most program hosts had not previously provided their proposed interventions to Medicare beneficiaries in the fee-for-service sector. For example, although some had developed sophisticated procedures for their commercial clients that they used to support care coordination, the four commercial disease management providers (CorSolutions, QMed, Quality Oncology, and Washington University's partner, Status One) had worked prior to the demonstration primarily with working-age, managed care plan members. The other demonstration programs generally had less sophisticated procedures in place when the demonstration started and were taking advantage of the opportunity afforded by demonstration participation to refine their procedures. As a result, most programs reported having to adapt their procedures as they learned about the needs of the patients they had enrolled and the particular difficulties of working with older patients, many of whom had multiple chronic health problems, complex medication regimens, and difficulty communicating or learning new information.

It is difficult to improve the self-care behavior of elderly beneficiaries. While nearly all the programs devoted considerable attention to patient education, it is exceedingly difficult to change people's behavior. At this point, there is no uniformly agreed-upon method of engaging beneficiaries and getting them to adhere rigorously to prescribed diet, exercise, and medication regimens. Studies of the effectiveness of diets for adults of any age find that average weight loss is slight at best (Dansinger et al. 2007; Heshka et al. 2003) and that losses are difficult to maintain (Mann et al. 2007). Adherence to medication regimens is similarly difficult to improve. The experience some of the programs had was with younger populations, but it may be even harder to change the behaviors of the elderly, because (1) they have been practicing these behaviors longer and may perceive less value to changing their lifestyle, since they have less time to live; and (2) many have difficulty remembering to adhere or are physically unable to.

There may have been little opportunity to improve the adherence of program enrollees. Although it is difficult to change the behavior of non-adherents to self-care regimens, even the strongest efforts might have had a negligible impact if beneficiaries who volunteered for the demonstration tend to be ones who were already the most able and willing to adhere to their prescribed regimens. For example, about 90 percent of beneficiaries in both the treatment and the control groups reported not missing a dose of their medication in the week prior to the survey, a rate that leaves little room for improvement.

Some of the participating programs are part of medical care systems that already have well-coordinated care. In some cases, programs may have had little opportunity to improve on the existing system. For example, Carle Clinic was rated highly on many aspects of care. Virtually all of the enrollees in the Carle program were patients of Carle Clinic, however, so the control group is likely to have benefited from Carle's electronic health records, the sharing of information among providers, and strong patient education, even without the benefit of the care coordinators. While the incremental effect of the program did lead to improved quality of care outcomes, Carle was unable to reduce hospitalizations below that experienced by the clinic's patients in the control group, perhaps because the clinic's usual care system already minimizes preventable hospitalizations.

Improvements in the quality of care do not necessarily result in reductions in hospitalizations or costs, even over a three-year period. The results for Carle Clinic in particular illustrate the fact that improvements in preventive care do not necessarily produce cost savings. While this program showed significant gains on more quality of care indicators than any other program, it had no impacts on hospitalizations or expenditures. Even the reduction in preventable CHF hospitalizations produced by several of the programs did not lead to significant effects on hospitalizations overall or Part A expenditures for most of the programs. While improved care may improve the well-being or quality of life of beneficiaries, and may result in eventual reductions in hospitalizations, it also may identify the need for additional procedures or hospitalizations beyond what would have occurred in the absence of the program, leading to higher costs. The treatment group's significantly higher rate of "preventable" hospitalizations for diabetes is consistent with such program effects.

Most programs had limited ability to coordinate medical care across providers or change provider behavior. Though the demonstration was intended as a test of care coordination broadly defined, in an attempt not to burden or antagonize physicians, most programs required little information from patients' primary and specialty care providers. Instead, programs relied on patients to describe problems they had with providers (such as receiving conflicting advice) or their need for care coordination. Feedback from the physicians of patients in the MCCD interviewed for this study illustrates this. While the physicians generally gave good ratings to the programs, few reported that the programs made various aspects of care coordination "a lot better." Fifteen percent or fewer gave high ratings to the programs' coordinating care across physicians, helping patients deal with contradictory information from other health care providers, and reducing duplicative tests.

Having someone play a more active role in coordinating medical care would seem to be a promising feature of care coordination because Medicare beneficiaries with one or more of eight chronic illnesses received Medicare-covered services from an average of 17 different physicians per year during 2002-2005, and there was often no one physician responsible for a beneficiary's care (Pham et al. 2007). It is plausible that a greater focus on coordinating medical care across providers might improve chronic care. For example, programs might enhance their ability to coordinate care by (1) obtaining medical treatment plans from all physicians treating the patient, comparing them to evidence-based guidelines, making recommendations to change treatments, and sharing a summary with all providers, and (2) obtaining and sharing input from all providers about whether the planned care was delivered, any diagnostic findings, patient adherence, the patient's response to treatment, and any observed changes in the patient's condition. This input could be shared through electronic medical records, regularly scheduled case conference calls, or established agreements with providers to furnish such information routinely to the care coordinator. However, in a fee-for-service setting such as this demonstration, electronic health

records are usually not available to all providers who treat a patient, and programs typically have no authority over providers to require that they share information. While incentives such as compensation for sharing data could be offered, doing so increases the cost of the program and the magnitude of savings that must be achieved for cost neutrality.

4. Possible Implications of the Findings for the Future of the Demonstration Programs

The Balanced Budget Act of 1997, which mandated the demonstration, specifies that the Secretary of Health and Human Services "may issue regulations to implement, on a permanent basis, the components of the demonstration project that are beneficial to the Medicare program." The Act defines programs as beneficial if they reduce total Medicare expenditures or improve the quality of care and increase beneficiary and provider satisfaction, without increasing total Medicare expenditures. The results here suggest that only three programs appear to have met these criteria—Health Quality Partners, Georgetown University, and QMed. Georgetown, which dropped out of the program before the end of its authorized four-year implementation time period, was never able to enroll many patients, and received a very sizeable additional (non-CMS) development grant to establish its program. These factors suggest that it is not a viable model to replicate.

