Contract No.: 500-95-0047 (09)
MPR Reference No.: 8756-420

MATHEMATICA
Policy Research, Inc.

The Evaluation of the Medicare Coordinated Care Demonstration: Findings for the First Two Years

March 21, 2007

Randall Brown
Deborah Peikes
Arnold Chen
Judy Ng
Jennifer Schore
Clara Soh

Submitted to:

Centers for Medicare & Medicaid Services
Office of Strategic Planning
C3-20-17
7500 Security Boulevard
Baltimore, MD 21244

Project Officer:
Carol 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

The analyses upon which this publication is based were performed under Contract Number 500-95-0047, entitled "Managed Care Research and Demonstration Task Order Contracts," sponsored by the Centers for Medicare & Medicaid Services, Department of Health and Human Services.

ACKNOWLEDGEMENTS

This report has benefited greatly from the cooperation, insights and support of many people. First among these are three CMS staff: Carol Magee (our project officer), Cynthia Mason (the project officer for the demonstration), and Renee Mentnech (the director of CMS's Research and Evaluations Group in the Office of Research, Development and Information). Their helpful comments, wise counsel, and collegial attitudes greatly enriched the report and our experience in producing it. We also thank the directors and staff of the 15 demonstration programs evaluated here. Their willingness to discuss the estimates of their program's effectiveness and give their opinions on the likely reasons for its successes or shortcomings enhanced our ability to assess what features really lead to effective care coordination. Finally, we thank our numerous colleagues at Mathematica Policy Research, whose assiduous programming, reviewing, editing, and typing were essential and greatly appreciated. Any remaining errors are the authors' own. The opinions reflected herein are those of the authors, and do not necessarily reflect the opinions of CMS, the U.S. Department of Health and Human Services, or the programs evaluated.

CONTENTS

CONTENTS *(continued)*

Chapter **Page**

CONTENTS *(continued)*

Chapter **Page**

CONTENTS *(continued)*

Chapter **Page**

TABLES

TABLES *(continued)*

TABLES *(continued)*

TABLES *(continued)*

FIGURES

EXHIBITS

EXECUTIVE SUMMARY

Care for beneficiaries with chronic illnesses, such as heart disease and diabetes, is a major expense to the Medicare program, and a major detriment to beneficiaries' quality of life. For example, just under one-half of all beneficiaries in 1997 were treated for one or more of eight categories of chronic illnesses, and they accounted for three-fourths of all Medicare spending in 1998 (Brown et al. 2004) Furthermore, beneficiaries often have multiple chronic illnesses, which compounds the cost and complexity of their care. The 12 percent with three or more of the eight chronic health problems accounted for one-third of all Medicare spending. Coordinating the care these patients require is difficult, because patients with chronic illnesses see an average of 11 different physicians per year (Anderson 2002). Despite these alarming statistics, many of the acute health problems caused by chronic illnesses can be prevented if (1) patients are provided with medical care that is consistent with recommended standards; (2) patients adhere to recommended diet, medication, exercise, and self-care regimens; and (3) providers communicate better with each other and with 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). This potential has led many health maintenance organizations and indemnity insurers to develop their own programs or to contract with disease management or case management providers for such programs (see Villagra and Ahmed 2004 for evidence of the effectiveness of disease management for diabetic patients in a managed care setting). However, the Medicare fee-for-service program does not cover such services.

The Medicare Coordinated Care Demonstration (MCCD) tests whether case management and disease management programs can lower costs and improve patient outcomes and well-being in the Medicare fee-for-service population. In January 2002, the Centers for Medicare & Medicaid Services (CMS) selected 15 demonstration programs in a competitive awards process, under which each was allowed to define its own intervention and target population, within broad parameters. Each program began enrolling patients between April and September of that year and was authorized to operate for 4 years. Beneficiaries who agree to participate are randomly assigned by the evaluator, Mathematica Policy Research, Inc., to either the treatment group, which received the intervention, or the control group. Both groups continued to obtain their traditional Medicare–covered services from fee-for-service providers in the usual manner.

This report synthesizes findings from the first 2 years of the demonstration programs' operations, focusing on program impacts over the first year after enrollment for beneficiaries who enrolled during the first year, and over the first 25 months of operations for all enrollees. Findings presented include program-specific estimates of impacts on (1) survey-based measures of patients' health status, knowledge, behavior, satisfaction with their health care, quality of care, and quality of life; and (2) claims-based measures of patients' Medicare service use and expenditures, and the quality of care received. The report links differences across programs in these impacts to differences in the interventions and the target populations in order to draw inferences about "what works" and "for whom." This synthesis of findings draws on an earlier

report to Congress that described the types of programs and beneficiaries participating in the demonstrations, the interventions the programs have implemented, and how well patients and physicians like the programs (Brown et al. 2004). This report updates that information and adds analyses of Medicare service use and expenditures and a scoring methodology developed specifically for this evaluation to rate the quality of each program's intervention on several dimensions.

The findings in brief indicate that patients and physicians were generally very satisfied with the program, but few programs had statistically detectable effects on patients' behavior or use of Medicare services. Treating only statistically significant treatment-control differences as evidence of program effects, the results show:

- Few effects on beneficiaries' overall satisfaction with care
- An increase in the percentage of beneficiaries reporting they received health education
- No clear effects on patients' adherence or self-care
- Favorable effects for only two programs each on: the quality of preventive care, the number of preventable hospitalizations, and patients' well-being
- A small but statistically significant reduction (about 2 percentage points) across all programs combined in the proportion of patients hospitalized during the year after enrollment
- Reduced number of hospitalizations for only 1 of the 15 programs over the first 25 months of program operations
- No reduction in expenditures for Medicare Part A and B services for any program

Despite the absence of statistically significant treatment-control differences in Medicare expenditures for traditional services, it is possible that some of the programs are cost neutral to date. This could be true because the large variation in Medicare expenditures and the small number of beneficiaries enrolled in some programs make it difficult to draw definitive conclusions—for nine programs, treatment-control differences over the first 25 months of operations are not statistically different from zero, but they are also not significantly different from the average fee paid to the programs. Based on the patterns of differences in hospitalizations, Medicare Part A and B expenditures, and total Medicare expenditures including the care coordination fees, six of the programs are not cost neutral, four probably are not, and five may be cost neutral, over their first 25 months of operations.

The results presented here are not the final word on the programs' impacts—changing ingrained behaviors of physicians and patients and improving communications among non-integrated fee-for-service providers are all difficult tasks to achieve. Furthermore, even if achieved, such improvements in the processes of care may not yield statistically discernable improvements in patients' well-being or reductions in Medicare costs over the first 2 years of

program operations. Thus, the estimates presented here may differ from those that will be observed over the full 4 years of operations. Nonetheless, this report provides (to our knowledge) the largest single random assignment study to date of disease management/case management programs, and only the second evaluation ever conducted of such programs in a Medicare fee-for-service setting (the first was Schore et al. 1999).

A. WHAT TYPES OF PROGRAMS AND BENEFICIARIES ARE PARTICIPATING?

The 15 MCCD programs were selected from 58 proposals responding to CMS's solicitation. Programs' hosts had to have experience operating a disease management or case management program that had been shown to reduce hospitalizations or costs in some population or setting. CMS took this approach to maximize the potential for showing, in a time-limited demonstration, that successful care coordination programs used in other settings (typically managed care) could be implemented in a Medicare fee-for-service environment. Each program is offered only to patients living in its catchment area and meeting its approved eligibility criteria—typically, having a particular chronic illness. (Some programs further restrict enrollment to patients who have had a hospitalization during the year or 2 years preceding enrollment.)

In return for providing the care coordination intervention described in its CMS-approved operational protocol, each program receives a negotiated monthly payment for each beneficiary who chooses to enroll and is randomized to the treatment group. Payments to the programs range from $50 per enrollee per month for low-risk patients with one or more of several chronic illnesses in one program to $437 per month for the first 9 months for all patients with congestive heart failure (CHF) enrolled in another program. The negotiated rates were based on the programs' estimates of the cost of their interventions; however, to increase the likelihood that each program would generate net savings to CMS, the rates also were tied to the projected costs of the programs' proposed target populations. If a 20-percent savings in these projected Medicare costs would not be enough to offset the cost of the intervention, either a program restricted the proposed target population to higher-risk cases (such as beneficiaries with a recent hospitalization) or CMS reduced the proposed program payment to meet this constraint. Five programs had monthly fees exceeding $300; six had fees below $175.

The evaluation's 2004 report to Congress on the MCCD showed that the 15 selected programs varied widely in their organizational structures, target populations, and interventions, and that they had varied levels of success in recruiting patients (Brown et al. 2004). The participating organizations include five commercial disease management vendors, three hospitals, three academic medical centers, an integrated delivery system, a hospice, a long-term care facility, and a retirement community (see Table 1). The programs operate in 16 states (mostly in the northeast and Midwest) and in the District of Columbia; five serve beneficiaries living in sparsely populated rural areas. The programs also vary widely in the numbers and types of chronic conditions they target, with six programs targeting only a single condition, three serving patients with less-specific problems (for example, high-risk patients identified from administrative data by an algorithm), and the six other programs falling between these two extremes. Ten programs required that a patient have a hospitalization for the target condition in the year (or less) prior to enrollment.

TABLE 1

CARE COORDINATION PROGRAMS PARTICIPATING IN THE EVALUATION

Host Organization (Average Monthly Fee)	Organization Type	Service Area	Targeted Diagnoses	Number Ever Enrolled After 12 and 24 Months	Medicare Expenditures per Month During Year Before Enrollment
Programs Starting in April 2002					
Carle Foundation ($159)	Integrated delivery system	Rural counties in east central Illinois and west central Indiana	Heart conditions, diabetes, chronic lung disease	2,283 2,642	$521
CenVaNet ($80)	Care coordination provider	Richmond, Virginia	Heart conditions, diabetes, chronic lung disease, cerebrovascular disease	1,074 1,305	$953
Charlestown Retirement Community ($244)	Retirement community	3 retirement communities in the Baltimore area	Heart conditions, diabetes, COPD	430 802	$1,159
Health Quality Partners ($108)	Care coordination provider	Eastern Pennsylvania (rural)	Heart conditions, diabetes, asthma, moderate to severe hyperlipidemia or hypertension	498 1,140	$414
Medical Care Development ($297)	Hospital consortium	Rural areas of Maine	Heart conditions	393 876	$1,718
Mercy Medical Center/North Iowa ($257)	Hospital	Rural areas of Iowa	CHF, chronic lung disease, liver disease, stroke, vascular disease, renal failure	627 865	$1,315
Programs Starting in June 2002					
Avera Research Institute/Avera McKennan Hospital and University Health Center ($316)	Hospital	Rural counties in Iowa, Minnesota, Nebraska, and South Dakota	CHF	318 624	$1,615
CorSolutions ($444)	Care coordination provider	Harris County (Houston), Texas	CHF	671 2,162	$2,644
Georgetown University Medical School ($320)	Academic institution	Washington, DC, and parts of Maryland and Virginia	CHF	108 199	$2,530
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	543 766	$1,450
University of Maryland Medical School ($350)	Academic institution	Baltimore	CHF	58 137	$3,299
Programs Starting in July Through September 2002					
Hospice of the Valley ($224)	Hospice	Maricopa County, Arizona (greater Phoenix)	CHF, COPD, cancer, neurological conditions	470 814	$2,174
QMed ($96)	Care coordination provider	2 counties in northern California	CAD	1,404 1,454	$507
Washington University School of Medicine ($173)	Academic institution with care coordination provider	St. Louis, Missouri	High-risk patients who are clinically unstable , targeted through proprietary algorithm	1,425 2,038	$2,263
Quality Oncology, Inc. ($140)	Care coordination provider	Broward County, Florida (Miami)	Cancer	63 141	$2,885

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

The mix of patients enrolled varied across programs on some characteristics, but on others the programs were quite similar.

- The most common primary conditions of program patients were CHF (29 percent of enrollees), coronary artery disease (CAD) (24 percent), and diabetes (13 percent)
- Four programs drew a high proportion of beneficiaries who were older than age 85, and one program targeted and enrolled a high proportion of younger beneficiaries with disabilities
- Compared with all Medicare beneficiaries, the programs' patients generally were substantially more highly educated and had higher incomes
- Most programs enrolled relatively few black or Hispanic patients, few patients younger than age 65, and few patients who also were enrolled in Medicaid

Many of the programs had unexpected difficulty enrolling the target number of patients, with only four exceeding the first-year target of 686 patients that was set by Mathematica Policy Research, Inc. as being the minimum necessary for the evaluation (although two others had over 600; see Table 1). Several programs enrolled less than one-half their targeted number of patients for the first year, citing initial overestimates of the number of eligible patients from their referral sources, physicians' failure to encourage their patients to enroll, high patient refusal rates, and limited care coordinator time to both recruit patients and serve those already enrolled. The programs that were most successful in enrolling patients were those that had a close relationship with physicians before the demonstration started and those with access to databases (such as clinic or hospital records) to identify potentially eligible patients. By the end of the second year, 12 of the 15 programs had over 600 enrollees.

Most of the programs succeeded in enrolling patients with serious chronic illnesses, but a few programs enrolled relatively healthy patients. Preenrollment Medicare expenditures averaged more than $2,000 per month during the year preceding enrollment for first-year participants in six programs, but less than $600 per month for three other programs (average Medicare expenditures for noninstitutionalized beneficiaries nationally was $505 per month in 2002). The programs with low-cost enrollees are likely to have difficulty achieving large enough savings to offset the cost of their interventions. In one-half (eight) of the programs, enrolled patients had an average of one or more hospitalizations per year during the 2 years before enrollment. (Three of these programs averaged two hospitalizations per patient per year.) In 13 of the programs, the enrolled patients had higher costs than did diagnostically eligible nonparticipants in the same geographic area during that year. However, the two programs whose enrollees had the lowest preenrollment Medicare costs (about $500 per month) enrolled patients with preenrollment costs and admission rates that were lower than those of eligible nonparticipants. The program with the greatest preenrollment discrepancy between participants and nonparticipants enrolled sizable numbers of beneficiaries it identified as eligible through chart reviews, but many of these enrollees did not meet diagnostic eligibility criteria according to claims data examined here, for the year prior to enrollment.

B. WHAT INTERVENTIONS ARE THE PROGRAMS DELIVERING, AND HOW ARE THEY DOING IT?

The 15 programs differed widely in both how they implemented their care coordination interventions with patients and their involvement with patients' physicians and other providers. Information about the interventions came from interviews with program staff as well as data recorded by care managers on their contacts with patients. Interviews were conducted at three points: by telephone at 3 months after enrollment began; during in-person visits 6 months after the telephone interviews; and by telephone again, roughly 3 years after startup.

The programs differed in their relative emphasis on four major vehicles for achieving better outcomes for patients: improving patients' adherence to treatment and self-care regimens, improving coordination and communication among providers, improving physician practice, and increasing access to support services. All but 1 of the 15 programs stressed patient education to improve adherence and coordination, but most devoted less attention to convincing physicians to change their practices or to improving access to support services.

The programs varied greatly in their approach to care coordination. They differed on the mode and intensity of contacts, staff credentials, ratio of staff to patients, method of monitoring, patient education methods, and approaches to improving communications between physicians and patients and among physicians.

All but two programs required all their care coordinators to be registered nurses, but caseload size varied widely. Thirteen programs required care coordinators also to have specific experience with cardiac, geriatric, medical-surgical, or community nursing. Caseload sizes ranged from a low of 36 patients per care coordinator to a high of 200.

All programs began care coordination with assessments of patients' needs and condition, after which they developed patient care plans. Of the 15 programs, 12 conducted at least part of their assessment in person, even though most of their invention was conducted over the telephone. Ten programs initiated their assessments within 3 weeks after enrollment on average. Only one program (Jewish Home and Hospital) took longer than 6 weeks on average to begin its assessment. The assessments culminated in care plans to fill the gaps in the patients' knowledge and treatment. These plans were developed collaboratively with patients and, when appropriate, with the patients' families.

Most (12) of the programs contacted patients one to three times per month on average (mostly by telephone), but 2 had more frequent contacts. Six of the programs averaged 1.2 to 1.5 contacts per patient per month during the first year after enrollment; another six averaged between 2.2 and 2.9. Avera, however, contacted patients over 8 times per month on average. The great majority of contacts were by telephone, except in Mercy, whose care coordinators conducted over two-thirds of their contacts in person. Patients initiated about 10 percent or less of the contacts in most programs.

Six programs used home telemonitoring devices, although three of these did so on a very limited basis. Electronic devices transmitted patients' weights, other clinical indicators, and symptom reports to their care coordinators on a daily basis. A seventh program provided

ambulatory ischemia monitoring. In addition, 13 programs required care coordinators to contact all of their patients at least monthly by telephone or in person.

All but one program provided patient education; almost all used standard curricula and had processes for assessing the effectiveness of the education. Over 85 percent of enrollees in the 14 programs featuring patient education (only the University of Maryland did not) received contacts for educational purposes during their first year in the program. The educational materials were part of electronic databases for some of the programs, and some assessed patients to identify specific learning barriers. Programs assessed effectiveness by reviewing clinical indicators or home monitoring data for evidence of improving health or relied on patients' self-reported behavior changes or responses to questions about their knowledge.

Most programs sought to improve communication between patients and providers by training patients, and they sent physicians regular written reports on patients. Some programs taught patients to take prepared lists of questions to their office visits, while others gave them schedules of tests they should be receiving. While most programs communicated with patients' physicians via written reports, one held formal conferences with participating physicians and one had its quality manager visit physicians to discuss adherence to evidence-based practice, using data obtained from ambulatory ischemia monitoring and physicians' medical records to make their point. Five programs had care coordinators practice in the same location as physicians, enhancing the coordinators' ability to communicate face-to-face with them. Seven programs arranged to have hospitals notify care coordinators when the hospital admitted program enrollees or had their care coordinators review hospital or emergency room (ER) admission lists.

Programs sought to minimize the burden on patients' physicians. Only four of the programs listed improvement of provider practice as one of their approaches for improving patient health. They did so mostly by providing recommendations for specific patients when treatment plans deviated from evidence-based guidelines. One program, however, provided education about such guidelines and offered physicians incentives to participate. Some programs used opinion leaders or advisory boards to encourage physicians' active participation, paying either a monthly stipend per patient ($20 to $30 typically) or paying for participation in meetings or for delivery of medical records.

The programs devoted relatively little attention to increasing patients' access to needed support services. All but one program provided such assistance, such as referring patients to transportation services or home-delivered meals, but only five ever did so for more than one-half their patients.

Programs varied widely in the sophistication of their electronic systems to manage data on patients and program activities. Thirteen programs used these systems to support their work with patients. Among those, 11 generated reports from those systems reminding coordinators about when to contact patients, and 12 used the systems to provide reports on patients' clinical indicators and outcomes.