Neither Health Quality Partners nor QMed generated savings for Medicare, but for both programs, total Medicare expenditures for the treatment group, including care coordination fees, were virtually identical to those for the control group, and both had significant effects on several of the quality indicators examined. In both programs, the treatment group members rated their ability to get adequate explanations of treatments significantly higher than the control group did. Health Quality Partners also increased patients' ratings of their providers keeping in touch with each other, and increased the proportion of patients with diabetes who received routine hemoglobin tests and the proportion of both patients with CAD and those with diabetes who received cholesterol tests. QMed substantially reduced the proportion of patients with CHF who incurred preventable hospitalizations for this disease, and significantly increased both the proportion of all patients receiving colon cancer screening and the proportion of CAD patients receiving blood lipid tests. Thus, these two programs appear to be beneficial to the Medicare program.

5. Implications for Other Medicare Disease Management and Care Coordination Interventions

The findings from this third report to Congress suggest that it will be difficult for other interventions like the 15 tested in the MCCD to show either improvements in quality or reductions in total expenditures for Medicare beneficiaries, especially in a fee-for-service setting. In particular, the findings do not bode well for efforts to use care coordination or disease management to improve quality and reduce overall expenditures on a larger scale within the Medicare program. If the MCCD demonstration's smaller, more manageable programs, some in very integrated environments, were unable to generate savings, the likelihood of doing so in large commercial programs that are population-based and serving thousands of patients seems remote.[9]

[9] Population-based programs often have difficulty engaging many of the individuals in the target group, because the individual did not volunteer for the program.

On the other hand, one encouraging sign is that two of the six most promising MCCD programs charged lower fees than most of the other programs, suggesting that generating reductions in hospitalizations does not require an expensive intervention. Also, it is interesting that the only two demonstration programs that are probably cost neutral and were of sustainable size (HQP and QMed) were both hosted by commercial providers, even though commercial entities comprised only one-third of the 15 program hosts in the demonstration.

As suggested in a report prepared to help design the demonstrations (Chen et al. 2000), there appears to be no "magic bullet" that will guarantee success, and no one best design for a successful, cost-effective program. Changing patient behavior and provider practice is difficult, for the reasons cited above, as is coordinating care, given the problems with the lack of efficient methods of sharing data on patients. However, our interviews with program staff have suggested that even if such changes in behavior and practice cannot be achieved, identifying problems more quickly through frequent communication with patients, especially with routine in-person contacts, could lead to timely medication changes or physician visits that could avert a hospital stay.

The results of the MCCD demonstration to date lead the evaluators to conclude that efforts to improve patients' self-care and the communications among their medical providers should not be abandoned, but they do require substantial refinement. The reductions in hospital admissions, though small, are somewhat encouraging, as are the modest improvements in some quality of care indicators (which may eventuate in savings over longer periods of time than studied) for some programs. However, the favorable results are too small and restricted to too few programs to warrant rapid expansion of such programs.

While the Balanced Budget Act of 1997 authorized the Secretary to permanently implement program components that would benefit Medicare, the analysis presented here suggests that it is extremely difficult to identify such components. Furthermore, even if it were possible to identify a few features (such as monthly in-person contacts in addition to telephone contacts) that seemed to be strongly associated with program success in reducing hospitalizations and expenditures, there is no guarantee that tacking these features onto an existing program would yield a successful intervention. On the surface, the programs that were not cost neutral do not seem markedly different from the ones that were. It appears that it is the details of how interventions are conducted that make the difference between success and failure.

These conclusions, if correct, suggest that it is not yet feasible to specify clearly a care coordination benefit that would be highly likely to produce improvements in quality and generate at least enough savings in Part A and B costs to cover its cost. Rather, the evaluation results suggest that a more fruitful approach would be to allow the two most promising active demonstration programs (HQP and QMed) to continue operating so that these interventions can be observed and documented in much more detail than was possible when 15 programs were being studied with equal intensity. A demonstration could then be conducted to determine whether strict adherence to either of these models can consistently lead to replication of their favorable results when delivered by other organizations in other environments.[10]

[10] Given Mercy Medical Care's statistically significant and sizable reductions in hospitalizations, it too would warrant consideration, but only if it and CMS were confident that the same impressive effects could be obtained at a fee less than half the size paid during the demonstration.

Although both programs are hosted by commercial service providers, they represent very different models for care coordination. HQP, with offices in Doylestown Hospital in Pennsylvania, serves beneficiaries in suburban and rural communities. The program focused primarily on improving patient adherence and self care, and improving patients' ability to communicate with their physicians. It relied on providers to identify patients for the program, and sent the patients letters on their physician's letterhead inviting them to participate. Enrolled patients, who had asthma, CAD, diabetes, CHF, hyperlipidemia, or hypertension, were risk stratified into three groups (moderate, high without geriatric frailty, and high with frailty) that received different intensity interventions. HQP tailored its education—rated by the evaluation team to be among the best of the demonstration programs—to individual patients' needs, risk category, and readiness to change, and taught them to ask questions of their providers and request needed care. Care coordinators were required to be registered nurses with at least five years of relevant experience, such as community nursing or hospice. The program trained care coordinators in how to deliver the intervention and to probe for evidence the patient understood fully, and included caregivers in the sessions for patients with cognitive difficulties. The program also provided patients with low-literacy visual aids in place of some written materials. Care coordinators often saw patients in person, both at home and in physicians' offices. The program tried to minimize its demands on physicians who have affiliated themselves with the program, most of whom were familiar with it from HQP's pilot program and previous work with Doylestown Hospital. HQP created rapport by assigning care coordinators to patients based on their physician, so that each physician tends to interact with the same coordinator for all of his or her program patients. Care coordinators sent reports of each patient contact to the patient's physician, by mail, email, or fax (depending on the physician's preferences), and met regularly with the physicians.

QMed differed from HQP in many ways. While HQP focused its intervention on improving patient adherence and communication skills, QMed focused on improving physician practice. QMed is a disease management services provider based in Eatontown, New Jersey. Its MCCD program targeted beneficiaries with coronary artery disease (CAD) who lived in several northern California counties covering urban, suburban, and rural areas. To recruit patients, QMed first recruited several large physician group practices with whom it had worked through the company's managed care contracts. Practice office staff then developed lists of eligible patients and the program sent invitation letters to them on practice letterhead signed by their own physician or the practice medical director. QMed's program intervention, based on its commercially available product, focused on improving physician practice by providing physicians with regular patient-specific reports comparing his or her medical treatment plan (based on the program's abstraction of the physician's medical records) with readings from the program's ambulatory ischemia monitor and evidence-based guidelines for the treatment of CAD. QMed's California-based quality assurance managers also met regularly with physicians to discuss their adherence to the guidelines and alerted practice medical directors if physicians in their practice were failing to follow more than a quarter of the program's recommendations. In addition to periodic ischemia monitoring, program patients were contacted by New Jersey-based care coordinators (who were registered nurses or experienced licensed practical nurses) who monitored their health and provided education about CAD. The program also sent patients wallet cards to prompt questions to physicians during visits and reminders of needed tests, preventive exams and vaccinations. QMed was not rated highly by the evaluation team on most of the 10 domains examined, but had one of the highest rankings of all the programs on improving provider practice, and also ranked highly on monitoring and quality measurement and management. Because the estimated savings

in Part A and B expenditures for QMed were not statistically significant, it may be necessary to lower the fee for this program to increase confidence that cost neutrality will be achieved.

While terminating all of the demonstration programs would ensure that no new cost increases will be created for Medicare from care coordination fees, failing to pursue effective programs such as Health Quality Partners or QMed (and possibly Mercy Medical Center, if it can operate with substantially reduced program fees) may mean a missed opportunity to substantially improve the quality of care for chronically ill beneficiaries at no increase in cost to Medicare. If the interventions can maintain their effectiveness at lower cost, it may even be possible to generate net savings for Medicare. Furthermore, the benefit of identifying successful interventions could be great for Medicaid as well, as nearly all states are investing in disease management programs, typically with little or no evidence that the programs will generate the savings that commercial programs promise. Thus, honing in on a detailed, concrete description of successful interventions for those with chronic illnesses and testing the replicability of these interventions seems to warrant serious consideration.

REFERENCES

Anderson, Gerard, and Jane Horvath. "Chronic Conditions: Making the Case for Ongoing Care." Baltimore, MD: The Johns Hopkins University, December 2002.

Bagchi, Ann D., Dominick Esposito, Myoung Kim, James Verdier, and Deo Bencio. "Adherence to Medications for the Treatment of Congestive Heart Failure and Its Effect on Medicaid Expenditures." Paper presented at the Academy Health Annual Research Meeting, Orlando, FL, June 2007.

Baicker, Katherine, and Amitabh Chandra. "Medicare Spending, the Physician Workforce, and Beneficiaries' Quality of Care." *Health Affairs*, Web Exclusive, 2004, W184-197.

Bodenheimer, T., K. Lorig, H. Holman, and K. Grumbach. "Patient Self-Management of Chronic Disease in Primary Care." *Journal of the American Medical Association*, vol. 288, no. 19, November 20, 2002, pp. 2469–2475.

Brown, Randall, Deborah Peikes, Arnold Chen, Judy Ng, Jennifer Schore, and Clara Soh. "The Evaluation of the Medicare Coordinated Care Demonstration: Findings for the First Two Years." Second report to Congress. Princeton, NJ: Mathematica Policy Research, Inc., March 21, 2007.

Brown, Randall, Jennifer Schore, Nancy Archibald, Arnold Chen, Deborah Peikes, Karen Sautter, Bridget Lavin, Sherry Aliotta, and Todd Ensor. "Coordinating Care for Medicare Beneficiaries: Early Experiences of 15 Demonstration Programs, Their Patients, and Providers." First report to Congress. Princeton, NJ: Mathematica Policy Research, Inc., December 2003.

Chen, Arnold, Randall Brown, Nancy Archibald, Sherry Aliotta, and Peter Fox. "Best Practices in Coordinated Care." Princeton, NJ: Mathematica Policy Research, Inc., February 29, 2000.

Dansinger, Michael L., Joi Augustin Gleason, John L. Griffith, Harry P. Selker, and Ernst J. Schaefer. "Comparison of the Atkins, Ornish, Weight Watchers, and Zone Diets for Weight Loss and Heart Disease Risk Reduction: A Randomized Trial." *Journal of the American Medical Association*, vol. 293, no. 1, January 5, 2005, pp. 43-53.

Galbreath, Autumn Dawn, Richard A. Krasuski, Brad Smith, Karl C. Stajduhar, Michael D. Kwan, Robert Ellis, and Gregory L. Freeman. "Long-Term Healthcare and Cost Outcomes of Disease Management in a Large, Randomized, Community-Based Population With Heart Failure." *Circulation*, November 7, 2004, pp. 3518-3526.