While information on *what* programs are doing as their interventions and how they do it can be useful for understanding why some programs are more effective than others, it may be more important to know how heavily they focus on particular dimensions of care coordination and

how well designed the interventions are on these dimensions. On the surface, many of the programs in this demonstration appear to implement quite similar interventions, yet in-depth discussions with the programs reveal a number of important differences in the intensity of their intended efforts to provide patient education or service arrangement or other possible components of their intervention. To address this issue, the evaluation developed a scoring algorithm for rating each program's interventions on 10 separate domains:

- Program Staffing
- Initial Assessment
- Patient Education
- Improving Communication and Coordination
- Improving Provider Practice
- Service and Resource Arrangement
- Information Technology and Electronic Records
- Ongoing Monitoring
- Quality Management and Outcome Measurement

These ratings were developed independently of the survey and claims data on program outcomes, and without regard to data on contacts supplied by the programs. Researchers scoring the programs relied solely on the information collected during in-person and telephone discussions with the programs about their intervention; estimates of program impacts were not shared with scorers until after they had completed their ratings. Scores were normalized to range from 0 (intervention did not address this domain) to 100 (intervention was extremely well-designed on this domain).

Programs varied widely on each of these domains, especially Quality Management and Outcome Measurement, for which scores ranged from 5 to 91, and Improving Provider Practice, which ranged from 0 to 77. Scores varied less widely across programs on the Problem Identification and the Initial Assessment domains. Average scores were highest for the Initial Assessment and the Monitoring domains, and lowest on average for Improving Provider Practice, reflecting the lesser attention given to this area by most of the programs.

While individual programs often scored extremely well on some domains and poorly on others (at times because a particular domain was not part of its intervention), a few programs had high scores on several domains and others had consistently low ratings across most of the domains. Carle was scored in the top quintile of programs (the 3 highest) on 6 of the 10 domains, and Mercy and Quality Oncology each had 4 scores in the top quintile. The Jewish Home and Hospital and the University of Maryland scored in the bottom quintile on nine and seven of the domains, respectively. Yet both of these programs scored in the top quintile on one domain each. The importance of these rankings is not to identify those programs that do particularly well or poorly across measures, but to determine whether having strong designs in certain domains is consistently associated with having favorable impacts on Medicare costs or the quality of care.

C. HOW DO PATIENTS AND PHYSICIANS LIKE AND RESPOND TO THE PROGRAMS?

Survey data collected on patients in the 12 programs with over 300 enrollees by the end of their first year and on enrollees' physicians in all 15 programs suggest that the programs are popular with both patients and physicians. The patient surveys generally were conducted 7 to 12 months after patients enrolled. Physicians were surveyed in two waves, once about 12 to 15 months after the program in which their patients were enrolled began operations, and a second wave about 18 to 21 months after program startup.

About two-thirds of treatment group patients on average across programs were aware of the program; 15 percent of control group members also reported receiving some care management. Most treatment group members were aware they were receiving care coordination although the percentage varied widely across programs, ranging from only 30 percent in QMed to 81 percent in Mercy saying "yes" when asked, "During the past 12 months, did someone like a nurse, social worker or geriatric nurse help arrange or coordinate your health care?" Thus, the programs generally were successful in establishing a relationship with the patients. However, 3 to 28 percent of the control group also answered "yes" to this question, suggesting that the interventions are not the only source of professional care coordination assistance available in the programs' service areas. Among those saying they received this type of assistance, those in the treatment group generally reported higher levels of satisfaction with the help received. Nonetheless, the fact that about one-third of treatment group members did not report receiving care coordination and that some control group members reported they did receive such assistance makes it more difficult for the programs to demonstrate a significant impact on the treatment group.

Treatment group patients were generally very satisfied with the care coordination they received. Coordinators were rated on four different dimensions—support and monitoring, knowledge and ability to get answers, ability to explain adherence to recommended self-care, and help arranging services—each with three or four specific indicators. About one-third to one-half of the patients surveyed rated their coordinators as excellent on the 14 indicators examined, and most of the rest rated them as "very good." Very few patients (less than 10 percent in nearly all instances) rated the programs as only fair or poor on any of the measures. Care coordinators received especially high marks on indicators of the emotional support and monitoring they offered, especially their "caring attitude," with over 60 percent of the patients on average giving their programs an excellent rating. Patients also rated programs highly on staying in touch (over one-half rating it excellent, on average). Patients gave somewhat lower, but still quite positive, ratings on average for programs' including them and their families in decisions, and for helping them cope with their illness and avoid complications.

Patients were somewhat less impressed with the help they received from programs in arranging appointments or services. Across most programs, about 35 to 40 percent of the patients gave an excellent rating. Exceptions include Carle's higher ratings, and two programs that received markedly lower ratings (the same two programs with low marks on support and monitoring). Substantial minorities of patients (10 to 24 percent) gave the programs a fair or poor rating on this domain. These less favorable ratings are likely to be due to most programs' focusing their attention more on monitoring and education than on arranging services.

Patients had high praise for the care coordinators' knowledge. Over one-half the patients on average rated care coordinators' knowledge as excellent, and only two programs had less than 43 percent giving an excellent rating. About 40 to 43 percent of patients on average rated their programs' care coordinators as excellent on their ability to explain symptoms or get physicians to answer questions or help them to identify early warning symptoms; these rates were similar across most programs.

Finally, a modest proportion of patients gave excellent ratings to care coordinators' ability to explain recommended diet, medication, and exercise regimens. Of all the measures, patients were least likely to give coordinators very high marks on their ability to explain exercise regimens (although few patients rated the programs as fair or poor). The somewhat less enthusiastic ratings on these measures may be due to care coordinators' focusing their education efforts less intensely on exercise than on other patient behaviors.

Overall, a consistent pattern emerges from these numerous patient ratings of the care coordination interventions, with Health Quality Partners consistently receiving notably higher marks than other programs. These high patient ratings were consistent with the evaluation's scoring results, in which Health Quality Partners had the highest score of the 15 programs on patient education and ranked among the top on monitoring as well. Carle and Avera also were rated highly on some patient survey measures, especially those related to providing emotional support and service arrangement. Avera's high ratings on explaining early warning signs is consistent with the scoring algorithm's strong ranking of this program (third highest) on patient monitoring and its use of home telemonitoring, which likely generated follow-up conversations between care coordinators and patients about heart failure and symptoms. Carle's high ratings from patients on getting answers from physicians is consistent with its top score among all programs on improving communications and coordination among providers and the relatively close relationship its program staff had with their patients' physicians. Carle's patients' high ratings of the program on service arrangement is also consistent with Carle having one of the top three scores on service arrangement in the scoring algorithm.

Most of the programs received high ratings from their patients' physicians on most dimensions, although there were clear differences across the dimensions and across programs. Physicians were asked to rate the programs on numerous factors, including their effects on the physician's practice (medical practice, time and paperwork burden, and financial impact if any), patients' education and behavior, service arrangements for patients, care coordination, physicians' relationship with patients, and patient outcomes and behavior. Physicians were also asked to rate care coordinators' clinical competence.

Program physicians widely agreed that the programs made things easier overall for patients and did a good job of monitoring and followup, but they were not always as positive about the usefulness of program reports (42 percent responded these were "very useful") or about the programs' effects on other aspects of their practice. Table 2 illustrates the wide range of responses across programs. Similar wide variation across measures and programs occurred in each of the other categories.

TABLE 2

PHYSICIANS' SATISFACTION WITH CARE COORDINATION
(Percentages)

	Mean	Minimum	Maximum
Program Reports on Patients Very Useful	42	0	91
Medical Practices a Little or a Lot Better on:			
Reducing problems with polypharmacy	56	11	81
Reducing telephone time	55	4	88
Making things easier for staff	56	22	86
Making care more evidence-based	49	20	95
Making it easier overall to care for patients	75	33	100
Monitoring and Followup Very Good/Excellent	71	38	100

Note: The mean is the average across the 15 programs.

The patients' primary physicians in general were pleased with the program overall. Across the 15 programs, on average, a majority (67 percent) of physicians felt that the program increased patients' overall quality of care, and 80 percent said they would recommend the program to patients and colleagues (about 60 percent said they would "definitely" recommend the program and the remainder said they would "probably" recommend the program).

There were some major variations across programs in physicians' ratings. For example, 95 percent of physicians in Charlestown found the program improved patients' quality of care and would definitely recommend the program to others, while only 11 percent of physicians in Quality Oncology were as impressed on either measure. Charlestown consistently received higher ratings from its patients' physicians than did other programs, while three programs (CenVaNet, QMed, and Quality Oncology) consistently received lower ratings from their physicians than did the other programs.

In general, physician satisfaction ratings corresponded with scoring algorithm results based on discussions with program staff and physicians. For example, Carle's and Charlestown's physicians, who consistently rated their programs more highly than did the others on physician practice effects, also scored in the top quintile for improving provider practice (Carle being the top scorer in this category). Mercy's program received higher physician ratings than the other programs on perceived service arrangement and care coordination effects, consistent with its scoring in the top two quintiles for the categories of service and resource arranging, and for improving communication and coordination. Similarly, at the other end of the spectrum, QMed and Quality Oncology had overall physician satisfaction ratings across all categories that were consistently lower than the cross-program average by more than 1 standard deviation, which coincides with their scoring algorithm ratings that place them in the bottom two quintiles.

Few significant differences were observed between treatment and control group members on satisfaction with the process of care. Despite the generally favorable rating that treatment group patients and physicians gave to most of the programs' care coordination efforts, the treatment group did not consistently report higher satisfaction than control group members with indicators of the quality of the health care they received from the various providers they saw. The indicators include ratings of the degree of choice in treatment that patients feel they have, the extent to which providers keep in touch with each other, the explanations received from specialists, explanations of side effects, explanations of treatments, explanations of tests, and the quickness of receiving test results. The treatment group members were significantly more likely than the control group members to report feeling they had a choice in the treatment of their condition in only 1 of the 12 programs included in the survey (Avera). Differences favoring the treatment group occurred most often for providers keeping in touch (5 of the 12 programs). Treatment group members in four of the programs also gave more favorable ratings than the corresponding control group on explanation of treatments. Satisfaction with explanations of side effects and explanations from specialists were significantly greater for the treatment than the control group for only two and three of the programs, respectively, and with explanation of tests for only one. None of the programs had impacts on the timeliness with which test results were delivered, according to the treatment-control differences.

A few programs appeared to have more impact than others on patients' satisfaction with their overall 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 had significant differences on two of the six measures, and three programs had significant effects on one of the measures. The four other programs included in the survey had no discernable effect on patients' satisfaction with care.

D. HOW DO THE PROGRAMS AFFECT ADHERENCE AND QUALITY OF CARE?

The care coordination programs were expected to improve patients' adherence to recommendations and their quality of care, which, in turn, was expected to lead to improvements in patients' health and well-being. The evaluation compared the treatment and control groups' receipt of health education, knowledge and behavior about self-care, quality of care, and health status and well-being to determine whether the programs had the intended effects. Measures of preventive care and preventable hospitalizations over the year after enrollment were constructed from Medicare claims data for all first-year enrollees enrolled in 14 programs. (The measures were not appropriate for Quality Oncology, which targeted patients with cancer.) The analysis also draws on the patient survey responses to examine receipt of education, knowledge, behavior, adherence, receipt of care, and functioning. Table 3 summarizes the results.

Overall, the programs appeared to have no consistent discernible effect across numerous measures of behaviors and outcomes except receipt of health education. 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 true effects in any area besides receipt of health education. Favorable effects were observed for 1 or 2 measures of health status and well-being (out of the 9 examined) for 8 of the 12 programs.

TABLE 3

TREATMENT-CONTROL DIFFERENCES ON QUALITY OF CARE, AMONG FIRST YEAR ENROLLEES

Category of Outcomes (Number of Measures)	**AVE**	**CAR**	**CEN**	**CCI**	**COR**	**GEO**	**HOS**	**HQP**	**JHH**	**MCD**	**MER**	**QMD**	**UMD**	**WSH**
Health Education (5)	●●	●●●●	●●●●●		●●		●	●●●●●	○	●●	●●●●●	●		●○
Knowledge and Behavior (8)	○		●					●●	○	●		●○		●
Service Arrangement and Unmet Needs (7)				●	●●		●	●			○	○		
General and Disease-Specific Preventive Care (12)	○○	●●●●	●○				○○	●●●●						○
General and Disease-Specific Potentially Preventable Hospitalizations (8)	○		●●○			●●	●○	●		○				○○
Functional Status (9)	○	○○○		○	○				○	●	●	●		○
Health Status and Well-being (8)	●		●	○	●●●		●●	●	●	○	●	●		

Source: Treatment-control differences from patient survey and Medicare claims data. The Georgetown 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, and so it is not shown.

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.

Abbreviations for MCCD Programs (Columns)
AVE = Avera
CAR = Carle
CEN = CenVaNet
CCI = Charlestown
COR = CorSolutions
GEO = Georgetown University
HOS = Hospice of the Valley
HQP = Health Quality Partners
JHH = Jewish Home and Hospital
MCD = Medical Care Development
MER = Mercy
QMD = QMed
UMD = University of Maryland
WSH = Washington University

The large effects on health education did not lead to effects on self-reported knowledge, adherence, or health-related behaviors. The treatment groups in all but 1 of the 12 surveyed programs were significantly more likely than their corresponding control groups to report having received education on health behaviors. The most common effects were on receipt of education about diet and exercise, followed by the receipt of health educational materials, education on recognizing when to seek urgent care, and education on the importance of medication adherence.

Four of the programs (Carle, CenVaNet, Health Quality Partners, and Mercy) had favorable treatment-control differences across four or more of the five measures of patient education examined. Despite the treatment group members being more likely to say 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 one or two programs. Across measures, four programs (Carle, CenVaNet, Health Quality Partners, and CorSolutions) had somewhat more favorable treatment-control differences than the other programs.

The programs had no discernable effects on service arrangements or unmet needs. While treatment group members in all 12 programs included in the survey were more likely than control group members to report receiving care coordination services (not included in Table 3), as intended, only 3 programs exhibited significant favorable treatment-control differences on other measures of unmet needs or service arrangements. Furthermore, two programs each had one outcome measure for which the *control* group had significantly better outcomes than the treatment group.

Only two programs appear to have made clear improvements in the quality of preventive care (Carle and Health Quality Partners), or to have reduced the number of preventable hospitalizations (Georgetown and Hospice of the Valley). The treatment groups were more likely than the control groups in Carle and Health Quality Partners to receive vaccination and (for women) screening mammography, and recommended blood and urine tests among beneficiaries with diabetes and coronary disease. Georgetown and Hospice of the Valley had significantly fewer "preventable" hospitalizations per beneficiary overall in their treatment groups than in the respective control groups. (Potentially preventable hospitalizations are inpatient admissions for common, acute medical conditions that, in the consensus of expert clinicians, generally should not progress to requiring inpatient care if treated in a timely fashion with adequate outpatient primary care; see Kozak et al. 2001.)

Only two programs (CorSolutions and Hospice of the Valley) had favorable effects on multiple measures of patient well-being, and these were only for selected measures. 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. In addition, only three programs had a favorable treatment-control difference on any of the nine survey-based measures of functioning (for example, ability to eat independently), and, for six programs, the treatment group reported significantly *worse* health status on one or more measures. However, it is difficult to conceive of a mechanism by which programs would adversely affect patients' functioning. Furthermore, one

should expect about one-half the sites to have one significant negative estimate out 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.

E. HOW DO THE PROGRAMS AFFECT MEDICARE SERVICE USE AND COST?

By improving patient adherence, the timeliness of response to worsening symptoms, or other aspects of the quality of care, care coordination programs are expected to reduce hospitalizations, the key factor in reducing Medicare expenditures for beneficiaries with chronic illnesses. On the one hand, the need for emergency room care and other expensive Medicare services that often follow hospitalizations (such as that provided by skilled nursing facilities and home health agencies) may also be reduced. On the other hand, some types of service use and expenditures could increase if the programs increase patients' visits to physicians for preventive care or to address symptom exacerbations. To measure these effects, the evaluation compared the treatment and control groups in each program on Medicare service utilization and expenditures. The measures were constructed for the year after enrollment for patients enrolled during the first year of program operations, and for all patients during the programs' first 25 months of operations.

Only 1 of the 15 programs (Mercy) showed a statistically significant reduction in hospitalizations, and none of the programs had significantly lower expenditures for Medicare Part A and Part B services. In eight other programs, the treatment group had fewer hospitalizations than controls during the first 25 months of program operations, but the observed differences could not be attributed with confidence to the intervention, rather than to chance. Four of these programs had 10 to 18 percent fewer hospitalizations among treatment group members than among control group members, but none of these differences were statistically significant. Furthermore, there was no difference in Medicare expenditures for two of these programs and the other two had very few enrollees. However, hospitalizations and Medicare expenditures were 14 and 21 percent higher, respectively, for the treatment than the control group in the Charlestown program, the only program for which a statistically significant difference in expenditures was observed. For the first year after enrollment, for all programs combined, the treatment group had 2 percent fewer patients admitted to the hospital, a statistically significant difference, but the differences in both the number of hospitalizations and Medicare expenditures were very small and not significantly different from zero.

The treatment group's significantly (27 percent) lower hospitalization rate in Mercy did not result in a statistically significant difference in Medicare expenditures, although expenditures were 13 percent lower for the treatment group over the 25-month period since startup. Medicare expenditures for Part A and B services were lower for the treatment group than the control group by at least 10 percent in two other programs (Georgetown and QMed), but neither difference was close to being statistically significant.

Cost neutrality cannot be rejected for some of the programs. These results suggest that none of the demonstration programs is cost neutral—that is, none has generated statistically significant evidence of savings in Medicare expenditures that could offset the fees paid to the

program. However, that conclusion is less clear when one considers the large variance in the estimates. That is, while the evaluation cannot reject the hypothesis that savings in Medicare Part A and B expenditures are zero, for some programs it also cannot reject the hypothesis that savings are large enough to cover the average fee paid to the programs for care coordination. For six programs (shown in the bottom panel of Table 4), cost neutrality can be rejected statistically—net costs have increased for these programs. For the nine other programs, the evaluation cannot formally reject the hypothesis that total average Medicare expenditures per month for the treatment group, including the care coordination fee, are equal to expenditures for the control group (cost neutrality). However, for the four programs with small (less than 10 percent) treatment-control differences in hospitalization or expenditures on Part A and B services, cost neutrality seems unlikely. Failure to reject the cost neutrality hypothesis in these cases may be due to low statistical power resulting from small sample sizes and high variance of Medicare expenditures. Four other programs, however, have treatment-control differences in hospital admissions of 10 percent or greater, and (in two cases) differences in Part A and B expenditures that are large enough to essentially offset the fees. The difference in hospitalizations is smaller for a fifth program (QMed), but the fee for this program is quite low and is almost fully offset by the treatment group's 12 percent lower Medicare expenditures for traditional services. Thus, these five programs may actually be generating savings in Part A and B expenditures that are sufficient to offset the program fees. However, the estimates are too imprecise at this time for the evaluation to definitively conclude that there are such savings, or that they are large enough to cover the average fee paid for care coordination.