Goetzel, Ron Z., Ronald J. Ozminkowski, Victor G. Villagra, and Jennifer Duffy. "Return on Investment in Disease Management: A Review." *Health Care Financing Review,* vol. 26, no. 4, summer 2005, pp. 1-19.

Gravelle, Hugh, Mark Dusheiko, Rod Sheaff, Penny Sargent, Ruth Boaden, Susan Pickard, Stuart Parker, and Martin Roland. "Impact of Case Management (Evercare) on Frail Elderly

Patients: Controlled Before and After Analysis of Quantitative Outcome Data." *British Medical Journal,* vol. 334, no. 7583, January 6, 2007, pp. 31-34.

Health Care Financing Review. *Medicare and Medicaid Statistical Supplement, 2005*. U.S. Department of Health and Human Services, Centers for Medicare & Medicaid Services, Pub. no. 03474, July 2006.

Heshka, Stanley, James W. Anderson, Richard L. Atkinson, Frank L. Greenway, James O. Hill, Stephen D. Phinney, Ronette L. Kolotkin, Karen Miller-Kovach, and F. Xavier Pi-Sunyer. "Weight Loss with Self-help Compared with a Structured Commercial Program: A Randomized Trial." *Journal of the American Medical Association,* vol. 289, no. 14, April 9, 2003, pp. 1792-1798.

Jencks, S.F., E.D. Huff, and T. Cuerdon. "Change in the Quality of Care Delivered to Medicare Beneficiaries, 1998-1999 to 2000-2001." *JAMA*, vol. 289, no. 3, January 15, 2003, pp. 305-312.

Kilpatrick, Kerry E., Kathleen N. Lohr, Sheila Leatherman, George Pink, Jean M. Buckel, Caroline Legarde, and Lynn Whitener. "The Insufficiency of Evidence to Establish the Business Case for Quality." *International Journal for Quality in Health Care,* vol. 17, no. 4, August 1, 2005, pp. 347-355.

Leatherman, Sheila, and Douglas McCarthy. "Quality of Health Care for Medicare Beneficiaries: A Chartbook." The Commonwealth Fund, vol. 815, May 2005. Available at [www.cmwf.org/publications/publications_show.htm?doc_id=275195], accessed May 1, 2007.

Lorig, K.R., P. Ritter, A.L. Stewart, D.S. Sobel, B.W. Brown, Jr., A. Bandura, V.M. Gonzalez, D.D. Laurent, and H.R. Holman. "Chronic Disease Self-Management Program: 2-Year Health Status and Health Care Utilization Outcomes." *Medical Care*, vol. 39, no. 11, November 2001, pp. 1217–1223.

Lorig, K.R., D.S. Sobel, A.L. Stewart, B.W. Brown, Jr., A. Bandura, P. Ritter, V.M. Gonzalez, D.D. Laurent, and H.R. Holman. "Evidence Suggesting That a Chronic Disease Self-Management Program Can Improve Health Status While Reducing Hospitalization: A Randomized Trial." *Medical Care*, vol. 37, no. 1, January 1999, pp. 5–14.

Mann, Traci, A. Janet Tomiyama, Erika Westling, Ann-Marie Lew, Barbra Samuels, and Jason Chatman. "Medicare's Search for Effective Obesity Treatments: Diets Are Not the Answer." *American Psychologist,* vol. 62, no. 3, April 2007, pp. 220-233.

Matheson, D., and D. Psacharopoulos. *Realizing the Promise of Disease Management: Payer Trends and Opportunities in the United States.* Boston, MA: Boston Consulting Group, 2006.

Ofman, J.J., E. Badamgarav, J.M. Henning, K. Knight, A.D. Gano Jr, R.K. Levan, S. Gur-Arie, M.S. Richards, V. Hasselblad, and S.R. Weingarten. "Does Disease Management Improve Clinical and Economic Outcomes in Patients with Chronic Diseases? A Systematic Review." *American Journal of Medicine*, vol. 117, no. 3, August 1, 2004, pp. 182-192.

Pham, Hoangmai H., Deborah Schrag, Ann S. O'Malley, Beny Wu, and Peter B. Bach. "Care Patterns in Medicare and Their Implications for Pay for Performance." *New England Journal of Medicine*, vol. 256, no. 11, March 15, 2007, pp. 1130-1139.

Smith, Brad, Emma Forkner, Barbara Zaslow, Richard A. Krasuski, Karl Stajduhar, Michael Kwan, Robert Ellis, Autumn Dawn Galbreath, and Gregory L. Freeman. "Disease Management Produces Limited Quality-of-Life Improvements in Patients with Congestive Heart Failure: Evidence from a Randomized Trial in Community-Dwelling Patients." *American Journal of Managed Care*, vol. 11, no. 11, November 2005, pp. 701-713.

Sokol, M.C., K.A. Mcguigan, R.R. Verbrugge, and R.S. Epstein. "Impact of Medication Adherence on Hospitalization Risk and Healthcare Cost." *Medical Care,* vol. 43, no. 6, June 2005, pp. 521–530.

Villagra, Victor, and Tamim Ahmed. "Effectiveness of a Disease Management Program for Patients with Diabetes." *Health Affairs,* vol. 23, no. 4, July-August 2004, pp. 255-266.

Wagner, E.H., B.T. Austin, C. Davis, M. Hindmarsh, J. Schaefer, and A. Bonomi. "Improving Chronic Illness Care: Translating Evidence into Action." *Health Affairs,* vol. 20, no. 6, 2001, pp. 64-78.

APPENDIX A

CLASSIFICATION INDEXES FOR INTERVENTION FEATURES

Classification domains. While information on what programs were doing and how they were doing it can be useful for understanding why some programs were more effective than others, it is also important to know the degree to which programs focused on particular dimensions of care coordination and how well designed the interventions were on these dimensions. On the surface, many of the programs in this demonstration appear to have implemented similar interventions, yet in-depth discussions with the programs revealed a number of important differences in the intensity of their intended efforts to provide patient education, better coordinate care, or implement other intervention components. To address this issue, the evaluation developed an algorithm for classifying each program's interventions on 10 separate domains:

Program staffing	Improving provider practice
Initial assessment	Service and resource arrangement
Problem identification and care planning	Information technology and electronic records
Patient education	Ongoing monitoring
Improving communication and coordination	Quality management and outcome measurement

These classification indexes were developed independently of data on program outcomes and data the programs supplied regarding the number and nature of contacts the coordinators had with patients. Researchers making the classifications relied solely on the information programs gave them during in-person and telephone discussions. (It seems unlikely that any changes programs reported by email in their final year would have affected a program's classification category for any domain.) Estimates of program impacts were not shared with these researchers until after they had completed their ratings. Classification indexes were normalized to range from 0 ("intervention did not address this domain") to 100 ("intervention was extremely well designed on this domain"). Within each domain, programs were arranged into quintiles based on their index values. Rather than judging program effectiveness a priori, the indexes and quintiles were meant to provide a parsimonious summary of program implementation that could aid in understanding why some programs appeared to have been more effective than others.

Overview of classification results. Programs varied widely on each of the 10 domains, especially Quality Management and Outcome Measurement (index values ranged from 5 to 91), and Improving Provider Practice (values ranged from 0 to 77). Values varied less across programs on the Initial Assessment and the Problem Identification domains. Average values were highest for the Initial Assessment and the Monitoring domains and lowest for Improving Provider Practice, which reflects the lesser attention most programs paid to this area.

While individual programs often had high index values for some domains and low ones for others (at times because a particular domain was not part of an intervention), across several domains, a few had high values and others had consistently low ones (see Table A.1 Carle was in the top quintile of programs for 6 of the 10 domains, and Mercy and Quality Oncology each had 4 values in the top quintile. (Each quintile contained 3 of the 15 programs.) The Jewish Home and Hospital and the University of Maryland were in the bottom quintile on 9 and 7 of the domains, respectively. Yet both these programs were in the top quintile on 1 domain each.

TABLE A.1

PROGRAM CLASSIFICATION INDEX QUINTILES, BY DOMAIN

	Program Staffing	Initial Assessment[a]	Problem Ident. & Care Plan.[a]	Patient Education	Impr. Comm. and Coord.	Improving Provider Practice	Service & Resource Arrange.	Info. Tech. & Elec. Records	Ongoing Monitoring	Qual. Mgt. & Outcome Meas.
Avera	**5**	**2**	**1**	**2**	**4**	**2**	**3**	**4**	**1**	**5**
Carle	**1**	**2**	**2**	**2**	**1**	**1**	**1**	**1**	**3**	**1**
CenVaNet	**3**	**1**	**4**	**3**	**3**	**3**	**4**	**2**	**1**	**2**
Charlestown	**3**	**4**	**2**	**3**	**1**	**4**	**5**	**2**	**3**	**4**
CorSolutions	**2**	**2**	**2**	**1**	**4**	**1**	**2**	**2**	**2**	**3**
Georgetown	**2**	**1**	**1**	**3**	**2**	**2**	**3**	**3**	**4**	**2**
Health Quality Partners	**2**	**1**	**3**	**1**	**2**	**4**	**4**	**5**	**2**	**3**
Hospice of the Valley	**4**	**4**	**3**	**2**	**2**	**5**	**3**	**5**	**4**	**3**
Jewish Home and Hospital	**5**	**5**	**5**	**5**	**5**	**5**	**1**	**5**	**5**	**5**
Medical Care Development	**4**	**3**	**3**	**4**	**3**	**2**	**2**	**3**	**5**	**4**
Mercy	**1**	**3**	**1**	**1**	**1**	**5**	**2**	**4**	**4**	**4**
QMed	**5**	**5**	**4**	**4**	**4**	**1**	**4**	**4**	**2**	**2**
Quality Oncology	**1**	**3**	**5**	**4**	**5**	**3**	**5**	**1**	**1**	**1**
University of Maryland	**4**	**5**	**5**	**5**	**5**	**4**	**5**	**1**	**5**	**5**
Washington University	**3**	**4**	**4**	**5**	**3**	**3**	**1**	**3**	**3**	**1**

Source: Index quintiles shown for each program are based on the means of values independently developed by two evaluator research staff. Staff consulted program documents, telephone and site visit interview notes, and evaluation case studies and evaluation first-year reports to complete structured assessment forms. The forms asked a series of questions on the 10 domains listed in the column headings.

Notes: Because there were 15 programs, each quintile consists of 3 programs. On each measure, quintile 1 contains the three highest values, quintile 5 the three lowest values. Quintiles 1 and 2 are shaded.

[a]Given the fair to poor correspondence between the two evaluator staff on scoring for these two domains, less importance or weight should be given to these.

Problem Ident. & Care Plan. = Problem Identification and Care Planning; Impr. Comm. & Coord. = Improving Communication and Coordination; Service & Resource Arrange. = Service and Resource Arrangement; Info. Tech. & Elec. Records = Information Technology and Electronic Records; Qual. Mgt. & Outcome Meas. = Quality Management and Outcome Measurement.

The importance of these rankings is not to identify those programs that do particularly well or poorly across measures, but to determine whether having a strong design for certain domains is consistently associated with favorable impacts on Medicare costs or the quality of care. It may well be that programs can have favorable effects even if they ignore 9 of 10 domains but have a strong intervention in the tenth.