F. SYNTHESIZING THE FINDINGS: WHAT WORKS, AND WHAT DOESN'T?

Given that few of the programs have shown convincing evidence to date of reducing beneficiaries' need for hospitalizations and saving money or of improving the quality of care received, there is relatively little assessment that can be done yet of "what works." The one program for which there were statistically significant estimates of reductions in hospital use (Mercy Medical Center in Iowa) differed from the other programs in that it had by far the highest proportion of contacts conducted in person (two-thirds), and it excelled at Problem Identification and Care Planning, Patient Education, and Improving Communications and Coordination between patients and physicians. The program also had large impacts on patient education, as judged from the patient survey, and was rated highly by the patients' physicians.

In the evaluation's follow-up discussions with the programs, Mercy's staff attributed the reductions in hospitalizations they achieved primarily to getting patients to see their physicians quickly when symptoms worsened or problems arose. By identifying looming problems before they became severe and convincing patients of the urgency of seeing a physician (or contacting physicians directly on behalf of patients when necessary), Mercy staff felt they were able to prevent the patients' health from deteriorating to the point where a hospital admission would be necessary. They felt this preventive effect typically arose through quickly getting patients on needed medications or different dosages of their current medications.

The four other programs for which the treatment group had 10 to 20 percent fewer hospitalizations than the control group (although these differences were not statistically significant) also scored highly on one or more domains. For example, Georgetown and Health Quality Partners both scored in the top quintile on Initial Assessment. Quality Oncology scored

TABLE 4

TREATMENT-CONTROL DIFFERENCES IN HOSPITALIZATIONS AND MEDICARE EXPENDITURES, WITH AND WITHOUT PROGRAM FEES, OVER THE 1ST 25 MONTHS OF PROGRAM OPERATIONS
(Percentages)

		Monthly Medicare Expenditures	
	Annual Number of Hospitalizations	Without Care Coordination Fees	Including Care Coordination
May Be Cost Neutral			
Mercy Medical Center	***-27***	-13	8
Quality Oncology	-18	-2	0
Hospice of the Valley	-14	0	9
Georgetown University	-12	-12	1
QMed	-4	-12	1
Probably Not Cost Neutral			
CorSolutions	-5	-8	4
University of Maryland	-1	0	10
Jewish Home and Hospital	0	-6	8
Medical Care Development	1	-2	9
Not Cost Neutral			
Health Quality Partners	-10	0	***17***
Carle Foundation	-4	-1	***21***
Avera	4	-5	***14***
CenVaNet	4	6	***14***
Washington University	6	4	***12***
Charlestown	14	***21***	***44***
Overall	**-4**	**-2**	***11***

Note: Bolded italicized numbers denote statistically significant treatment-control differences at the 10-percent level for hospitalizations and Medicare expenditures without fees, and at the 20-percent level for expenditures including care coordination fees. Negative estimates imply that hospitalizations or Medicare expenditures (with or without the fee included) are lower for the treatment group, a favorable outcome. Positive estimates suggest that the treatment group used more services and cost Medicare more than the control group.

in the top quintile on four domains—Staffing, Information Technology, Ongoing Monitoring, and Quality Management.

Programs that seemed to improve preventive care (Carle and Health Quality Partners) also scored well on patient survey indicators and tended to receive high ratings on the scoring algorithm for various aspects of their interventions. Carle scored higher than all other programs on 5 of the 10 indicators. Health Quality Partners scored at the top on patient education. However, neither of these programs generated reductions in Medicare expenditures for traditional services; thus, both significantly increased net costs to Medicare. This lack of reduction in expenditures may be due in part to the fact that patients in these two programs had far lower preenrollment Medicare expenditures than all but one of the other programs.

Programs that exhibited no effects on hospitalizations, costs, or quality-of-care indicators gave a range of reasons why they were unable to reduce the need for hospitalizations. Reasons included the still-short time frame over which the analysis was conducted; the belief that some of their patients were either too debilitated or not sick enough to benefit from their interventions; and the belief that physicians in their service areas had an intractable tendency to send patients to the emergency room, which is more expensive, rather than to find time for office visits when patients exhibited worsening symptoms.

Looking across the characteristics of the five programs most likely to be cost neutral over the first 2 years of operation and the two that appear to have improved the quality of care seems to confirm the finding in Chen et al. (2000) that no single program feature or characteristic seems to be associated with a greater likelihood of program "success." Nor does the absence of a particular feature seem to doom a program to relative failure. However, how *well* programs perform their functions (based on information obtained from program staff and assessed by the evaluator) does appear to be associated with program success.

While no firm conclusions can be drawn as yet about which MCCD programs really are effective (because samples are still relatively small and the follow-up period relatively short), those programs that are most promising to date share few common structural features. Two of the programs with the most success in improving quality (Health Quality Partners and Carle) operate in rural areas, as does Mercy, the sole program with statistically significant effects on the number of hospitalizations. Yet Avera and Medical Care Development also operate in rural areas and show no such promising results to date. Two of the programs with the most favorable expenditure results (Quality Oncology and Georgetown) have fewer than 100 treatment group members—Medical Care Development is the only other program serving fewer than 300 patients. However, the results for these two programs may be due more to the imprecision of the estimates than to the excellence of the interventions. The five other relatively promising programs have substantially more patients. All four programs whose care coordinators have average caseloads of 50 or fewer patients are among the most effective programs, but the three other relatively effective programs have average caseloads in the highest range (over 75 patients). Three of the five programs operated by commercial disease management programs were among the seven promising programs, but the four other promising programs had hospitals, clinics, or academic medical centers as hosts. Other program characteristics examined seem equally unrelated to whether a program was one of the more effective seven.

How well designed programs were on various dimensions appeared to have a somewhat stronger association with performance than did structural characteristics. Strong performance in

any particular domain does not appear to be necessary or sufficient for a program to be relatively successful. However, there are some clear patterns of association between how programs scored on the 10 domains examined and the programs' ability to improve quality or generate reasonably favorable expenditure comparisons. The domains most strongly associated with the promising programs are Staffing (the five programs with the highest ratings on staffing were all among the seven most promising programs), Improving Communications and Coordination (five of the six top programs on this domain were promising programs), Patient Education (four of the top five programs were promising), and Quality Management and Outcome Measurement (four of the top five programs were promising). Characteristics decidedly *not* associated with stronger quality or cost performance included Improving Provider Practice, Service and Resource Arranging, Information Technology, and (perhaps surprisingly) Ongoing Monitoring. For each of these characteristics, only one or two of the five top-rated programs were among the seven programs classified as most promising to date.

Finally, the characteristics of the patients enrolled appeared to be unrelated to the relative success of the programs to date. Three of the seven promising programs targeted patients with a single disease; the other four targeted multiple diseases. All three of the programs that enrolled patients with average preenrollment Medicare expenditures of under $600 per month were among the top seven performers, but three others of the top performers were among the six programs whose patients had average expenditures in excess of $2,000 per month. None of the other patient characteristics examined (age, education, income, race) appeared to be related to programs' likelihood of success.

The current findings suggest that hiring excellent staff and performing certain key functions well are the most important determinants of the likelihood that a program might successfully improve patient outcomes or save enough in Medicare expenditures to cover the cost of its intervention. The results to date are thus consistent with findings from Chen et al. (2000) that a few factors were common to most successful programs, including hiring well-trained, experienced nurses with at least a baccalaureate degree, but many other factors, such as having sophisticated electronic health records, were not required.

G. LONGER FOLLOWUP AND MORE OBSERVATIONS ARE NEEDED FOR DEFINITIVE FINDINGS

Due to the small sample sizes, the high variability in Medicare costs, and (for some programs) the small amount of savings required to cover the cost of the intervention, there remains uncertainty over whether nine of the programs generate savings, and if so, whether they are large enough to offset the fees. We cannot conclude with confidence that any of the programs *generate savings* in Medicare expenditures on the normal Part A and B services, because none of the estimated treatment-control differences are significantly different from zero. However, we also cannot conclude with confidence that these programs *increase net costs* to CMS. That is, there is a nontrivial possibility that these programs do generate enough savings in Medicare Part A and B expenditures to offset the modest program fees (typically 2 to 13 percent of the Part A and B expenditures), despite the fact that none of the estimates of such savings are statistically significant. The wide confidence intervals around the estimated savings in Part A and B expenditures encompass both zero (implying no effect) and the average fee paid (implying savings large enough to offset the fee). The conservative inference is that the programs were not

cost neutral over the first 25 months, but there is a substantial possibility**,** given that the statistical power to detect true net savings in these nine programs ranges from only 11 percent to 77 percent, that such a conclusion is not correct for some of the programs. Furthermore, effects may yet emerge for some programs as the program and patients gain more experience and as any cumulative effects of the interventions on patient and provider behavior begin to be reflected in outcomes.

Although none of the impact estimates available at this time suggest that the demonstration programs are having large effects on patients' behaviors or outcomes, effects on Medicare service use and expenditures might be observed when the full 4 years of data on all patients become available. Physicians have been responding favorably to the programs—an important factor, given the widespread recognition that few care coordination programs are likely to succeed without significant cooperation and reinforcement from patients' physicians (Chen et al. 2000; Schore et al. 1999). Even more important, patients appear to have formed a bond with their care coordinators, and to trust their judgment.

The absence of large effects on the patient adherence measures may be somewhat discouraging for programs, but it does not necessarily imply that the programs are having no effect on patients' behavior. Relative to the control group, patients of several programs reported better access to information and appointments and better communication among their providers. Furthermore, the finding that program patients were not significantly more likely to report eating a healthy diet or exercising regularly may have a positive explanation—it is possible that, as a result of program education, the treatment group had higher standards as to what constitutes "healthy" or "regular." If that is true, their actual adherence may be better than the control group's, but the survey measures reported here may not reflect it. In addition, in many cases, behavioral change takes time; some changes do not occur until patients have experienced an adverse event that makes them recognize the value of adhering to advice from their physicians or care coordinators. Programs report that they expect it to take a few years to observe changes in patients' behavior and the effects of those behaviors on the patients' health and service use. The observed improvements in preventive care in some programs also may not result in lower hospitalizations or costs for a few years. Thus, there is reason to believe that some programs may have effects over the longer run.

The final evaluation report will assess the effectiveness of the demonstration programs by estimating program impacts on Medicare service use, expenditures, and quality of care over the first 4 years of program operations. The report will also describe the features of the program or target populations associated with effectiveness (if any). CMS has extended the end dates by 2 years to 2008, for the 11 demonstration programs that requested extensions. The four other programs will end in 2006 as originally planned. CMS granted the extensions because the Balanced Budget Act of 1997 authorizes CMS to continue any programs that are found to be cost-effective after the demonstration ends. The Act defines cost-effectiveness as either (1) reducing Medicare expenditures, or (2) not increasing Medicare expenditures while increasing the quality of services furnished and beneficiaries' and providers' satisfaction. The new end dates allow 11 of the demonstration programs to continue operating until the final evaluation findings are available. This extension allows any of the programs that the final evaluation report finds to be cost-effective to remain operating, rather than shutting down in 2006 and having to restart later.

I. INTRODUCTION

Chronic medical conditions contribute disproportionately to rising health care costs, morbidity, and mortality among Medicare beneficiaries. The definition of chronic illness is the subject of considerable debate, but it is generally accepted that a chronic disease (1) is persistent and incurable, although controllable with treatment; (2) if uncontrolled, leads to repeated acute health crises and hospitalizations and induces steady physical deterioration; and (3) requires substantial, sustained efforts by patients and providers to control (Brown et al. 2005). Disease management and case management interventions, which have been implemented widely in both the commercial sector and managed care plans, seek to provide better care for the millions of people with chronic conditions, thereby improving the health and quality of life of these patients and reducing their health care costs. A number of demonstrations testing the effectiveness of different types of disease management interventions for Medicare beneficiaries have been sponsored by the Centers for Medicare & Medicaid Services (CMS).

This congressionally mandated report describes the findings from the evaluation by Mathematica Policy Research, Inc. (MPR) of the 15 programs participating in the Medicare Coordinated Care Demonstration (MCCD); the evaluation was mandated by the Balanced Budget Act of 1997. The report covers the experiences of the programs during their first 2 years of operations and presents detailed descriptions of the programs' interventions; experiences with enrollment and disenrollment; and estimated effects on patients' quality of care, satisfaction, and Medicare service use and expenditures. The findings across the 15 programs are then synthesized to draw inferences about what program features appear to be associated with improved patient outcomes and lower Medicare expenditures.

A. COSTS OF CHRONIC CARE AND SHORTCOMINGS OF THE HEALTH SYSTEM

Chronic conditions can have severe adverse effects on the quality of the lives of both people who have the conditions and their caregivers, and, as many researchers have shown, the cost of treating those conditions is disproportionately high. Medicare beneficiaries with five or more chronic conditions accounted for two-thirds of total 1999 Medicare spending ($167 billion), roughly half of which was for inpatient hospital care (U.S. Department of Health and Human Services 2003; Anderson and Horvath 2002). Between 1990 and 1998, the age/sex-adjusted rate of hospital admissions among people aged 65 or older for 12 preventable conditions rose 15 percent (Kozak et al. 2001). Furthermore, from 1992 to 2000, there have been no reductions in the rates of preventable hospitalizations for congestive heart failure (CHF) (McCall et al. 2004); CHF is the leading cause of hospitalization among Medicare beneficiaries.

Nearly 80 percent of beneficiaries in the top quartile of 2001 Medicare spending were diagnosed as having at least one of seven chronic conditions: (1) asthma, (2) chronic obstructive pulmonary disease, (3) chronic renal failure, (4) CHF, (5) coronary artery disease (CAD), (6) diabetes, or (7) senility. Nearly one-half (48 percent) had more than one of those conditions (Congressional Budget Office 2005). Spending for the top quartile accounted for 88 percent of Medicare fee-for-service expenditures, and most of that spending was concentrated in the top 5 percent of beneficiaries, who accounted for 48 percent of total expenditures (Medicare Payment Advisory Commission 2004; Figure 2.1). Brown et al. (2005) found that nearly one-half of Medicare beneficiaries in 1997 were treated for one or more of eight, often chronic, conditions: (1) anemia, (2) CAD, (3) cancer, (4) diabetes, (5) heart disease, (6) liver/kidney problems, (7) pulmonary disease, or (8) stroke. In 1998, the average annual cost to Medicare for beneficiaries in this group was roughly $8,500, more than three times the average for beneficiaries without any of the conditions ($2,700). Costs for this group accounted for three-fourths of total Medicare costs for that year.

The current fee-for-service health care system does not support the substantial and sustained efforts required of patients and providers to manage chronic illnesses effectively. The system was originally developed to treat short-term acute conditions, and it currently focuses on the provision of services by individual providers, rather than on the coordination of services among providers. It does not pay for the time-consuming, ongoing education that many patients must receive if they are to adhere to their physicians' treatment recommendations. The system also does not pay for professionals to coordinate care across the numerous providers that chronically ill patients typically see, nor does it pay for providers to adopt electronic health records that might help to coordinate care, reduce duplicated tests, avoid prescription errors and interactions, and prompt providers when follow-up care is required (Moreno 2005).

The current Medicare fee-for-service system does not reward physicians and other providers for adhering to evidence-based guidelines for chronic care treatment (nor does it penalize them for failing to adhere). Physician reimbursement under the current system is based on the numbers and types of Medicare-covered procedures that are performed, rather than on adherence to practice guidelines, which is more difficult to measure. For example, one study examining care received by nearly 7,000 people in 12 metropolitan areas concluded that "Americans get substandard care for their ailments about half the time" and noted that care for diabetes was particularly poor (Kerr et al. 2004). Similarly, a study of elderly people with heart disease enrolled in two managed care plans found that these patients received appropriate care only 55 percent of the time, and that those with dementia or malnutrition received it only 31 percent of the time, because "providers overlook some common problems of old age" (Wenger et al. 2003). Providers, many of whom already are familiar with evidence-based treatment guidelines, would benefit from reminders to adhere to the guidelines for *all* their patients, mechanisms that encourage them to communicate with each other and that facilitate the sharing of test results on their patients (for example, between emergency room physicians and cardiologists), and systems

to measure care outcomes in order to identify and remedy care gaps (Lowenstein 2005). A former director of CMS's Office of Research Development and Information summarized several key barriers to better care for Medicare beneficiaries with chronic illnesses: poor data collection and tracking, lack of prescription drug coverage, decentralized program administration, difficulty communicating with beneficiaries, and difficulty integrating physicians into the improvement process (Guterman 2004).

B. INTERVENTIONS TO IMPROVE CHRONIC CARE AND THEIR EFFECTIVENESS

A wide variety of "disease management" or "case management" programs have been found to improve the delivery of chronic care (Villagra and Ahmed 2004; Bodenheimer et al. 2002; Chen et al. 2000). However, many others have been shown to have no impacts, leading to considerable debate over the effectiveness of such programs in general (Congressional Budget Office 2004). Many of these programs are operated by providers of disease management services or by hospitals or other entities that serve people with chronic illnesses. Most of the programs that have been assessed with any methodological rigor have served a small number of patients. The successful programs have been shown to have a number of features in common. Many rely on two methods to improve patients' health and reduce the likelihood of hospitalization: (1) patient education on treatment regimens and self-care recommendations, and on the importance of adhering to them; and (2) telephone or in-person monitoring of patients' symptoms, adherence, and self-care between physician office visits (see, for example, Riegel et al. 2002; Rich et al. 1995; Wasson et al. 1992). Some programs have shown that encouraging physicians to use evidence-based practices and feeding back to them patient information obtained during monitoring calls or home visits have both reduced medical costs and improved care delivery (see, for example, Sidorov et al. 2002; West et al. 1997). Many of the most

successful programs also develop mechanisms to improve communication across providers, such as team meetings, telephone updates by case managers, and the sharing of medical records, thereby reducing the fragmentation of care and the amount of conflicting advice given to patients (Chen et al. 2000). Finally, successful programs sometimes help patients to follow treatment regimens by guiding them to (or providing) support services and care-related goods that the patients may not have realized were available, such as pharmacy assistance, subsidized transportation, home-delivered meals, and pill cassettes and other medication scheduling aids (Brown et al. 2001). More specifically, research during the past decade suggests (but by no means shows conclusively) that successful care coordination programs typically share some broad features: effective patient identification; a well-designed, evidence-based, structured intervention; highly qualified staff; physician buy-in; and financial incentives aligned with program goals (Exhibit I.1) (Chen et al. 2000).

Another approach to improving the care of chronic conditions is to base patient support in physicians' offices. The underlying rationale is that, for most patients, the physician is the primary link to the current health care system. Like disease management, the "chronic care model," as it is known, aims to improve patient self-management by communicating with patients in their homes; unlike disease management, however, it also calls for physicians to reorganize their office practices to focus more on chronic, rather than acute, illnesses. This focus is achieved by providing patients who have chronic conditions with self-management support (such as patient education); employing multidisciplinary teams that include nonphysician staff; providing case management for the most seriously ill or frailest patients; and developing clinical information support systems that remind physicians to use evidence-based practice guidelines, and that can provide the teams with feedback about their performance (Casalino 2005). A review of 39 chronic care model programs for patients with diabetes showed that 32 improved at

EXHIBIT I.1

COMMON FEATURES OF MOST SUCCESSFUL
CARE COORDINATION PROGRAMS

Targeting High-Risk People. People with recognized high-cost diagnoses, such as heart failure, but also those with prevalent geriatric syndromes, such as physical inactivity, falls, depression, incontinence, misuse of medications, and undernutrition (Fox 2000; Rector and Venus 1999)

Having a Comprehensive, Structured Intervention Adaptable to Individual Patients

- Multifaceted assessment whose end product is a written care plan that can be used to monitor a patient's progress toward specific long- and short-term goals, and that is updated and revised as the patient's condition changes (Chen et al. 2000)
- Process for providing aggregate- and patient-level feedback to care coordinators, program leaders, and physicians about patient outcomes to enable the program to modify or intensify the intervention if it appears that it is not having the expected effect on intermediate or ultimate outcome indicators (Chen et al. 2000)
- Patient education combining provision of factual information with techniques to help patients to change self-care behaviors and better manage their care, as well as addressing affective issues related to chronic illness, such as depression (Aubry 2000; Williams 1999; Lorig et al. 1999; Vernarec 1999; Roter et al. 1998)
- Structures and procedures for integrating fragmented care and for facilitating communication among providers, to address the complexities posed by patients with several comorbid conditions and, when necessary, to arrange for community services (Chen et al. 2000; Hagland 2000; Bodenheimer 1999)

Having Highly Trained Staff and Actively Involved Providers. Disease managers who are at least baccalaureate-prepared nurses or who have case management or community nursing experience. Active support and involvement of patients' physicians to encourage the patients' cooperation with disease managers, and to respond to disease managers' requests when urgent patient problems arise (Chen et al. 2000; Schore et al. 1999)

Using Financial Incentives. To compel programs to look for creative ways to meet patients' goals, and to reduce total health care costs by reducing preventable hospital stays (Schore et al. 1999)

---

least one patient process or outcome measure (Bodenheimer et al. 2002). A review of 27 chronic care model programs for patients with asthma, diabetes, or heart failure showed that 18 reduced the patients' health service use and costs (Bodenheimer et al. 2002). Despite these positive outcomes, the chronic care model has the disadvantage of requiring that physician practices have the financial resources to make the necessary organizational changes, even though the current health care financing system does not explicitly provide those resources.

Although managed care plans have embraced disease management as a means of improving care for enrollees with chronic illnesses, it is unclear whether disease management and care coordination programs can improve health outcomes for and reduce the Medicare costs of their chronically ill beneficiaries. Specifically, disease management might fail to work if interventions aimed at altering patients' or physicians' behaviors either do not change the behaviors or, if they do change them, do so in ways that do not lead to changes in service use. Even if the programs do work, it still is necessary to understand how best to implement disease management or care coordination programs in a fee-for-service setting. Additional research also is necessary to assess the relative importance of specific disease management features (such as patient identification, patient engagement, the use of multiple clinical guidelines, and the integration of disease management with physician practice), as well as to identify best practices for those features (Villagra and Ahmed 2004).

C. MEDICARE DEMONSTRATIONS AND INITIATIVES TO IMPROVE OUTCOMES FOR BENEFICIARIES WITH CHRONIC ILLNESSES

CMS is currently funding or is planning to fund a series of important demonstrations, evaluations, and studies to expand the evidence base on whether and how disease management, case management, and other care coordination interventions can improve care for Medicare beneficiaries with chronic illness in its fee-for-service program. Exhibit I.2 summarizes the features of those demonstrations.

EXHIBIT I.2

CURRENT AND PLANNED DEMONSTRATIONS AND EVALUATIONS TO IMPROVE CARE FOR BENEFICIARIES WITH CHRONIC ILLNESSES

Program (Operational Period)	Demonstration	Identification of Potential Enrollees	Eligible Beneficiaries Volunteer (Opt-In Model)	Number of Beneficiaries Participating (Number of Sites)	Quality and Satisfaction Performance Criteria	Program Carries Financial Risk	Randomly Assigned Treatment Group?
Medicare Coordinated Care Demonstration (2002–2008)	Yes	Provider and self-referral	Yes	20,088 as of 9/24/05 (15)	No	No	Yes
Informatics for Diabetes Education and Telemedicine (2000–2008)	Yes	Provider and self-referral	Yes	2,164 as of 9/30/05 (2)	No	No	Yes
Medicare Disease Management Demonstration (2004–2008)	Yes	CMS provides names; provider and self-referral	Yes	25,921 as of 9/24/05 (3)	Yes[a]	Yes[b]	Yes
LifeMasters Demonstration Program (2005–2008)	Yes	CMS provides names of dual eligibles	No	30,375 as of August 2005 (1)	Yes[a]	Yes[c]	Yes
Medicare Health Support Program (formerly, the Chronic Care Improvement Project) (2004–2009)	No	CMS provides names; all in geographic area who meet criteria are covered	No	180,000 (expected) (9)	Yes[d]	Yes[c]	Yes
Medicare Physician Group Practice Demonstration (2005–2008)	Yes	CMS provides names of eligible beneficiaries per PGP	No	5,000 physicians; more than 200,000 Medicare fee-for-service beneficiaries (10)	Yes[e]	No	No
Medicare Lifestyle Modification Program Demonstration (1999–2006)	Yes	Provider and self-referral	Yes	Up to 3,600 (18)	No	No	No
Medicare Care-Management for High-Cost Beneficiaries (2005–2008)	Yes	CMS provides names based on claims data	Yes	At least 25,000 (expected) (6)	Yes[f]	Yes[g]	Yes
End-Stage Renal Disease Management (2005–2009)	Yes	Provider and self-referral	Yes	TBD	Yes[h]	Yes	No

EXHIBIT I.2 *(continued)*

Program (Operational Period)	Demonstration	Identification of Potential Enrollees	Eligible Beneficiaries Volunteer (Opt-In Model)	Number of Beneficiaries Participating (Number of Sites)	Quality and Satisfaction Performance Criteria	Program Carries Financial Risk	Randomly Assigned Treatment Group?
Medicare Care Management Performance Demonstration (forthcoming, 3 years)	Yes	CMS provides names of eligible beneficiaries per practice	No	800 practices; at least 140,000 fee-for-service beneficiaries (expected) (4 states)	Yes[i]	No	No

[a]Programs participating in the Medicare Disease Management and LifeMasters demonstrations are required to collect patient-level quality indicators, but their payment from CMS does not depend on quality outcomes.

[b]Programs participating in the Medicare Disease Management Demonstration must guarantee to CMS that they will reduce overall Medicare spending; if they do not produce these savings, they must fully reimburse the Medicare program for any losses.

[c]LifeMasters and organizations participating in the Medicare Health Support Program are required to reduce overall Medicare spending; if they do not produce these savings, they must return all or part of their fees.

[d]Organizations participating in the Medicare Health Support Program are required to collect patient-level quality and satisfaction indicators and are paid partially based on quality performance and patient satisfaction.

[e]The demonstration will use 32 measures that focus on common chronic illnesses and preventive services.

[f]The performance standards are clinical quality, beneficiary and provider satisfaction, and savings guarantee. Each organizational agreement will specify standards.

[g]Participating organizations face financial risk in the event of failure to meet agreed-on performance guarantees for clinical quality, beneficiary and provider satisfaction, and savings targets that will include all fees and gain-sharing payments.

[h]Determinations of incentives will be based on the end-stage renal disease clinical performance measures.

[i]The demonstration will use 26 measures that focus on common chronic illnesses and preventive services.

CMS = Centers for Medicare & Medicaid Services; PGP = physician group practice; TBD = to be decided.

1. Precursors to the MCCD

The MCCD was preceded by the random-assignment-based Medicare Case Management Demonstrations for high-cost, fee-for-service beneficiaries, which included three demonstration programs operating between 1993 and 1995. Although none of the programs produced the intended improvements in beneficiary self-care, health status, or cost-savings, several important lessons emerged contrasting the features of the demonstration programs with those of a successful program operating at about the same time (Rich et al. 1995). The lessons included the importance to successful disease management of (1) having active physician involvement; (2) using well-defined, goal-oriented interventions; (3) providing ongoing feedback to program staff and physicians on progress toward the goals; and (4) providing financial incentives to programs and providers to reduce the need for expensive hospital stays and thereby generate Medicare cost savings (Schore et al. 1999).

2. The MCCD

The design for the MCCD, the subject of this report, was based on CMS's assessment of best practices in coordinated care (Chen et al. 2000). That evaluation concluded that there is no optimal approach to care coordination. Successful programs varied widely in the types of interventions used and in the structural characteristics of the organizations implementing the programs. However, most of the successful programs shared several features, namely (1) a focus on well-developed care planning and patient education, (2) strong patient-case manager relationships, (3) a proactive emphasis on preventing health problems, (4) use of evidence-based intervention guidelines, and (5) having experienced nurses serve as care coordinators.

In July 2000, CMS issued a Request for Proposals soliciting organizations to participate in the MCCD project, a demonstration mandated by the Balanced Budget Act of 1997. Applicants were expected to have experience operating a disease management or case management

program, and to present some evidence that they had been able to reduce hospitalizations or costs. CMS took this approach to maximize the potential for showing, in a time-limited demonstration, that a successful care coordination program could be adapted effectively to a Medicare fee-for-service environment and population. Of the 58 proposals submitted, 15 were selected as demonstration sites.

The 15 demonstration programs all serve chronically ill Medicare beneficiaries, but they target different diseases and have developed widely differing interventions. The demonstration allowed the programs to design their own interventions (which typically included patient education to improve adherence to treatment recommendations), and to define appropriate target populations (which included beneficiaries with diabetes, heart failure, and other types of heart disease, as well as other conditions associated with morbidity or frailty in elderly people).

The MCCD evaluation has three goals. The evaluation is intended to (1) provide CMS with unbiased estimates of the ability of the 15 demonstration programs to provide better and more cost-effective care for chronically ill Medicare beneficiaries; (2) assess the extent to which the effectiveness of care coordination depends on patient and program characteristics; and (3) provide guidance on the feasibility, desirability, and possible structure of a Medicare coordinated care benefit.

The demonstration is now in its third year. The MCCD programs originally were authorized to operate for 4 full years, and to enroll new patients through the 42nd month. However, the end date has been extended to 2008 for 11 programs that requested an extension. The extension was granted because the legislation requiring the demonstration authorizes the continuation of any programs that are shown to lower net costs to Medicare or to improve quality and beneficiary and provider satisfaction without increasing net costs to Medicare. The new end dates allow the 11 demonstration programs to continue operating until the evaluation findings are available.

This change allows any of the programs that are found to be effective to remain operating rather than shutting down in 2006 and then having to restart later.

During the demonstration, programs are paid a capitated rate per month for each patient who is enrolled in the treatment group until the patient dies or disenrolls. The rates in the first year varied from $50 to $437 across the 15 programs.[1] In return for the capitation payment, programs must provide the intervention that was described in their operational protocols approved by and established with CMS, but they are not required to guarantee costs savings for the Medicare program.

The programs started enrolling patients between April and September of 2002, after receiving approval from the Office of Management and Budget. Six programs started enrolling in April, five started in June, one did so in July, two began in August, and one began in September. In each program, Medicare beneficiaries who expressed an interest in participating in the demonstration and who met the program's eligibility criteria were randomly assigned (by MPR) to either the treatment group, which received the intervention as well as their normal Medicare benefits, or to the control group, which received only their normal Medicare benefits. The evaluation estimates the effect of the demonstration programs by comparing outcomes for the treatment and control groups. The impact analyses test whether the programs (1) reduce Medicare payments and service use, (2) improve the quality of care, and (3) improve patients' and physicians' satisfaction. The analysis of Medicare payments estimates program impacts on costs to the Medicare program (including care coordination program costs) and program impacts on Medicare service use. The analysis of the quality of care assesses the care delivery process

[1] Five programs have multiple rates. The rate that is applicable for a particular patient depends on the patient's diagnosis, acuity level, or length of time in the program.

and the clinical outcomes of Medicare beneficiaries. In the satisfaction analysis, both patients' satisfaction and physicians' satisfaction are covered.

3. Other CMS Disease Management Initiatives

A number of other CMS demonstrations also will provide estimates of program effectiveness. The Medicare Disease Management Demonstration, which began in 2004, was authorized by Section 121 of the Benefits Improvement and Protection Act of 2000. That demonstration requires that its three demonstration programs target beneficiaries with advanced-stage diabetes, CAD, or heart failure. The demonstration intervention includes both disease management (as defined by each program) and a pharmacy benefit that covers nearly all prescription drugs, and that charges modest patient copayments. (Its design pre-dates legislation for the Medicare prescription drug benefit that will begin in 2006.) In addition, the three participating demonstration programs must guarantee savings for Medicare (although the amount of savings is not specified, so this requirement is essentially that a program be cost neutral). The demonstration is being evaluated using a random assignment design.

CMS also is testing larger-scale "population-based" approaches to delivering disease management, on the grounds that these approaches would be more operationally efficient: CMS could hold a single entity accountable for improving health and for reducing costs for all Medicare beneficiaries who reside in a particular geographic area and who have a particular illness. Furthermore, holding a single disease management provider accountable for *all* beneficiaries in a defined population would eliminate concerns about favorable selection (that is, that a program would enroll patients who would be most likely to adhere to recommendations

and, therefore, most likely to have favorable outcomes).[2] With a population-based approach, however, the task for the provider changes from that of identifying beneficiaries who want the services to that of encouraging all patients with the target condition (and their physicians) to participate, including patients who initially decline or who are difficult to contact (for example, because the program is unable to obtain their telephone numbers). Two initiatives, the LifeMasters Demonstration Program and the Medicare Health Support Program (formerly known as the Chronic Care Improvement Program), are currently testing a population-based approach.

The LifeMasters program targets beneficiaries who are dually eligible for Medicare and Medicaid; live in Florida; and have severe CAD, CHF, or diabetes. CMS's evaluation contractor identifies all beneficiaries eligible for LifeMasters based on program-provided criteria and randomly assigns them to either the treatment group (which has the opportunity to receive the program's disease management intervention) or the control group (which does not). The program is then responsible for engaging and providing disease management services to all eligible beneficiaries in the treatment group in exchange for a monthly payment for each of those beneficiaries. The program is required to be cost neutral and must share any net reductions in Medicare costs with CMS.

The Medicare Health Support Program, which is expected to operate in nine sites located in nine states and the District of Columbia, targets beneficiaries with CHF or diabetes. Two key features distinguish this pilot program from previous demonstrations. First, its scale is much larger. The selected sites must operate in regions that, together, encompass 10 percent of all

[2] If a program is paid based on savings generated, its incentive is to enroll patients in whom it is most likely to be able to induce the desired behavior changes (sometimes referred to as "low-hanging fruit"). Patients' self-selection behavior is likely to exacerbate these tendencies, as the patients most likely to enroll voluntarily are those who are most open to accepting the program's advice. A population-based approach forces programs to engage all patients in the target group, and to seek ways to improve their outcomes.

beneficiaries in the fee-for-service Medicare program; each site is expected to identify at least 30,000 eligible beneficiaries. Second, the sites will have clear performance targets for quality, cost savings, and patient satisfaction, and they will be at risk for all fees if they fail to meet their targets. Sites in previous demonstrations (the Medicare Disease Management Demonstration and LifeMasters Demonstration Program) are at risk only for fees.

D. PURPOSE OF AND METHODOLOGY FOR THIS REPORT

This report is a synthesis of findings covering the first 25 months of the MCCD programs, about halfway through the demonstration. The goals of the evaluation are to estimate the impacts of each of the 15 programs, and to assess which program features appear to be associated with program success. The first report to Congress provided a preliminary synthesis of findings from the first year of the demonstration programs' operations (Brown et al. 2004). It described the programs and beneficiaries participating in the demonstrations, the interventions that the programs implemented, and early feedback from physicians and patients about the programs. At that time, it was too early to produce estimates of program effects on enrolled patients' service use or costs, as an insufficient number of observations were available, and data were available for only the first 6 months of program operations. This report presents treatment-control differences for outcomes measured over the first year after the month of enrollment for beneficiaries who enrolled during the programs' first year; it also presents estimates of treatment-control differences, by calendar month, over the first 25 months of operations. The report also describes the methods that the programs used to recruit beneficiaries, the characteristics of the first-year program enrollees, the nature and focus of the programs' interventions, and the relative intensity and quality of the interventions. The findings are summarized in the study's Second Report to Congress (Brown et al. 2006).

Data are drawn from several sources. Program descriptions from the evaluation's implementation analysis are based primarily on information from site visits, telephone interviews, review of program documents, and analysis of data that the programs provided specifically for the evaluation. The participation analysis and treatment-control differences presented in this report are based on Medicare enrollment and claims data available through June 2005. Additional measures of the processes of care, quality of care, patient symptoms and well-being, and satisfaction with care are based on surveys of patients and physicians conducted by MPR by telephone. A final report to Congress, due in 2008, will present findings on program impacts on quality, use, and cost of care provided over the full 4-year demonstration period, based on claims data.

E. THE REST OF THIS REPORT

Chapter II compares the key features of the 15 demonstration programs as planned and as implemented and scores each of the 15 in terms of how well designed it is on 10 different dimensions. Chapter III describes the demonstration enrollees and analyzes participation and disenrollment patterns. Chapter IV presents patients' and physicians' perceptions about the programs. Chapter V describes treatment-control differences in the processes and quality of care, as well as in health status, and Chapter VI estimates the treatment-control differences in Medicare service use and expenditures. Chapter VII synthesizes the lessons from the study to date and describes the content and timing of the third evaluation report.