APPENDIX B

SUPPLEMENTARY TABLES

TABLE B.1

QUALITY OF CARE INDICATORS

Health Education (5)
Reported being taught how to:
- Follow healthy diet
- Exercise
- Take medications
- Recognize warning signs to seek urgent care

Reported receiving materials about condition or treatment

Knowledge and Behavior (8)
Understands diet
Follows healthy diet most or all of the time
Understands proper way to exercise
Exercises regularly
Missed doses of medication $\geq$2 times per week
Visits physicians with a list of questions
If smoker, tried to quit
If drinker, tried to cut down

Service Arrangement and Unmet Needs (7)
If unable to arrange own care, received help with:
- Telephone
- Transportation
- Shopping
- Preparing meals
- Housework
- Taking medications
- Handling money

Functional Status (9)
Can do independently:
- Eat
- Dress
- Bathe
- Use telephone
- Prepare meal
- Perform housework
- Shop
- Take medications
- Handle finances

Health Status and Well-being (8)
Most or all of the time:
- Felt calm and peaceful

TABLE B.1 *(continued)*

- Felt downhearted or blue
- Slept poorly
- Pain interfered with usual activities
- Primary condition interfered with enjoyment of life
- Primary condition was a burden on family

SF-12 Physical health summary score
SF-12 Mental health summary score

Preventive Care (11)

- **General Preventive Care (5)**
- Flu shot (survey)
- Pneumonia vaccine (survey)
- Colon cancer screening (one each, survey and claims)
- Mammography (claims)

- **Disease-Specific Preventive Care (6)**
- For patients with diabetes:
  - Diabetes education
  - Eye examination
  - Cholesterol or lipid test
  - Hemoglobin A1c test
  - Urine test for protein
- For patients with CAD:
  - Cholesterol or lipid test

General and Disease-Specific Potentially Preventable Hospitalizations (8)

General or non-disease specific potentially preventable hospitalizations
Among patients with diabetes

- For CAD
- For diabetes
- For CHF
- Microvascular complication

Among patients with CHF

- For fluid or electrolyte problems
- For CHF

Among patients with CAD

- For CAD

TABLE B.2

AVERAGE ANNUALIZED NUMBER OF HOSPITAL ADMISSIONS PER YEAR, BY COHORT
(Regression Adjusted)

Follow-up Period:	One Year				Two Years				Three Years				Cumulative Through June 2006			
Cohort Enrolled Through:	June 2005				June 2004				June 2003				June 2005			
	Control Group	Treatment-Control Difference	Percent Change	*p*-Value	Control Group	Treatment-Control Difference	Percent Change	*p*-Value	Control Group	Treatment-Control Difference	Percent Change	*p*-Value	Control Group	Treatment-Control Difference	Percent Change	*p*-Value
1,100 or More Treatment Group Members Enrolled Through June 2005																
Carle	0.54	-0.03	-4.7	0.55	0.54	-0.01	-1.6	0.81	0.54	-0.02	-4.2	0.50	0.54	0.00	-0.2	0.97
CorSolutions	1.85	-0.04	-2.0	0.67	1.80	-0.07	-4.0	0.39	1.91	-0.06	-3.3	0.66	1.77	-0.05	-3.0	0.46
Washington University	1.36	0.06	4.3	0.50	1.29	0.05	3.9	0.51	1.38	0.00	0.3	0.96	1.36	-0.03	-2.2	0.66
415 to 725 Treatment Group Member Enrolled Through June 2005																
Avera	1.37	0.00	-0.1	0.99	1.38	0.04	3.3	0.75	1.39	0.02	1.2	0.92	1.29	0.03	2.3	0.78
CenVaNet	0.65	0.05	7.1	0.49	0.67	0.02	3.5	0.68	0.68	0.06	8.5	0.32	0.66	0.05	7.2	0.32
Charlestown	0.64	0.10	15.0	0.25	0.63	0.13	20.3	0.07*	0.67	0.12	17.7	0.13	0.65	0.08	12.3	0.19
Health Quality Partners	0.39	-0.04	-10.1	0.40	0.40	-0.03	-6.9	0.55	0.45	-0.07	-16.5	0.23	0.45	-0.06	-13.6	0.12
Hospice of the Valley	1.36	0.03	2.1	0.82	1.43	-0.20	-14.1	0.11	1.39	-0.28	-20.0	0.06*	1.35	-0.14	-10.7	0.13
Jewish Home and Hospital	0.84	0.06	7.5	0.54	0.87	0.03	3.9	0.74	0.85	0.07	8.2	0.54	0.87	0.04	4.4	0.65
Medical Care Development	1.26	0.04	3.5	0.69	1.16	-0.03	-2.9	0.75	1.36	-0.09	-7.0	0.54	1.15	-0.04	-3.3	0.66
Mercy	1.01	-0.25	-24.5	0.01***	0.98	-0.20	-20.2	0.02**	1.01	-0.17	-16.4	0.06*	0.98	-0.17	-17.0	0.02**
QMed	0.39	-0.04	-9.6	0.45	0.38	0.01	2.5	0.82	0.41	-0.02	-5.9	0.53	0.43	-0.03	-7.4	0.38
Under 115 Treatment Group Members Enrolled Through June 2005																
Georgetown	2.04	-0.39	-19.3	0.20	2.03	-0.65	-31.8	0.02**	1.76	-0.12	-6.6	0.75	2.00	-0.48	-24.1	0.06*
Quality Oncology	1.49	-0.40	-26.9	0.20	1.36	-0.06	-4.4	0.89	1.06	0.25	23.8	0.76	1.11	-0.02	-1.8	0.94
University of Maryland	2.34	-0.47	-19.9	0.28	1.95	-0.26	-13.1	0.55	2.62	-0.59	-22.5	0.55	2.11	-0.34	-16.1	0.35
All Programs	**1.03**	**-0.03**	**-3.2**	**0.18**	**0.97**	**-0.04**	**-4.1**	**0.07***	**0.89**	**-0.04**	**-4.3**	**0.10**	**0.96**	**-0.04**	**-4.5**	**0.02****

TABLE B.2 *(continued)*

Sources: Medicare Enrollment Database and Standard Analytic File.