II. DEMONSTRATION PROGRAM STRUCTURE AND KEY FEATURES

Nearly all of the program hosts for the Medicare Coordinated Care Demonstration (MCCD) were experienced providers of disease management or case management interventions that had some evidence of having reduced hospitalizations or costs.[3] For the demonstration, each host was free to select a target patient population, and, based on its previous experience, to design interventions to improve patients' health and reduce costs in the fee-for-service Medicare program. Although the 15 programs shared some features, such as conducting initial patient assessments and regularly monitoring patients' symptoms, they differed in many important ways—who they targeted, how they provided patient education, and the extent to which they involved physicians in their programs, to name just a few.

This chapter describes the organizations that hosted the 15 demonstration programs. It then summarizes key program intervention features and presents data that each program collected for the demonstration describing care coordinator contacts with patients randomly assigned to the demonstration treatment group (referred to in this chapter simply as "patients"). It concludes with a discussion of an approach for assigning quantitative "scores" or ratings for the features of each program's intervention.

A. PROGRAM STRUCTURE, EXPERIENCE, AND RELATIONSHIP WITH PHYSICIANS

The MCCD programs were hosted by a diverse range of organizations that differed both in their level of experience delivering the specific interventions offered to demonstration program

[3] This evidence generally consisted of pre- and postintervention studies, studies comparing enrolled patients with historical controls, or studies comparing enrolled patients with groups of roughly comparable patients. Patients in the studies were most often insured by commercial or Medicare managed care health plans.

target populations and in their existing links to the community of physicians who would be serving their patients. Previous experience is likely to facilitate the ability of a program to generate favorable effects quickly; pre-existing links may affect the extent to which a program enjoys critical support from patients' physicians.

1. Program Host Organizations

Program hosts included a wide variety of provider and organizational types. The program hosts consisted of five commercial disease management providers, three hospitals, three academic medical centers, one integrated delivery system, a hospice, a retirement community, and a long-term care facility (Exhibit II.1).

2. Program Service Areas and Target Populations

The programs served patients in 16 states and in the District of Columbia, from Maine to Arizona, and from northern California to south Florida (Figure II.1). Four programs served patients primarily in rural areas; the other 11 served patients in cities and suburbs.

Six of the 15 programs targeted only a single condition. Four of the six targeted congestive heart failure (CHF), one targeted coronary artery disease (CAD), and one targeted cancer (Exhibit II.1). Another program targeted CAD and CHF. Each of the eight remaining programs targeted several diagnoses. Three of the eight cast particularly wide nets by targeting many diagnoses or by targeting beneficiaries who were frail or otherwise considered to be at high risk for hospitalization.

Most of the programs excluded certain types of beneficiaries, including people with medical conditions that were unrelated to the programs' target diagnoses and the treatment of which might dominate the treatment for the target diagnoses (such as end-stage renal disease), those who had terminal illnesses, and those who had conditions or limitations that would make it

EXHIBIT II.1

PROGRAM HOSTS, TARGET DIAGNOSES, AND SERVICE USE REQUIREMENTS

	Avera	Carle	CenVaNet	Charles-town	CorSolutions	George-town	Health Quality Partners	Hospice of the Valley	Jewish Home & Hospital	Medical Care Development	Mercy	QMed	Quality Oncology	University of Maryland	Washington University	Number of Programs
Host Organization Type																
DM/CC Provider			X		X		X					X	X			5
Hospital	X									X	X					3
Academic						X								X	X	3
Other[a]		X		X				X	X							4
Targeted Diagnoses																
CHF	X	X	X	X	X	X	X	X	X	X	X			X		12
Other Heart Diseases[b]		X	X	X			X	X		X		X				7
COPD/ Asthma		X	X	X			X	X	X		X					7
Diabetes		X	X	X			X		X							5
Cancer								X	X				X			3
Stroke								X	X		X					3
Dementia								X	X							2
Renal Failure											X					1
Other[c]			X				X	X	X		X				X	6
Service Use Requirements[d]																
Hospital, Target Diagnosis	X				X	X				X	X			X		6
Hospital, Any Diagnosis		X		X				X	X							4
Other Service Use		X	X						X							3
None							X					X	X		X	4

[a]Other organizations include an integrated delivery system (Carle), a retirement community (Charlestown), a hospice (Hospice of the Valley), and a long-term care facility (Jewish Home and Hospital).

[b]Other heart diseases include coronary artery disease, atrial fibrillation, ischemic heart disease, and hypertensive heart disease.

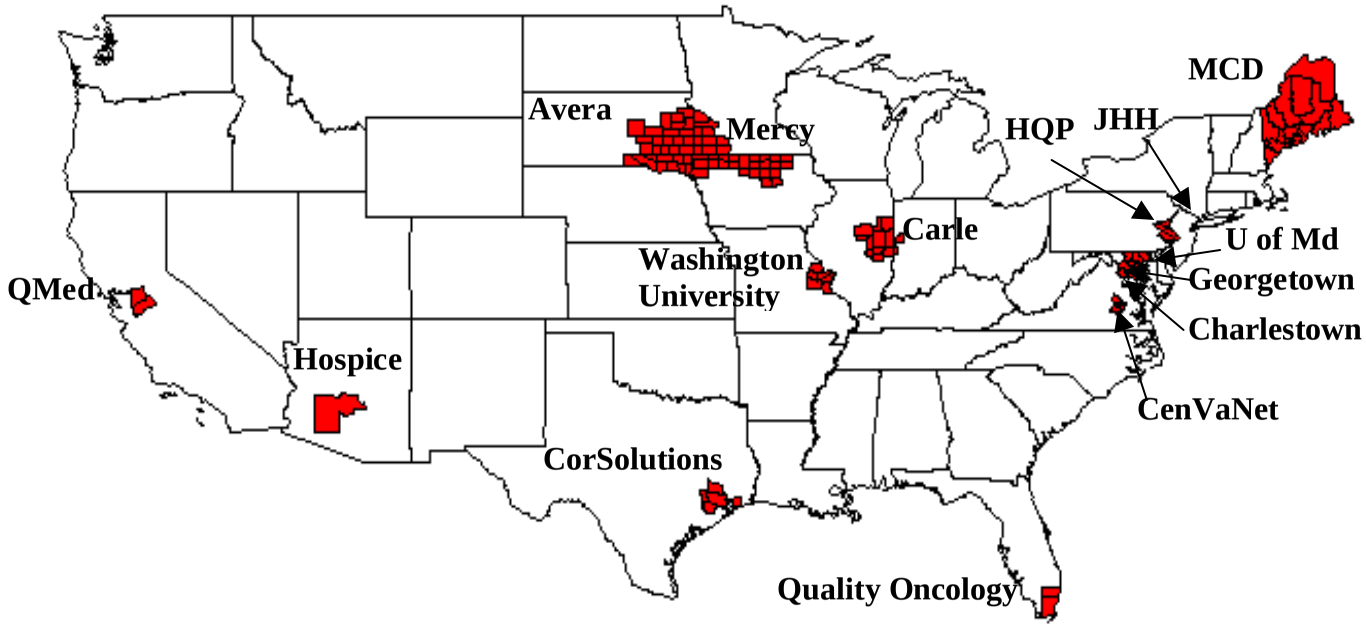
[c]Other diagnoses include cerebrovascular disease (CenVaNet); hypertension and hyperlipidemia (Health Quality Partners); Parkinson's disease and amyotrophic lateral sclerosis (Hospice of the Valley); liver disease, other cerebrovascular disease, psychotic disorders, and major depression (Jewish Home and Hospital); and other cerebrovascular disease and liver disease (Mercy). To identify patients, Washington University used a proprietary algorithm developed by its demonstration partner, StatusOne, to target Medicare beneficiaries who were likely to become clinically unstable, and to require hospitalization during the next 12 months.

[d]Service use is required during the year before enrollment with the following exceptions: hospitalization could be during the 2 years before enrollment (Charlestown), and hospitalization must be within the last 60 days (Medical Care Development). See Appendix Exhibit A.1 for additional detail on service use requirements.

CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; DM/CC = disease management or care coordination.

FIGURE II.1

THE MCCD SERVICE AREAS



Hospice = Hospice of the Valley; HQP = Health Quality Partners; JHH = Jewish Home and Hospital Lifecare System; MCD = Medical Care Development; U of Md = University of Maryland.

difficult for them to benefit from a program's intervention (such as dementia or a serious mental disorder). (See Appendix Exhibits A.1 and A.2 for detailed descriptions of the programs' eligibility criteria.)

3. Host Organizations' Experience with Care Coordination

The Centers for Medicare & Medicaid Services (CMS) selected demonstration hosts that were experienced providers of care coordination, disease management, or related services to maximize the chances that the hosts would be able to become operational quickly, and that they would succeed at improving patient health. Most hosts (12 of 15) had provided disease management or care coordination in programs that served as prototypes to their demonstration programs (Exhibit II.2). The remaining three had provided cardiac rehabilitation or web-based telemedicine or had provided consultation services to a disease management program.

Most of the hosts had experience primarily with patients in managed care. Only 4 of the 15 prototype programs had served patients in the fee-for-service sector. Managed care typically affords providers more influence with patients' physicians and access to medical records than is usually available in the fee-for-service sector. The fragmented care typically found in the fee-for-service sector can present difficulties for programs whose staff expect to obtain patients' clinical data from medical records, or to easily gain the support of physicians.

4. Host Organizations' Relationship to Patients' Physicians

The literature has found that physician support is critical to care coordination efforts in terms of both encouraging patients' initial participation and validating the advice that care coordinators provide to patients (Chen et al. 2000). Physician support also is key to facilitating communications between care coordinators and physicians, to establishing the credibility of care coordinators in the eyes of the physicians, and to fostering physicians' trust in the care

EXHIBIT II.2

DEMONSTRATION PROGRAM PROTOTYPES AND PHYSICIAN LINKS

	Program Prototypes	Physicians' Relationship with Host and Previous Experience with Program Staff
Avera	Certified cardiac rehabilitation and short-term health management for working-age managed care population with cardiovascular disease and diabetes	Some physicians employed by host Some physicians not employed by host Have also worked with staff
Carle	Geriatric Team Care program for high-risk managed care population	Most physicians employed by host Program administrators worked with physicians
CenVaNet	Care management program for Medicare managed care members with CHF, COPD, or diabetes	Physicians are part of host's physician network Program administrators worked with physicians
Charlestown	Care management program for Medicare managed care members	All physicians employed by host Program administrators and care coordinators worked with physicians
CorSolutions	Commercial disease management product provided to Medicare managed care members with heart failure	No ties No experience
Georgetown	Web-based telemedicine program for patients with diabetes	Some physicians employed by host No experience
Health Quality Partners	Commercial care management product for Medicare managed care members	No ties Program administrators worked with physicians
Hospice of the Valley	PhoenixCare care management for managed care members with terminal illness	No ties Program administrators worked with physicians
Jewish Home and Hospital	Geriatric Outreach program for individuals aged 80 or older with chronic illness	Two hospital-based physician practices collaborating with host Program administrators worked with one of the practices
Medical Care Development	MECare cardiac disease management	Most physicians employed by participating hospitals Care coordinators are hospital-based nurses
Mercy	Outpatient hospital case management program	Most physicians employed by host Program staff worked with physicians
QMed	Commercial disease management product provided to Medicare managed care members with CAD	No ties Many physicians worked with program staff
Quality Oncology	Commercial disease management product provided to Medicare managed care members with cancer	No ties Many physicians worked with program staff

EXHIBIT II.2 *(continued)*

	Program Prototypes	Physicians' Relationship with Host and Previous Experience with Program Staff
University of Maryland	Consultants to disease management program for managed care members with heart failure operated by home health agency	No ties No experience
Washington University	Care management for high-risk Medicare managed care members developed jointly with StatusOne, a commercial disease management provider (and the university's partner for this demonstration)	All physicians employed by host Program administrators and care coordinators worked with physicians

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

coordinators. Credibility and trust enable physicians to feel comfortable about sharing important patient information with care coordinators, asking care coordinators to intervene with patients when necessary, and responding to issues that care coordinators have raised.

As discussed in detail in this chapter, programs can take a number of steps to build relationships between physicians and care coordinators. However, they may have a head start if program staff and patient physicians already have established organizational links, such as a shared employer (for example, if the host is a medical center), or if the physicians are familiar with the program's administrative staff or care coordinators. Both circumstances increase the likelihood that physicians and program staff share a common vision of patient care, and they may give the programs some leverage over physicians' behavior (for example, in encouraging physicians to refer patients or to cooperate with care coordinators).

For most of the programs (9 of 15), at least some physicians were employed by either the program host or organizations affiliated with the host for the demonstration (Exhibit II.2). Physicians of patients in 12 of the 15 programs had worked previously with program leadership or care coordinators. In only two programs were physicians neither affiliated with program hosts nor familiar with staff.

B. PROGRAM INTERVENTIONS

The demonstrations are not a test of a single intervention in 15 sites, but rather, a test of 15 different interventions. CMS decided on this design because it found that successful programs shared some common features, but did not necessarily follow a common approach (Chen et al. 2000). Care coordination is predicated on the belief that the failure of patients and physicians to properly manage chronic illnesses results in uncontrolled symptoms and acute exacerbations that could have been avoided, but instead, lead to expensive treatment. All of the MCCD programs shared the broad goal of improving patient health as a means of reducing the

need for emergency room (ER), inpatient hospital, and other acute care services. Each program started with a comprehensive patient assessment for each patient, followed by a plan for addressing the patient's knowledge and care gaps and a process for ensuring that gaps would be filled.

The program interventions collectively took four basic approaches to improving patient health: (1) improving adherence to treatment recommendations, usually through patient education; (2) improving communication and coordination, including identifying worsening symptoms before they required hospital care; (3) improving physician practice; and (4) increasing access to support services. The programs varied, however, as to whether they focused on a particular approach or approaches, and, if so, the extent of their focus.

The programs also differed in the mode and intensity of their contacts with patients who enrolled during their first year of operations. Most programs (11 of 15) contacted patients once or twice per month during the patients' first year after random assignment, with 6 of the 11 averaging 1.2 to 1.5 contacts per month and the other 5 averaging 2.2 to 2.6 contacts per month (Table II.1). Three programs contacted their patients substantially more frequently, between four and eight times per month. Each of these three programs used home telemonitors (as discussed in section 3 below), and some portion of those contacts were likely for the purpose of inquiring about out-of-range monitor readings. It appears that the remaining program contacted patients only about once every 2 months. (Staff from the program speculated that care coordinators were not recording all their patient contacts.) Care coordinators (rather than patients) initiated most contacts (between 73 and 98 percent) during the patients' first year, and most contacts were by telephone for most programs. Seven programs also provided over one-fourth of their contacts in person; the program with the highest proportion provided nearly 70 percent of its contacts in person (56 percent in patients' homes and 13 percent in other locations, such as clinics and physicians' offices).

TABLE II.1

CARE COORDINATORS' CONTACTS FOR ANY PURPOSE DURING THE YEAR AFTER RANDOM ASSIGNMENT, AMONG 1ST-YEAR ENROLLEES

	Mean Number of Contacts per Month	Percentage of Patients with 4 or More Contacts per Month	Percentage of Contacts Initiated by Care Coordinators	Percentage of Contacts that Were In-Person and in Patients' Homes	Percentage of Contacts that Were In-Person and Elsewhere[a]	Number of Patients
Avera	8.2	89.3	97.5	1.4	0.2	157
Carle	1.4	2.2	85.3	6.6	24.8	1,151
CenVaNet	1.4	6.2	86.8	17.6	0.5	538
Charlestown	2.3	8.6	89.8	17.1	14.8	212
CorSolutions	2.6	9.5	88.9	3.7	0.0	366
Georgetown	5.9	78.0	90.8	13.8	0.3	53
Health Quality Partners	2.2	3.3	95.1	10.6	31.0	243
Hospice of the Valley	2.5	13.4	93.3	32.2	4.9	236
Jewish Home and Hospital	2.5	18.2	74.9	15.2	25.0	271
Medical Care Development	1.5	2.6	95.5	1.5	27.9	196
Mercy	1.4	1.6	85.0	56.4	12.8	317
QMed	1.2	0.0	93.2	0.0	7.6	698
Quality Oncology	0.6[b]	0.0	95.3	0.0	0.0	31
University of Maryland	3.9	17.2	91.4	0.0	6.5	29
Washington University	1.2	0.7	72.8	2.7	2.0	715

Source: Data describing program contacts with treatment group patients were prepared by programs and submitted quarterly to Mathematica Policy Research, Inc.

[a]Some programs instructed care coordinators to meet with patients in clinics or in the offices of the patients' physicians.

[b]Staff from the program speculated that care coordinators were not recording all their patient contacts.

1. Care Coordinators' Qualifications, Training, and Caseloads

All but two of the programs required their care coordinators to be experienced registered nurses (RNs) (Exhibit II.3). Of the two programs, one also used social workers as care coordinators, and the other used experienced licensed practical nurses as well as RNs. Three of the programs that required care coordinators to be RNs also required them to be at least baccalaureate-prepared. Almost all of the programs required previous nursing experience in specific areas, most commonly cardiac care, community nursing (such as home health, public health, or hospice nursing), geriatric care, or medical-surgical nursing.

The intensity and types of training that the programs provided to incoming care coordinators varied substantially. All but two offered formal (classroom) orientation that varied in length from a couple of days to 4 weeks. One of the two programs without formal orientation (the smallest program) employed only one care coordinator; because this care coordinator was a nurse practitioner with 30 years of cardiac care experience who had helped to design the program, she did not need orientation. The other program that did not provide formal orientation was small and trained its care coordinators solely by offering 6 to 8 months of mentoring. Seven of the 13 programs with formal orientation followed up the orientation with a period of mentoring by either the care coordination supervisor or a more experienced care coordinator.

Average caseloads also varied widely. Care coordinators at 4 programs had caseloads of 50 or fewer patients; the smallest average caseload was 36 patients. (Two of the four programs intentionally had small caseloads, whereas the other two were smaller than planned because they enrolled so few patients.) Seven programs had average caseloads of 60 to 90 patients per care coordinator, and 3 had caseloads of more than 90; the largest had a caseload of 200 patients per care coordinator. The 15th program relied on local care coordinators who contacted patients primarily by telephone, but who could meet face-to-face with patients if necessary, as well as on

EXHIBIT II.3

CARE COORDINATORS' QUALIFICATIONS, TRAINING, AND CASELOADS

	Education	Types of Nursing Experience Required[a]	Initial Program Training	Average Caseload[b]
Avera	RN; baccalaureate (preferred)	Cardiac, geriatric	Orientation by supervisor	1:88
Carle	RN	Community, medical-surgical	Three-week orientation; directed observation by supervisor	1:135
CenVaNet	RN; baccalaureate preferred	Case management, managed care	Two-week orientation; directed observation by supervisor	1:70
Charlestown	RN; baccalaureate preferred	Community, medical-surgical	Orientation by supervisor; worked with experienced mentor	1:60
CorSolutions	RN	Cardiac, critical care	Three-week orientation	1:145
Georgetown	RN; baccalaureate (required)	Cardiac, community, geriatric, medical-surgical	Worked with experienced mentor for 6 to 8 months	1:36
Health Quality Partners	RN; baccalaureate or masters (preferred)	Community, medical-surgical	Orientation; role-playing; supervisor mentors	1:90
Hospice of the Valley	RN; baccalaureate (preferred)	Cardiac, medical-surgical	One-week classroom orientation; worked with supervisor or experienced mentor for 2 to 6 weeks	1:40
Jewish Home and Hospital	RN; baccalaureate (required) Masters-prepared social worker	Community, geriatric	Orientation	1:70
Medical Care Development	RN, nurse practitioner, or physician's assistant	Cardiac, community	Orientation; worked with experienced mentor	1:70
Mercy	RN; baccalaureate or masters (required)	Specific experience not noted	Four-week orientation	1:50

EXHIBIT II.3 *(continued)*

	Education	Types of Nursing Experience Required[a]	Initial Program Training	Average Caseload[b]
QMed	RN or experienced licensed practical nurse	Specific experience not noted	Orientation	1:200
Quality Oncology	RN	Case management, community, oncology	Two-week orientation; close oversight by supervisor for 6 months	1:40
University of Maryland	Nurse practitioner	Cardiac	None noted	1:71
Washington University[c]	RN	Chronic care	Two-day orientation	1:50 for local 1:100 for telephone

[a]Community nursing includes home health, hospice, and public health nursing.

[b]Based on program staffs' reports of actual average caseloads after about 2 years of operations.

[c]The Washington University program uses St. Louis-based nurses who can see patients in person if necessary and StatusOne telephone call center nurses.

RN = registered nurse.

telephone call center-based care coordinators. The average caseloads of the local care coordinators and the center-based care coordinators were 50 patients and 100 patients, respectively.

2. Assessment and Care Planning

All programs began care coordination with a comprehensive patient assessment—a review of each new patient's medical and health service use history, current health, medications, health habits, functional status, and finances designed to identify the patient's barriers to improved health, and to determine his or her needs. Of the 15 programs, 10 conducted at least part of their assessments in person despite the fact that most of their interventions were conducted largely by telephone (Exhibit II.4). Nine programs expected to complete patient assessments within 2 weeks of enrollment, and four expected to complete them within 1 week. In fact, all but four programs came within 1 week, on average, of meeting their goals for timely assessment (data not shown). One program took more than 3 months, on average, to begin to assess patients. The program's staff reported that difficulty hiring care coordinators and scheduling assessment appointments with patients led to the delay.

All the programs used information from the assessments to develop written patient care plans that included patient goals, as well as care coordinator and patient activities required to meet those goals (Exhibit II.4). Programs updated care plans when patients met their current goals or experienced changes in health status or adverse events.

3. Monitoring

All 15 programs routinely monitored patients by telephone; a number also monitored them in person and with home telemonitoring devices (Exhibit II.4). Planned minimum monitoring frequency was difficult to categorize because, for some programs, it varied across patients (by

EXHIBIT II.4

KEY FEATURES OF ASSESSMENT, CARE PLANNING, AND MONITORING

	Assessment Mode	Expected Weeks to Assessment	Written Care Plan Developed	Care Plan Update Circumstances	Routine Monitoring Mode	Expected Monitoring Frequency[a]	Telemonitoring Used (Name)[b]
Avera	In person	Maximum of 2	Yes	Change in patient's status; patient met goals	Telephone	Weekly for first 6 months; twice monthly thereafter	Yes (HomMed)
Carle	Telephone primarily	Maximum of 2	Yes	Change in patient's status	Telephone primarily In person in clinic or at home	Weekly to quarterly by telephone; in person as necessary	No
CenVaNet	In person	Maximum of 2	Yes	Change in patient's status or every 6 months	Telephone primarily In person	At least monthly by telephone; at least every 6 months in person	Yes (Health Buddy)
Charlestown	Telephone primarily	Maximum of 2	Yes	Patient met goals	Telephone primarily In person at home or physician visit	Daily to monthly	No
CorSolutions	In person and telephone[c]	Less than 1	Yes	Patient met goals	Telephone	Every 2 weeks for first few months; monthly thereafter	No
Georgetown	In person	Maximum of 2	Yes	After adverse event or every 12 months	Telephone primarily In person at home	At least monthly	Yes (HomMed)
Health Quality Partners	Telephone primarily[d] In person at home for high-risk patients	Maximum of 2	Yes (except for lowest-risk patients)	At each contact	Telephone primarily In person for high- or moderate-risk patients	At least monthly	No
Hospice of the Valley	In person	Maximum of 1	Yes	Change in patient's status; patient met goals	Telephone and in person	As necessary during first 6 months; at least monthly thereafter	No
Jewish Home and Hospital	In person	Maximum of 2	Yes	At care coordinator's discretion[e]	Telephone and in person	At least monthly	Yes (Viterion)

EXHIBIT II.4 *(continued)*

	Assessment Mode	Expected Weeks to Assessment	Written Care Plan Developed	Care Plan Update Circumstances	Routine Monitoring Mode	Expected Monitoring Frequency[a]	Telemonitoring Used (Name)[b]
Medical Care Development	In person	Maximum of 4	Yes	After adverse event; patient met goals	Telephone primarily In person at cardiac rehabilitation center	Three or four times during first month; monthly thereafter	No
Mercy	In person	Maximum of 2	Yes	Annually	Telephone and in person	At least monthly	Yes (Tel-Assurance)
QMed	Telephone	Maximum of 1	Yes	At least every 6 months	Telephone In person for ischemia monitoring	Every other month	Yes (QMed OHMS/CAD)
Quality Oncology	Telephone	Maximum of 1	Yes	After adverse event; change in patient's status; patient met goals	Telephone	Weekly to monthly	No
University of Maryland	In person	Prior to random assignment[f]	Yes	Change in patient's status	Telephone	At least monthly	Yes (Phillips)
Washington University	Telephone	Maximum of 2	Yes	After adverse event; at least every 10 weeks	Telephone primarily In person	At least every 6 weeks	No

[a]Expected frequency includes contacts for education and service arranging as well as routine monitoring.

[b]Avera, Georgetown, and University of Maryland offered home telemonitoring devices to all their patients. CenVaNet, Jewish Home and Hospital, and Mercy offered devices to a small number of patients. QMed's device was an ambulatory ischemia monitor, rather than a home telemonitor.

[c]During the first 2 years of operations, CorSolutions contracted with home health nurses to conduct part of its assessment in person; telephone call center-based care coordinators conducted the remainder. During the 3rd year, the program made the in-person assessment optional at the discretion of the patients' physicians.

[d]The Health Quality Partners program first administered the Sutter Health Questionnaire to patients before randomization to gauge patients' "risk" (that is, the severity of their condition and the risk of adverse events); those determined to be at the lowest risk (well-controlled chronic conditions) were not eligible to participate. Patients in the next three risk levels (low, moderate, and high) were randomized within each risk stratum to treatment and control groups. Although this risk stratification process was not part of the initial assessment, treatment group members' risk levels were later used to determine the intensity of intervention.

[e]During the 3rd year of operations, Jewish Home and Hospital began updating patient care plans every 60 days.

[f]The University of Maryland's program is designed to test the effect of home telemonitoring only. Consequently, all consenting eligible beneficiaries are assessed and receive some education prior to random assignment.

formally or informally assessed level of patient risk) or over time (for example, frequently for an initial period and at least monthly thereafter). However, only two programs did not require at least monthly monitoring. Routine monitoring included discussion of symptoms and other health issues and the provision of emotional support. In 11 of the 15 programs, almost all patients enrolling during the first year of operations (98 percent or more) had at least one contact for routine monitoring during the first year after enrollment (Table II.2). (Some patients had no monitoring contacts because they died or disenrolled shortly after enrollment.) Programs varied widely as to whether their care coordinators provided emotional support to patients. Seven programs recorded providing emotional support to roughly three-fourths or more of their patients, whereas five recorded providing it to fewer than one-third of their patients.

Six programs provided patients with home telemonitoring devices that transmitted weights, other clinical indicators, and responses to patients' questions about symptoms to their care coordinators each day (Exhibit II.4).[4] Three of those programs offered the devices to all their patients. Between 88 and 97 percent of the three programs' patients had contacts with care coordinators to discuss abnormal results (out-of-range telemonitor readings or laboratory results). The three other programs used the devices on a more limited basis, either for only a few patients or for a short trial period.

[4] Another program (Mercy) provided telemonitoring through a computerized system that automatically dialed patients at specified intervals and asked pre-recorded questions to which patients responded by pressing numbers on their telephone keypads. This system did not require patients to have any devices at home.

TABLE II.2

CARE COORDINATORS' CONTACTS FOR ASSESSMENT AND ROUTINE MONITORING DURING THE YEAR AFTER RANDOM ASSIGNMENT, AMONG 1ST-YEAR ENROLLEES

	Percentage of Patients with Assessment Contacts	Average Weeks from Enrollment to 1st Assessment Contact	Percentage of Patients with Monitoring Contacts	Percentage of Patients with Contacts to Monitor Abnormal Results	Percentage of Patients with Contacts for Emotional Support	Number of Patients
Avera	100.0	2.7	93.2	97.3	46.9	157
Carle	99.2	3.4	98.6	32.8	59.9	1,151
CenVaNet	96.5	4.9	94.7	42.0	15.1	538
Charlestown	98.5	2.8	99.0	40.9	87.6	212
CorSolutions	88.5	1.6	100.0	45.8	0.0	366
Georgetown	100.0	3.0	98.0	88.0	98.0	53
Health Quality Partners[a]	99.5	2.3	99.5	58.4	60.4	243
Hospice of the Valley	89.1	1.0	100.0	17.3	25.9	236
Jewish Home and Hospital	96.4	13.3	85.3	9.4	97.6	271
Medical Care Development	91.4	4.9	86.6	31.0	73.2	196
Mercy	93.0	2.6	99.6	69.8	98.0	317
QMed	98.9	5.5	98.9	0.8	1.3	698
Quality Oncology	93.1	4.4	100.0	51.7	93.1	31
University of Maryland[b]	n.a.	n.a.	100.0	93.1	34.4	29
Washington University	95.0	2.2	98.3	72.5	94.7	715

Source: Data describing program contacts with treatment group patients were prepared by programs and submitted quarterly to Mathematica Policy Research, Inc.

[a]Health Quality Partners administered a risk stratification questionnaire to patients before randomization, and patients were randomized according to risk stratum. This risk stratification was not part of the initial assessment, however.

[b]The University of Maryland conducted its initial patient assessment prior to enrollment.

n.a. = not applicable.

4. Patient Education

All but one of the programs used patient education as the cornerstone of their approach to improving patients' health[5] (Exhibit II.5). Among the 14 programs that did include patient education, care coordinators were the staff primarily responsible for providing that service. Ten programs supplemented education by referring patients to community-based education programs, including workshops held at local hospitals, many of which were led by specialized staff, such as diabetes educators, pharmacists, or nutritionists.

As RNs, most of the care coordinators had received basic education that included instruction on how to provide patient education. In addition, many care coordinators had experience as community nurses, whose practice emphasizes patient education. Despite this background, nine programs provided additional training to incoming care coordinators on how to educate patients, which varied substantially in intensity and type of training. Five of the nine programs included patient education training as a part of their regular orientation for all new care coordinators. Four programs provided more-focused training on communication techniques, lifestyle/behavioral change, or learning theory.

All but 1 of the 14 programs that provided patient education used a standard curriculum based on nationally published guidelines. The curricula and educational materials were part of the electronic care coordination databases of five of the programs, which then guided care coordinators' educational contacts with patients. Six programs assessed patients individually on their need for education and learning barriers (such as low literacy or cognitive limitation) and then customized their educational interventions based on acuity, cognitive ability, or readiness to

[5] The 15th program differed markedly from the others in that its goal was to test home telemonitoring as a means of improving the health of patients with CHF. Thus, it neither provided education or coordination nor referred patients for support services.

EXHIBIT II.5

KEY FEATURES OF PATIENT EDUCATION

	Patient Educators	Program Training of Patient Educators[a]	Education Primarily Structured or Event-Based	Adaptation to Individuals	Education Effectiveness Assessment
Avera	Care coordinators Community education	Individualized CHF education training	Standard CHF curriculum and educational materials from pharmaceutical company	Worked with caregivers	Telemonitor data Knowledge assessment tool Self-reported behavior
Carle	Care coordinators Community education	Patient education part of overall staff training	Standard curricula developed by host	Worked with caregivers Non-English and picture versions of materials available	Clinical events and indicators Self-reported behavior
CenVaNet	Care coordinators Community education	Year 1: No special training provided Year 2: Provided training on lifestyle modification	Standard curriculum embedded in program software	Individualized based on acuity level, reading level, and cognitive ability	Clinical indicators Observation during visits Knowledge assessment tool Self-reported behavior
Charlestown	Care coordinators Community education	No special training provided	Standard curriculum based on published guidelines	Individualized based on initial assessment of educational need and cognition	Patients' health Observation during visits
CorSolutions	Care coordinators Community education	Communication techniques, behavior change, and learning theory training	Standard curriculum embedded in program software	Individualized based on cognitive level assessed during each contact	Clinical indicators Self-reported behavior and response to hypothetical situations
Georgetown	Care coordinators Hospital-based diabetes education	Patient education part of overall staff training	Standard curriculum based on published guidelines	Worked with caregivers Multi-reading level versions of materials available	Telemonitor data Observation during visits
Health Quality Partners	Care coordinators	Readiness-to-change assessment and patient education training	Established curricula developed by host based behavior change model	Individualized based on readiness-to-change assessment	Self-reported behavior
Hospice of the Valley	Care coordinators Community education	No special training provided	Standard curricula developed by host	Individualized based on readiness-to-change and cognition	Observation during visits Asked patients to explain concepts

EXHIBIT II.5 *(continued)*

	Patient Educators	Program Training of Patient Educators[a]	Education Primarily Structured or Event-Based	Adaptation to Individuals	Education Effectiveness Assessment
Jewish Home and Hospital	Care coordinators, nutritionists, and pharmacists conduct group education Nurse care coordinators provide one-on-one education to some patients	No special training provided	No curriculum; materials from a variety of sources	Individualized based on patient's needs and learning style	None
Medical Care Development	Care coordinators	Patient education and motivational interviewing conferences	Materials embedded in program software Education varied by hospital and care coordinator	Worked with caregivers Learning aids for low literacy available	Knowledge assessment tool
Mercy	Care coordinators Hospital-based education	Patient education part of overall staff training	Standard curriculum based on published guidelines	Worked with caregivers Non-English versions of materials available	Telemonitor data Observation during visits Self-reported behavior
QMed	Care coordinators	No special training provided	Standard curriculum based on published guidelines	Worked with caregiver Spanish-speaking staff and Spanish language materials available	Clinical indicators Self-reported behavior
Quality Oncology	Care coordinators	No special training provided	Standard curriculum embedded in program software	Spanish language materials available	Self-reported behavior Asked patients to explain concepts
University of Maryland	Patient education not part of program intervention[b]	n.a.	n.a.	n.a.	n.a.
Washington University	Care coordinators Community education	Patient education included in some monthly staff training sessions	Standard curriculum embedded in program software	Worked with caregivers Low-literacy and non-English materials available	Clinical indicators Self-reported behavior Providers' reports on behavior in some cases

[a]Training beyond the training that registered nurses will have had as a part of basic nursing education. Nurses with community nursing experience (for example, home health, public health, or hospice) will have additional experience providing patient education.

[b]The University of Maryland intervention does not include patient education; prior to random assignment, all enrolling patients receive written materials covering symptoms and dietary recommendations.

CHF = congestive heart failure; n.a. = not applicable.

change. Seven included caregivers in their education interventions if care coordinators were having difficulty educating patients directly.

All but one of the programs providing education had developed processes for assessing the effectiveness of the education. Eight programs reviewed clinical indicators or home telemonitoring data in order to determine whether the health of their patients was improving—the ultimate goal of patient education. The six others relied on patients' self-reports of behavior change (such as improvement in diet or medication adherence), direct observation of patients during care coordinators' visits, or the administration of a knowledge assessment tool.

Almost all the patients at the 14 programs providing education who enrolled during the first year of operations (85 percent or more) had a contact for education during the first year after random assignment (Table II.3). Care coordinators also explained tests, procedures, and medications to patients. Care coordinators for all but one program explained medications to at least one-half their patients during the patients' first year in the program; the care coordinators of eight of those programs explained medications to at least 80 percent of the patients. Care coordinators for nine programs explained tests or procedures to at least one-half their patients during the first year.

5. Communication and Care Coordination

Patients in a fee-for-service environment often receive care that is poorly coordinated across providers and settings. Of the 15 programs in the demonstration, 14 considered an important aspect of their interventions to be improving communication and coordination between providers and patients, either by teaching patients to do this themselves, or by doing it for them (Exhibit II.6). Twelve of the 14 programs taught patients how to manage their conditions themselves, and how to communicate more effectively with their physicians. They used a variety of techniques, including encouraging patients to take lists of questions to physician

TABLE II.3

CARE COORDINATORS' CONTACTS FOR EDUCATION DURING THE YEAR AFTER RANDOM ASSIGNMENT, AMONG 1ST-YEAR ENROLLEES

	Percentage of Patients with Education Contacts	Percentage of Patients with Contacts to Explain Tests or Procedures	Percentage of Patients with Contacts to Explain Medications	Number of Patients
Avera	95.9	14.0	84.5	157
Carle	95.2	81.1	82.4	1,151
CenVaNet	99.2	62.2	61.8	538
Charlestown	98.0	68.0	97.6	212
CorSolutions	86.0	0.0	81.2	366
Georgetown	100.0	78.0	98.0	53
Health Quality Partners	99.5	99.1	99.5	243
Hospice of the Valley	93.0	13.8	78.7	236
Jewish Home and Hospital	84.9	12.6	7.5	271
Medical Care Development	96.7	63.1	73.2	196
Mercy	99.3	35.7	68.8	317
QMed	94.7	53.1	94.4	698
Quality Oncology	89.6	72.4	55.1	31
University of Maryland	31.0	44.8	82.7	29
Washington University	99.4	95.4	98.1	715

Source: Data describing program contacts with treatment group patients were prepared by programs and submitted quarterly to Mathematica Policy Research, Inc.