Notes: Treatment and control group members who do not meet the demonstration-wide requirements of the Centers for Medicare & Medicaid Services (CMS) or were dead during the entire follow-up period, or who had an invalid Health Insurance Claim number on MPR's enrollment file, are excluded from this table because Medicare data showing their payments in the fee-for-service program were not available. Members of the same households as the research sample members are also excluded.

The outcomes are weighted according to the proportion of the follow-up period each sample member meets CMS's demonstration-wide requirements and is alive. CMS's requirements are as follows: being in fee-for-service, having both Part A and Part B coverage, and having Medicare as the primary payer. Weights are calculated separately for the treatment and control groups.

A statistically significant and negative treatment-control difference and percent change value indicate that the number of hospitalizations was lower for the treatment than control group. This signifies that the program is working as intended.

*Difference between the treatment and control groups is significantly different from 0 at the 0.10 level, 2-tailed test.
**Difference between the treatment and control groups is significantly different from 0 at the 0.05 level, 2-tailed test.
***Difference between the treatment and control groups is significantly different from 0 at the 0.01 level, 2-tailed test.

TABLE B.3

MONTHLY MEDICARE EXPENDITURES, BY COHORT
(Regression Adjusted)

Follow-up Period:	One Year				Two Years				Three Years				Cumulative Through June 2006			
Cohort Enrolled Through:	June 2005				June 2004				June 2003				June 2005			
	Control Group	Treatment-Control Difference	Percent Change	*p*-Value	Control Group	Treatment-Control Difference	Percent Change	*p*-Value	Control Group	Treatment-Control Difference	Percent Change	*p*-Value	Control Group	Treatment-Control Difference	Percent Change	*p*-Value
1,100 or More Treatment Group Members Enrolled Through June 2005																
Carle	655	11	1.7	0.81	663	48	7.3	0.25	686	17	2.5	0.66	711	42	5.8	0.24
CorSolutions	2,582	-48	-1.9	0.74	2,573	-5	-0.2	0.97	2,803	-84	-3.0	0.73	2,609	25	1.0	0.83
Washington University	1,915	94	4.9	0.43	1,814	125	6.9	0.24	1,908	101	5.3	0.38	1,897	71	3.7	0.44
415 to 725 Treatment Group Members Enrolled Through June 2005																
Avera	1,470	-197	-13.4	0.15	1,370	5	0.4	0.97	1,355	-79	-5.9	0.58	1,348	-12	-0.9	0.91
CenVaNet	723	78	10.7	0.23	805	26	3.2	0.67	839	52	6.2	0.43	861	43	4.9	0.44
Charlestown	942	179	19.0	0.13	952	216	22.7	0.03**	993	213	21.5	0.07*	1,023	156	15.2	0.08*
Health Quality Partners	659	-117	-17.7	0.10*	670	-68	-10.2	0.33	683	-58	-8.6	0.47	721	-100	-13.9	0.07*
Hospice of the Valley	2,035	112	5.5	0.45	2,084	23	1.1	0.89	2,044	-65	-3.2	0.76	2,069	-12	-0.6	0.93
Jewish Home and Hospital	1,817	61	3.3	0.79	1,739	23	1.3	0.91	1,682	111	6.6	0.58	1,751	78	4.4	0.63
Medical Care Development	1,578	-86	-5.4	0.57	1,381	-124	-9.0	0.36	1,484	-147	-9.9	0.41	1,379	-90	-6.5	0.40
Mercy	1,114	-151	-13.6	0.10*	1,136	-132	-11.7	0.11	1,195	-80	-6.7	0.37	1,197	-113	-9.4	0.13
QMed	682	-85	-12.5	0.36	700	-54	-7.7	0.51	762	-94	-12.4	0.20	788	-83	-10.5	0.21
Under 115 Treatment Group Members Enrolled Through June 2005																
Georgetown	2,508	-563	-22.4	0.20	2,550	-748	-29.3	0.06*	2,372	-252	-10.6	0.59	2,516	-335	-13.3	0.34
Quality Oncology	3,989	-496	-12.4	0.30	3,869	-251	-6.5	0.74	3,801	261	6.9	0.87	3,237	-126	-3.9	0.78
University of Maryland	2,647	1,200	45.3	0.38	2,576	-128	-5.0	0.85	3,697	-644	-17.4	0.70	2,610	803	30.8	0.44
All Programs	**1,428**	**-16**	**-1.1**	**0.66**	**1,351**	**-3**	**-0.2**	**0.94**	**1,220**	**1**	**0.1**	**0.96**	**1,369**	**-1**	**-0.1**	**0.96**

TABLE B.3 *(continued)*

Sources: Medicare Enrollment Database and Standard Analytic File.

Notes: Treatment and control group members who do not meet the demonstration-wide requirements of the Centers for Medicare & Medicaid Services (CMS) or were dead during the entire follow-up period, or who had an invalid Health Insurance Claim number on MPR's enrollment file, are excluded from this table because Medicare data showing their payments in the fee-for-service program were not available. Members of the same households as the research sample members are also excluded.