EXHIBIT II.6

KEY FEATURES OF EFFORTS TO IMPROVE COMMUNICATION AND COORDINATION

	Taught Patients or Coordinates for Patients	Had Regular Formal Communications with Physicians	Had Informal Communications with Physicians	Provided Co-Location and/or Caseload Allocation[a]	Had Role in Ensuring Adherence to Treatment Guidelines	Process for Learning About Adverse Events
Avera	Primarily taught patients	Program telemonitor trend reports	Telephone, fax as necessary	Neither	Physician reminded about preventive care Physician sent patient medication inventory	Missing telemonitor readings
Carle	Primarily taught patients	Conferences between care coordinator and physician at least annually, sometimes included patient	Regular in-person contact, telephone	Both	Physician notified if deviated from guidelines Care coordinators able to order tests	Email alerts concerning all patient encounters in Carle system
CenVaNet	Primarily taught patients	Program care plans Updates on patients' progress at least annually	Telephone for urgent patient problems	Neither	Patients received standard-of-care cards to enable them to remind physicians about necessary care	Patients'/caregivers' self-reports
Charlestown	Primarily coordinated for patients	Program care plans	Regular in-person contact, email	Co-location: yes Caseload allocation: no	Discussed on a patient-by-patient basis	A variety of host community-generated reports Patients' self-reports
CorSolutions	Primarily taught patients	Patient summaries every 90 days	Telephone for urgent patient problems	Co-location: no Caseload allocation: yes	Patient summary highlighted deviations from guidelines	Patients' self-reports
Georgetown	Primarily taught patients	Program telemonitor trend reports	Telephone for urgent patient problems	Neither	Multidisciplinary team reviewed cases and makes treatment suggestions	Missing telemonitor readings

EXHIBIT II.6 *(continued)*

	Taught Patients or Coordinates for Patients	Had Regular Formal Communications with Physicians	Had Informal Communications with Physicians	Provided Co-Location and/or Caseload Allocation[a]	Had Role in Ensuring Adherence to Treatment Guidelines	Process for Learning About Adverse Events
Health Quality Partners	Primarily taught patients	Initial assessment Encounter report after each patient contact	Occasional in-person contact, telephone	Both	Initial assessment included recommendations for care Discussed on a patient-by-patient basis	Patients'/caregivers' self-reports Year 2: began reviewing hospital admissions data
Hospice of the Valley	Primarily taught patients	Patient summaries every 6 months	Patient office visits, telephone	Co-location: no Caseload allocation: yes	Discuss on a patient-by-patient basis	Patients' self-reports
Jewish Home and Hospital	Primarily coordinated for patients	Year 1: quarterly updates sent to some physicians Year 2: updates sent to all physicians as necessary	Telephone, email for urgent patient problems	Neither	None	One participating hospital notified program Patients' self-reports for those referred from other hospitals
Medical Care Development	Primarily taught patients	Patient summaries periodically	In person during hospital rounds	Both	Patients received health checklists to enable them to remind physicians about necessary care	Daily reviews of hospital inpatient and emergency room lists Patients' self-reports
Mercy	Primarily taught patients	Patients' change-of-status reports	Telephone primarily, in person for clinic-based care coordinators	Both	Patients taught to remind physicians about necessary care	Reviews of hospital emergency room lists Patients' self-reports
QMed	Primarily taught patients	Reports with ischemia monitoring results every 6 months Program quality manager visited physician practices	Telephone for urgent patient problems	Neither	Ischemia monitoring report compares clinical outcomes with treatment guidelines and notes deviations	Patients' self-reports Physicians' chart reviews

EXHIBIT II.6 *(continued)*

	Taught Patients or Coordinates for Patients	Had Regular Formal Communications with Physicians	Had Informal Communications with Physicians	Provided Co-Location and/or Caseload Allocation[a]	Had Role in Ensuring Adherence to Treatment Guidelines	Process for Learning About Adverse Events
Quality Oncology	Primarily taught patients	None	Telephone, fax for patient problems	Neither	Discussed on a patient-by-patient basis	Patients'/caregivers' self-reports
University of Maryland	Coordinating care not part of program intervention	None	Telephone, email for urgent patient problems	Co-location: no Caseload allocation: yes	None	Missing telemonitor readings
Washington University	Primarily taught patients	None	Written updates for patients about whom care coordinators were concerned Telephone, fax for urgent patient problems	Neither	Discussed on a patient-by-patient basis Program software provided reminders about recommended care	Participating hospital notified program Patients'/caregivers' self-reports

[a]Care coordinators worked in the same location as patients physicians; patients were assigned to care coordinators such that all of a physician's patients were assigned to just one care coordinator.

appointments, providing patients with clinical guidelines or schedules of tests that they should be receiving regularly, and educating patients to make more-informed decisions about treatment options. The remaining 2 of the 14 programs sought to improve coordination by having care coordinators do so on behalf of patients.

Most of the programs (12 of the 15) directly communicated with physicians either through regular written reports or, less frequently, through regular face-to-face meetings. Ten of the 12 sent to the physicians the patients' care plans, home telemonitor trend reports, or patients' health status summaries. One program held formal conferences with participating physicians. One had its quality manager visit physicians to discuss adherence to evidence-based practice guidelines in the context of patients' home monitoring data trends. Finally, all of the programs contacted physicians via telephone, fax, or email to discuss urgent patient problems. Care coordinators for six programs also had the opportunity for informal, in-person contacts with physicians.

The ease with which care coordinators were able to communicate with physicians also depended on whether the care coordinators and physicians were physically located in the same place (co-location), as well as on whether care coordinators were assigned to monitor all the patients of particular physicians (caseload allocation). An organizational link between program hosts and physicians facilitated placing care coordinators in or near physicians' offices. Care coordinators for five programs worked in the same physical location as their patients' physicians all or some of the time; of these, only one program had no organizational link to the physicians.

Most of the programs (13 of 15) viewed assisting physicians in adhering to evidence-based treatment guidelines as a part of their approach to improving patients' health, although they did so in different ways. Five programs either had care coordinators remind physicians about necessary tests or suggested treatment changes that they sent to the physicians. Three programs

prompted patients to remind physicians about necessary care. The other five had care coordinators diplomatically discuss deviations from guidelines with physicians on a patient-by-patient basis.

An important shortcoming of fee-for-service care is that, when adverse events befall patients (for example, unplanned hospital admissions or ER visits), no single provider takes responsibility for determining the cause of the event and for developing strategies for avoiding its repetition. The demonstration programs established a variety of processes to learn about and address adverse events. Programs that provided patients with home telemonitors received the timeliest notification of adverse events. The failure of a patient to transmit a daily reading from his or her device signaled the care coordinator to call the patient or the patient's emergency contact; during the call, the care coordinator would have been able to determine whether the patient had been to the hospital. Eight programs received notifications of patient admissions from participating hospitals or providers or reviewed hospital admissions logs. Four programs relied solely on patients' or caregivers' reports of adverse events, but any program that did not contact patients frequently might not receive the reports until weeks after the events had occurred.

6. Service and Resource Arranging

Programs also sought to improve patients' health by increasing the patients' access to support services that are not covered by Medicare (such as home care; transportation; certain equipment and supplies; and disease-specific, diet, or smoking-cessation support groups). Although none of the programs considered improving access to such non-Medicare covered support services a primary focus of their efforts, they recognized that the availability of support services could be crucial for at least some of their patients. Thus, all but one program assessed patients' needs for non-Medicare support services or additional Medicare-covered services during the patients' first year in the program through contacts to identify these needs (Table II.4).

TABLE II.4

CARE COORDINATORS' CONTACTS TO ARRANGE FOR SERVICES DURING THE YEAR AFTER RANDOM ASSIGNMENT, AMONG 1ST-YEAR ENROLLEES

	Percentage of Patients with Contacts to Identify Medicare Service Needs	Percentage of Patients with Contacts to Identify Non-Medicare Service Needs	Percentage of Patients with Contacts to Monitor Service Receipt	Number of Patients
Avera	13.4	2.6	59.7	157
Carle	83.2	97.2	29.9	1,151
CenVaNet	24.6	43.7	28.6	538
Charlestown	67.1	50.0	70.9	212
CorSolutions	0.0	66.2	17.3	366
Georgetown	4.0	10.0	92.0	53
Health Quality Partners	16.4	11.1	19.3	243
Hospice of the Valley	3.4	35.0	11.2	236
Jewish Home and Hospital	21.3	80.6	41.8	271
Medical Care Development	30.4	12.8	32.6	196
Mercy	8.6	99.3	16.5	317
QMed	0.2	9.2	9.5	698
Quality Oncology	17.2	17.2	89.6	31
University of Maryland	0.0	0.0	3.4	29
Washington University	28.1	35.9	72.1	715

Source: Data describing program contacts with treatment group patients were prepared by programs and submitted quarterly to Mathematica Policy Research, Inc.

However, only five programs did so for more than half their patients (Table II.4). Of the 15 programs, 11 reported that they either had funds to pay for the home telemonitoring devices that were part of their interventions or had more limited funds to pay for other goods and services (or planned to provide them directly) (Table II.5). In fact, other than those who received the telemonitors, few program patients received goods and services during their first year in the program. One program paid for monitored exercise for roughly 30 percent of its patients; another program with a telephonic intervention paid home health nurses to conduct in-person assessments for almost two-thirds of its patients and provided medication cassettes for about 40 percent of its patients (Table II.5).

Access to prescription drugs is particularly important to chronic disease management, as even beneficiaries with drug coverage may have needs that exceed their coverage. Two programs had allocated limited funds to help patients to close drug coverage gaps.[6] In addition, one-half of CorSolutions's treatment group was randomly assigned to a separate arm of the study, under which they could receive coverage of all prescription medications if they had income less than 200 percent of the federal poverty level and lacked other prescription drug coverage. However, the program purchased prescription drugs for only 36 treatment group members (roughly 20 percent of the patients randomly assigned to the prescription coverage arm).

7. Efforts to Engage Physicians

Although most of the programs recognized physicians as a critical part of a patient's healthcare team, most programs tried to minimize the burden that their interventions placed on

[6] Staff from these two programs (Georgetown and QMed) noted that their programs had limited funds to pay for medications related to the programs' primary diagnoses; Table II.5 shows that Georgetown paid for medications for 6 percent of its patients during their first year in the program, but that QMed made no such purchases. Although Washington University staff reported they had *not* allocated funds to purchase goods or services for its patients, Table II.5 suggests that they did in fact purchase medications for 7 percent of them.

TABLE II.5

PROGRAMS' PROVISION OF GOODS AND SERVICES DURING THE YEAR AFTER RANDOM ASSIGNMENT, AMONG 1ST-YEAR ENROLLEES

	Percentage of Patients for Whom Program Provided or Paid for:							
	Exercise[a]	Home Health	Rx[b]	Medication Cassette	Transportation	Home Monitor[c]	Program Planned to Provide Goods or Services	Number of Patients
Avera	0.0	0.0	0.0	0.0	0.0	83.4	Yes	157
Carle	0.0	0.0	0.0	0.0	0.3	0.6	Yes	1,151
CenVaNet	0.0	0.0	0.0	0.0	0.0	0.0	No	538
Charlestown	0.0	0.0	0.0	11.8	0.0	0.0	Yes	212
CorSolutions	0.0	63.9	9.8	41.3	0.0	30.9	Yes	366
Georgetown	0.0	0.0	5.7	0.0	5.7	81.1	Yes	53
Health Quality Partners	0.0	0.0	0.0	0.0	0.0	0.0	No	243
Hospice of the Valley	0.0	0.0	0.0	0.0	0.0	0.0	Yes	236
Jewish Home and Hospital	3.3	1.5	0.0	4.8	18.5	0.0	Yes	271
Medical Care Development	29.6	0.0	0.0	0.0	0.0	0.0	Yes	196
Mercy	0.6	0.6	0.0	0.0	0.0	1.3	No	317
QMed	0.0	0.0	0.0	0.0	0.0	0.0	Yes	698
Quality Oncology	0.0	0.0	0.0	0.0	0.0	0.0	No	31
University of Maryland	0.0	0.0	0.0	0.0	0.0	0.0	Yes	29
Washington University	0.0	0.0	7.4	0.0	0.6	0.6	Yes	715

Source: Data describing program contacts with treatment group patients were prepared by the programs and were submitted quarterly to Mathematica Policy Research, Inc.

Note: A few programs provided other goods and services to patients, including home-delivered meals, personal care/homemaker services, durable medical equipment, and prescription medication set-up services.

[a]Fitness counseling or monitored exercise.

[b]Staff from Georgetown and QMed noted that their programs had limited funds to pay for medications related to the programs' primary diagnoses. Although Washington University staff stated that they did not have funds to purchase any goods or services for their patients, the data presented in this table suggest that they purchased medications (and monitoring devices) for some patients.

[c]Home monitors for Avera, Georgetown, and Mercy are telemonitoring devices. (The University of Maryland also provided telemonitoring devices for all its patients but did not record the provision of those devices in the data that it prepared for the evaluation.) Home monitors for CorSolutions are bathroom scales; for Carle, they are blood pressure cuffs, glucometers, and peak flow meters; for Washington University they are wearable emergency response devices.

Rx = prescription drugs.

physicians. Only four programs adopted direct improvement of provider practice as an approach to improving patient health (Exhibit II.7). They did so either by comparing the physicians' treatment plans with evidence-based guidelines and feeding back recommendations to the physicians or, in the case of one program, by directly providing both education about treatment guidelines and incentives to participate in that education.

Rather than trying to affect clinical practice directly, most programs regarded themselves as the physicians' "eyes and ears" outside office visits. Almost every program (14 out of the 15) asked physicians to review potential patients for program appropriateness, although only one-half of the programs expected physicians to actively refer or encourage their patients to participate in the demonstration. All of the programs expected physicians to respond to requests from care coordinators when contacted about specific urgent patient problems, but only four programs expected physicians to call care coordinators with new information about their patients. Most of the programs that expected physicians to actively encourage patients to enroll or to initiate contacts with care coordinators were largely disappointed and had to find strategies to compensate (for example, by devoting more time to encouraging potential patients themselves or relying on patients' self-reports of important health-related events). Only five programs asked physicians to provide care coordinators with standing orders to order routine tests or recommend changes in medication dosages for their patients.

Programs engaged in a variety of strategies to achieve physicians' support of and active participation in their care coordination efforts (Exhibit II.8). Five programs used physician opinion leaders to increase awareness of their programs, and to encourage physicians to participate. Three programs developed physician advisory boards that were responsible for either assisting with program design or eliciting feedback to improve the intervention. Nine programs paid physicians for their participation. Four programs provided a modest per patient

EXHIBIT II.7

PROGRAMS' EXPECTATIONS OF THEIR PHYSICIANS

	Improve Physician Clinical Practice	Review Patients for Program Appropriateness	Refer Patients Directly to Program/ Encourage Patients to Participate[a]	Participate in Program Care Planning	Respond to Care Coordinators' Requests	Call Care Coordinators with New Information	Provide Standing Orders to Care Coordinators[b]
Avera	Yes	Yes	Yes	Set telemonitor parameters	Yes	Yes	No
Carle	Yes	Yes	Yes	Provide input to initial plan Review revised plan annually	Yes	Yes	Yes
CenVaNet	No	Yes	Yes	Receive copies of care plan Review care plan if patient has special needs (DME, other) Set telemonitor parameters if patient uses one	Yes	No	No
Charlestown	No	Yes	No	Receive copies of care plan	Yes	No	No
CorSolutions	Yes	Yes	Year 1: yes Year 2: no	No	Yes	No	No
Georgetown	No	Yes	No	Approve initial care plan Set telemonitor parameters	Yes	No	Yes

EXHIBIT II.7 *(continued)*

	Improve Physician Clinical Practice	Review Patients for Program Appropriateness	Refer Patients Directly to Program/ Encourage Patients to Participate[a]	Participate in Program Care Planning	Respond to Care Coordinators' Requests	Call Care Coordinators with New Information	Provide Standing Orders to Care Coordinators[b]
Health Quality Partners	No	Yes	No	No	Yes	No	No
Hospice of the Valley	No	No	Yes	No	Yes	No	Yes
Jewish Home and Hospital	No	Yes	Yes	Set telemonitor parameters if patient uses one	Yes	Yes	No
Medical Care Development	No	Yes	Yes	No	Yes	No	No
Mercy	No	Yes	No	Review initial plan and revised plan annually	Yes	No	Yes
QMed	Yes	Yes	Yes	No	Yes	Yes	Year 1: no Year 2: yes
Quality Oncology	No	Yes	No	No	Yes	No	No
University of Maryland	No	Yes	Yes	No	Yes	No	No[c]
Washington University	No	Yes	No	No	Yes	No	No

[a]A notation of "yes" in this column indicates that the program expected physicians to be major sources of patient referrals or to take an active role in encouraging patients to enroll and participate. All programs accepted patients referred by physicians even if they did not expect the programs to be major referral sources.

[b]Standing orders included those to order routine tests or to recommend medication dosage changes to patients (most commonly, for diuretics).

[c]The University of Maryland's care coordinator is a nurse practitioner; as such, she has the authority to change a patient's medication dosage.

DME = durable medical equipment.

EXHIBIT II.8

PROGRAMS' APPROACHES TO ENGAGING PHYSICIANS

	Physician Opinion Leaders or Physician Advisory Board	Monetary Compensation	Other
Avera	Both	$30 per patient per month	Care coordinators met with physician after patient enrolls
Carle	Both	Paid to attend formal meetings with care coordinators[a]	Carle's medical directors encouraged physicians to actively participate in program
CenVaNet	Neither	No	
Charlestown	Neither	$26 per patient per month	
CorSolutions	Opinion leader: yes Advisory board: no	$50 per quarterly teleconference, $30 per additional teleconference	Year 1: actively marketed to area physicians
Georgetown	Opinion leader: program medical director communicates with peers Advisory board: no	$100 per in-person case conference with care manager	
Health Quality Partners	Opinion leader: no Advisory board: program medical director elicited feedback from participating physicians	No	Physicians asked to provide preferences for how and when care coordinators should contact them
Hospice of the Valley	Neither	No	Care coordinators attended physician office visits with patients
Jewish Home and Hospital	Opinion leaders: program medical directors encouraged physicians in their practices to participate Advisory board: no	$27.75 per patient per month for physicians in one practice; similar payments planned for the other practice	
Medical Care Development	Opinion leaders: local hospital medical directors encouraged physicians to participate Advisory boards: each hospital has MECare board	$20 per patient per month	
Mercy	Neither	No	
QMed	Opinion leaders: yes Advisory board: no	$25 for initial report review $50 for subsequent reports	

EXHIBIT II.8 *(continued)*

	Physician Opinion Leaders or Physician Advisory Board	Monetary Compensation	Other
Quality Oncology	Opinion leaders: added after year 1 Advisory board: no	$40 for providing medical records	
University of Maryland	Neither	$100 per patient per month	
Washington University	Opinion leaders: no Advisory board: yes	Planned to do so but did not do so in year 1	

[a]Carle's physicians are salaried; they receive patient care credit for attending meetings with care coordinators.

per month stipend ranging from $20 to $30, and one paid physicians $100 per month. Four others paid physicians between $30 and $100 for specific tasks, such as reviewing reports, participating in teleconferences, providing medical records, or meeting with care managers in person.