The outcomes are weighted according to the proportion of the follow-up period each sample member meets CMS's demonstration-wide requirements and is alive. CMS's requirements are as follows: being in fee-for-service, having both Part A and Part B coverage, and having Medicare as the primary payer. Weights are calculated separately for the treatment and control groups.

A statistically significant and negative treatment-control difference and percent change value indicate that expenditures were lower for the treatment than control group. This signifies that the program is working as intended.

*Difference between the treatment and control groups is significantly different from 0 at the 0.10 level, 2-tailed test.
**Difference between the treatment and control groups is significantly different from 0 at the 0.05 level, 2-tailed test.
***Difference between the treatment and control groups is significantly different from 0 at the 0.01 level, 2-tailed test.

TABLE B.4

MONTHLY TOTAL MEDICARE EXPENDITURES, INCLUDING PROGRAM FEES, BY COHORT
(Regression Adjusted)

Follow-up Period:	One Year				Two Years				Three Years				Cumulative Through June 2006				
Cohort Enrolled Through:	June 2005				June 2004				June 2003				June 2005				
	Control Group	Treatment-Control Difference	Percent Change	*p*-Value	Control Group	Treatment-Control Difference	Percent Change	*p*-Value	Control Group	Treatment-Control Difference	Percent Change	*p*-Value	Control Group	Treatment-Control Difference	Percent Change	*p*-Value	Cost Neutral
1,100 or More Treatment Group Members Enrolled Through June 2005																	
Carle	655	165	25.2	0.00***	663	199	30.0	0.00***	686	166	24.2	0.00***	711	190	26.7	0.00***	No
CorSolutions	2,583	225	8.7	0.12	2,573	204	7.9	0.14	2,802	117	4.2	0.63	2,609	217	8.3	0.07*	No
Washington University	1,915	261	13.6	0.03**	1,814	291	16.0	0.01***	1,908	265	13.9	0.02**	1,897	231	12.2	0.01**	No
415 to 725 Treatment Group Members Enrolled Through June 2005																	
Avera	1,470	85	5.8	0.54	1,370	280	20.5	0.03**	1,355	183	13.5	0.20	1,349	261	19.4	0.01**	No
CenVaNet	723	152	21.1	0.02**	805	98	12.2	0.11	840	122	14.5	0.07*	861	113	13.1	0.04**	No
Charlestown	942	416	44.2	0.00***	952	449	47.2	0.00***	992	443	44.6	0.00***	1,023	374	36.5	0.00***	No
Health Quality Partners	659	-12	-1.9	0.86	670	34	5.1	0.62	683	43	6.3	0.60	721	2	0.3	0.96	Yes
Hospice of the Valley	2,035	305	15.0	0.04**	2,085	207	9.9	0.21	2,045	109	5.3	0.61	2,069	167	8.1	0.18	No
Jewish Home and Hospital	1,818	308	17.0	0.18	1,740	270	15.5	0.17	1,682	343	20.4	0.09*	1,752	299	17.1	0.06*	No
Medical Care Development	1,578	102	6.5	0.50	1,381	14	1.0	0.92	1,484	-9	-0.6	0.96	1,379	36	2.6	0.74	Possibly
Mercy	1,114	101	9.1	0.27	1,136	117	10.3	0.15	1,195	168	14.1	0.06*	1,197	135	11.3	0.07*	No
QMed	684	5	0.8	0.96	701	37	5.2	0.65	764	-11	-1.5	0.88	790	-2	-0.2	0.98	Possibly
Under 115 Treatment Group Members Enrolled Through June 2005																	
Georgetown	2,509	-262	-10.4	0.55	2,550	-464	-18.2	0.24	2,371	18	0.8	0.97	2,516	-93	-3.7	0.79	Yes
Quality Oncology	3,989	-432	-10.8	0.36	3,869	-173	-4.5	0.82	3,795	373	9.8	0.81	3,237	-62	-1.9	0.89	Possibly
University of Maryland	2,646	1,511	57.1	0.27	2,569	170	6.6	0.80	3,683	-336	-9.1	0.84	2,605	1,080	41.5	0.30	No
All Programs	**1,430**	**166**	**11.6**	**0.00*** **	**1,352**	**163**	**12.1**	**0.00*** **	**1,221**	**155**	**12.7**	**0.00*** **	**1,369**	**154**	**11.3**	**0.00*** **	**No**

TABLE B.4 *(continued)*

Sources: Medicare Enrollment Database and Standard Analytic File.

Notes: Treatment and control group members who do not meet the demonstration-wide requirements of the Centers for Medicare & Medicaid Services (CMS) or were dead during the entire follow-up period, or who had an invalid Health Insurance Claim number on MPR's enrollment file, are excluded from this table because Medicare data showing their payments in the fee-for-service program were not available. Members of the same households as the research sample members are also excluded.

The outcomes are weighted according to the proportion of the follow-up period each sample member meets CMS's demonstration-wide requirements and is alive. CMS's requirements are as follows: being in fee-for-service, having both Part A and Part B coverage, and having Medicare as the primary payer. Weights are calculated separately for the treatment and control groups.

A statistically significant and negative treatment-control difference and percent change value indicate that the program generated savings. If the difference and percent change value are not statistically significant, the hypothesis that the program generated enough savings to offset the care coordination fees cannot be rejected. In other words, the program may have been cost neutral.

*Difference between the treatment and control groups is significantly different from 0 at the 0.10 level, 2-tailed test.
**Difference between the treatment and control groups is significantly different from 0 at the 0.05 level, 2-tailed test.
***Difference between the treatment and control groups is significantly different from 0 at the 0.01 level, 2-tailed test.