8. Data Systems and Reporting

Each program had some type of electronic system to manage data on patient enrollment and program activities, although the amount of data stored on and sophistication of the systems varied widely (Exhibit II.9). Four programs whose hosts (or collaborating partners) were disease management providers used the systems developed for their commercial clients. Six programs purchased commercial case management software products (Canopy®, HomeWorks®, InformaCare®, or Clinical Management System®). Four either used databases they had developed for previous case management projects or developed databases specifically for the demonstration (for example, using Microsoft Access). The remaining program, whose intervention was simply to provide home telemonitoring, relied on the data system associated with the telemonitor for its electronic recordkeeping needs. In addition to these primary data systems, six programs had access to systems that provided additional patient information, access to medical records, email alerts about medical encounters, and information on potential drug interactions.

The extent to which care coordinators used data systems with their daily work and program administrators used the systems to generate management reports varied across the 15 programs. Care coordinators in 12 of the 15 programs regularly used the primary data systems to support

EXHIBIT II.9

DATA SYSTEMS AND REPORTING

	Program's Primary System	Care Coordinators Routinely Used Primary System	Other Electronic Systems to Which Care Coordinators Have Access	Reports Generated		
				Program Process/ Reminders	Patient Behavior	Clinical Indicators/ Outcomes
Avera	Microsoft Access database	No	Microsoft Outlook HomMed telemonitor system	Yes	No	Yes
Carle	Carle Care Management Information System	Yes	Carle's medical records system Carle's physician appointment scheduler Email alerts on Carle's medical records system for ER visits, admissions, and other encounters	Yes	Yes	Yes
CenVaNet	InformaCare commercial disease management software	Yes	None	Yes	Yes	Yes
Charlestown	Canopy commercial web-based case management software	Yes	Web site that checks for drug interactions	Yes	No	No
CorSolutions	CorSolutions CorConnect	Yes	None	Yes	Yes	Yes

EXHIBIT II.9 *(continued)*

	Program's Primary System	Care Coordinators Routinely Used Primary System	Other Electronic Systems to Which Care Coordinators Have Access	Reports Generated		
				Program Process/ Reminders	Patient Behavior	Clinical Indicators/ Outcomes
Georgetown	Canopy commercial web-based case management software	Yes	HomMed telemonitor system	No	No	Yes
Health Quality Partners	Microsoft Access database	Yes	None	Yes	No	Yes
Hospice of the Valley	HomeWorks for Hospice commercial case management software	No	None	Yes	Yes	Yes
Jewish Home and Hospital	Canopy commercial web-based case management software	No	Viterion telemonitor system	No	No	No
Medical Care Development	Clinical Management Systems commercial disease management software	Yes	None	Yes	Yes	Yes
Mercy	Mercy Case Management Information System	Yes	Mercy hospital database Tel-Assurance telemonitor system	No	Yes	Yes
QMed	QMed's OHMS, PIMS, and PAT	Yes	None	Yes	Yes	Yes

EXHIBIT II.9 *(continued)*

	Program's Primary System	Care Coordinators Routinely Used Primary System	Other Electronic Systems to Which Care Coordinators Have Access	Reports Generated		
				Program Process/ Reminders	Patient Behavior	Clinical Indicators/ Outcomes
Quality Oncology	Quality Oncology Integrated Care Management System	Yes	None	Yes	Yes	Yes
University of Maryland	Philips telemonitor system	Yes	None	No	No	Yes
Washington University	StatusOne CareLink case management software	Yes	None	Yes	No	No

ER = emergency room; OHMS = On-line Health Management System; PAT = Patient Automated Tracking System; PIMS = Patient and Information Management System.

their work with patients, whereas those in the 3 other programs relied largely on paper records.[7] Eleven programs used their data systems to produce reports reminding care coordinators about when to contact patients, or when patients might be due for tests or treatment. Eight programs used their data systems to produce reports on patient behavior, and 12 programs had data systems capable of producing reports on clinical indicators and patient outcomes.

9. Summary

Although each of the programs provided a unique intervention, they shared several common features that the literature associates with success in improving patient health and in reducing health care costs. All the program interventions began with a comprehensive patient assessment, and all of them regularly monitored patients' symptoms either through the use of home monitoring devices or through regular contact with care coordinators, most of whom were experienced RNs. Almost all the programs (14 of 15) included patient education as an integral part of their interventions, and they attempted to decrease care fragmentation through a variety of means. The strategies included (1) teaching patients to communicate more effectively with their physicians, (2) encouraging physicians to adhere to evidence-based treatment guidelines, and (3) helping patients to avoid repeated adverse events. Finally, although relatively few patients who enrolled in the demonstration required support services or health-related goods, most of the programs had the capacity to identify the need for additional goods and services, and to help patients to arrange for their receipt.

[7] The three other programs also maintained electronic databases that recorded program activities and/or information on treatment group members, but the databases were used to monitor program performance; they were not electronic medical records useful for daily patient management. Nurse case managers in these programs thus relied on paper records and separately entered data into the databases.

Although all the programs had prior experience with care coordination or disease management, each of the program hosts differed in the amount of experience it had with fee-for-service Medicare, as well as in its organizational ties to patients' physicians. Programs with less experience providing services to fee-for-service patients often had to devote significant staff time and resources to modifying their outreach and intervention strategies, because they had less leverage over physicians and less access to medical records and other data than they did in their managed care contracts. Programs without an organizational link to patients' physicians had to spend more time marketing to local physicians, and cultivating the physicians' support. In the absence of that support, program staff had to tailor program interventions to the reality of minimal physician involvement.

Programs also differed in their ability to identify and respond to patients' problems. Those that provided patients with telemonitors had the advantage of timely notification of worsening symptoms, ER visits, or hospital admissions. Thus, they were able to ensure that patients with worsening symptoms saw their physicians promptly, which helped to avert hospitalization. When patients did have adverse events, the absence of a daily reading served as a notification to the programs that the events had occurred; program staff were therefore able to contact the patients immediately on release from the hospital to make certain that the patients understood instructions from hospital staff, as well as to identify strategies to minimize the chances of a recurrence of the event. Nevertheless, staff of one program that used telemonitors expressed concern that patients were becoming dependent on the devices, and they tried to teach the patients how to self-monitor symptoms without using the devices.

The programs differed in their capacity to meet face-to-face with patients. Although all the programs had a telephonic component, some included in-person visits for assessment or ongoing monitoring. In-person contact provided care coordinators with an opportunity to glean additional

information about the patients' health status, cognitive ability, and capacity to live safely at home that they might not have been able to obtain over the telephone.

Although almost all of the programs provided education as a means of improving patients' adherence to physicians' treatment recommendations, the literature has not identified one approach to patient education that is uniformly considered to be "the most effective." Some of the programs primarily provided factual information about target conditions and common comorbidities and taught patients appropriate self-care techniques for their conditions. Other programs adopted lessons from learning theories about readiness to change and approaches to health behavior modification and applied them to their education interventions. A few programs took the former approach during their first year of operations but began to consider the latter as potentially more effective and thus modified their interventions during their second year. This change in strategies could result in different program effects over time.

Finally, although a few programs made concerted, systematic efforts to improve physician clinical practice, most tried to minimize the burden they placed on physicians and made few demands on the physicians' time. For the most part, care coordinators contacted physicians only about specific and urgent patient problems. The programs also tried to minimize paperwork and requests to review documents or to meet with program staff. Only one program provided additional education to physicians; the effort was facilitated by the fact that the program host was an integrated health system that employed the participating physicians.

C. QUANTITATIVE SCORING OF INTERVENTION FEATURES

1. Scoring Approach

In addition to developing qualitative descriptions of program interventions, the evaluation sought an approach that would allow the program features and components, such as patient education or service arrangement, to be assigned numerical "scores." Examining the correlations

between these scores on the one hand and impacts on outcomes on the other would theoretically help to identify the features most likely to lead to program success.

The evaluator developed a structured assessment form (contained in Appendix B) for research team members to score the programs according to specified intervention features. The form asked questions about 10 domains (listed in Table II.6) chosen on the basis of previous research (Chen et al. 2000). Most questions asked about the presence or absence of specific program characteristics related to each domain. For example, under the Program Staffing domain, a question asked whether care coordinators were required to be RNs, and under the Patient Education domain, a question asked whether the program used a curriculum for patient education. However, the form also contained Likert scale questions that sought scorers' judgments on how well program components appeared to meet the needs of the target population both on an absolute basis and in relation to services available to enrollees in the absence of the program. As noted, these scores were meant to be quantitative descriptors. Several programs were expected to have low scores in one or more areas, as they had designed interventions that intentionally did not emphasize certain aspects of care coordination. (For example, several programs did not view improvement of provider practice as a focus of their interventions.)

To standardize responses, the form contained detailed definitions of terms and instructions on where to find the information for each question. The sources of information were program documents (ranging from the original program proposals to subsequent memos, care manager training materials, and protocols), notes from the evaluator's telephone and site visit interviews, and previous evaluation reports and program profiles prepared from these materials. Because these scores are based on the descriptions of program features as conveyed by documents and program staff, they portray the extent and intensity of the actual implementation of these elements only incompletely. The survey data of participants (presented in Chapters IV and V),

and of physicians (Chapter IV), in which respondents reported their perceptions of whether certain processes of care occurred, provide additional information on implementation of program features and thus complement the scores presented here. The correlations of the scores with impact estimates from participant survey data are examined in Chapter VII.

Answers within each domain were summed so that higher scores indicated a greater number of features, greater apparent intensity of those features, or both. The scores were rescaled to a 0-to-100 range. Scores for all 10 domains were summed to produce a combined score, also converted to a 0-to-100 range.

Each program was assessed independently by two research team members from a pool of five scorers. (Thus, not all 15 programs were assessed by the same 2 raters.) Scorers were not permitted to see the impact estimates for programs they were scoring. It is important to assess the degree to which scorers' subjectivity might have contributed to observed variation in the scores. A standard approach to quantitating scorers' consistency in assessing programs is to calculate a statistic called the "intraclass correlation coefficient" (McGraw and Wong 1996; Shrout and Fleiss 1979). The intraclass correlation coefficient measures the proportion of the total variation in scores that is due to variation between the subjects being scored (care coordination programs in this case), as opposed to that due to variation between scorers. The maximum possible value of an intraclass correlation coefficient is 1, indicating that all variation in scores is due to program variation (and none to inter-rater variability). Intraclass correlation coefficients as close to 1 as possible are thus desirable. A conventional classification scheme is to call intraclass correlation coefficient values from 0.8 to 1.0 "excellent," from 0.6 to 0.8 "good," from 0.4 to 0.6 "moderate," and 0.4 or less "fair to poor" (Landis and Koch 1977).

Table II.6 shows the intraclass correlation coefficients for each of the domains. (Appendix B contains additional details on the calculation of the intraclass correlation coefficients.) Four

TABLE II.6

INTER-RATER CONSISTENCY FOR PROGRAM DOMAINS, AS ASSESSED BY INTRACLASS CORRELATION COEFFICIENTS

Domain	Intraclass Correlation Coefficient	Conventional Classification
Program Staffing	0.69	Good
Initial Assessment	0.05	Fair to poor
Problem Identification and Care Planning	0.23	Fair to poor
Patient Education	0.90	Excellent
Improving Communication and Coordination	0.80	Good
Improving Provider Practice	0.82	Excellent
Service and Resource Arranging	0.72	Good
Information Technology and Electronic Records	0.93	Excellent
Ongoing Monitoring	0.57	Moderate
Quality Management and Outcome Measurement	0.83	Excellent
Combined Score	0.86	Excellent

Source: Intraclass correlation coefficients were calculated from two independent rating scores of each program by evaluator research staff. They were calculated using a two-way analysis of variance model (McGraw and Wong 1996; Shrout and Fleiss 1979) in which raters are considered random effects.

Note: Intraclass correlation coefficients are a measure of the consistency or reliability of scorers' assessments of programs. The closer their values are to 1, the greater the amount of the variation in scores is due to variation among programs, and not due to variation in scoring practices. The conventional classification scheme is from Landis and Koch (1977).

domains (Patient Education, Improving Provider Practice, Information Technology and Electronic Records, and Quality Management and Outcome Measurement) had excellent coefficients, as did the overall score. Three domains (Program Staffing, Improving Communication and Coordination, and Service and Resource Arranging) had coefficients in the good range, and one domain (Ongoing Monitoring) had a moderate coefficient. Finally, two domains (Initial Assessment and Problem Identification and Care Planning) had poor to fair coefficients. Less weight or importance was thus attached to the scores for these last two domains.

2. Rating Score Results

Not surprisingly, given the diversity of program approaches and the populations that the programs targeted, there was large variation within domains. Table II.7 shows the program rating scores for each domain, listed in descending order, as well as the average scores for each domain. The programs had the highest average scores for Ongoing Monitoring (mean of 68).[8] Consistent with the previous observation that few programs had sought to influence clinical practice, Improving Provider Practice had the lowest average score (32). Average scores for the remaining domains ranged from 54 to 65.

Table II.7 also shows the variability within each domain, as measured by the range of scores (the differences between the highest and lowest). The two domains with the narrowest ranges were the Problem Identification and Care Planning domain, in which scores ranged from 38 to 83, and the Initial Assessment domain, with scores between 50 and 96. As previously noted, however, the inter-rater reliability of these two scores was low. The two domains with the

[8] Initial Assessment also had a high average score of 78, but, as noted, the inter-rater reliability of Initial Assessment was only fair to poor.

TABLE II.7

DEMONSTRATION PROGRAMS IN DESCENDING ORDER OF RATING SCORES FOR EACH DOMAIN

	Program Staffing (ICC = 0.69)		Initial Assessment[a] (ICC = 0.05)		Problems[a] Identification and Care Planning (ICC = 0.23)		Patient Education (ICC = 0.90)		Improving Communication and Coordination (ICC = 0.80)
Carle	80	Georgetown	96	Mercy	83	Health Quality Partners	88	Carle	94
Quality Oncology	75	CenVaNet	95	Georgetown	81	Mercy	85	Mercy	81
Mercy	70	Health Quality Partners	91	Avera	76	CorSolutions	82	Charlestown	74
Health Quality Partners	65	Avera	89	CorSolutions	74	Carle	80	Georgetown	65
Georgetown	65	Carle	87	Carle	74	Hospice of the Valley	80	Health Quality Partners	61
CorSolutions	65	CorSolutions	87	Charlestown	69	Avera	77	Hospice of the Valley	61
Charlestown	65	Quality Oncology	85	Medical Care Development	69	Charlestown	77	Washington University	61
CenVaNet	65	Medical Care Development	82	Hospice of the Valley	67	CenVaNet	71	CenVaNet	59
Washington University	60	Mercy	80	Health Quality Partners	64	Georgetown	71	Medical Care Development	59
Medical Care Development	55	Charlestown	74	Washington University	64	Quality Oncology	71	CorSolutions	54
Hospice of the Valley	55	Hospice of the Valley	74	QMed	62	QMed	62	QMed	54
University of Maryland	50	Washington University	65	CenVaNet	62	Medical Care Development	59	Avera	52
Avera	50	Jewish Home and Hospital	59	Quality Oncology	60	Washington University	56	Quality Oncology	48
QMed	30	QMed	54	Jewish Home and Hospital	52	Jewish Home and Hospital	12	University of Maryland	30
Jewish Home and Hospital	20	University of Maryland	50	University of Maryland	38	University of Maryland	6	Jewish Home and Hospital	22
Range	**60**	**Range**	**46**	**Range**	**45**	**Range**	**82**	**Range**	**72**
Average	**58**	**Average**	**78**	**Average**	**66**	**Average**	**65**	**Average**	**58**

TABLE II.7 *(continued)*

	Improving Provider Practice (ICC = 0.82)		Service and Resource Arranging (ICC = 0.72)		Information Technology and Electronic Records (ICC = 0.93)		Ongoing Monitoring (ICC = 0.57)		Quality Management and Outcome Measurement (ICC = 0.83)
Carle	77	Washington University	76	Carle	88	Quality Oncology	100	Carle	91
QMed	73	Jewish Home and Hospital	69	Quality Oncology	75	CenVaNet	82	Washington University	86
CorSolutions	65	Carle	67	University of Maryland	71	Avera	77	Quality Oncology	82
Avera	50	CorSolutions	67	CenVaNet	71	CorSolutions	77	Georgetown	77
Medical Care Development	35	Medical Care Development	64	CorSolutions	69	Health Quality Partners	77	QMed	77
Georgetown	31	Mercy	64	Charlestown	63	QMed	77	CenVaNet	68
CenVaNet	23	Georgetown	60	Medical Care Development	63	Carle	73	Hospice of the Valley	64
Washington University	23	Hospice of the Valley	60	Washington University	63	Charlestown	73	Health Quality Partners	64
Quality Oncology	23	Avera	55	Georgetown	52	Washington University	73	CorSolutions	59
Charlestown	19	CenVaNet	55	Mercy	50	Hospice of the Valley	68	Mercy	55
University of Maryland	19	Health Quality Partners	55	QMed	50	Georgetown	64	Medical Care Development	50
Health Quality Partners	15	QMed	50	Avera	38	Mercy	64	Charlestown	46
Mercy	15	Charlestown	45	Jewish Home and Hospital	27	Medical Care Development	41	Avera	14
Hospice of the Valley	12	Quality Oncology	33	Health Quality Partners	19	University of Maryland	41	University of Maryland	9
Jewish Home and Hospital	0	University of Maryland	7	Hospice of the Valley	19	Jewish Home and Hospital	32	Jewish Home and Hospital	5
Range	**77**	**Range**	**69**	**Range**	**69**	**Range**	**68**	**Range**	**86**
Average	**32**	**Average**	**55**	**Average**	**55**	**Average**	**68**	**Average**	**56**

Source: Rounded means of two independent rating scores of each program by evaluator research staff. Raters consulted program documents, telephone and site visit interview notes, and evaluation case studies and evaluation 1st year reports to complete structured assessment forms. The forms asked a series of questions on the 10 domains listed in the column headings.

Notes: The ICC is a measure of the consistency or reliability of scorers' assessments of programs. The closer the values are to 1, the greater the amount of the variation in scores that is due to variation among programs, and the less it is due to variation in scoring practices among scorers. By convention, values from 0.8 to 1.0 are called "excellent," from 0.6 to 0.8 "good," from 0.4 to 0.6 "moderate," and 0.4 or less "fair to poor" (Landis and Koch 1977).

Each domain consisted of several questions. A few examples of question topics within each domain are: (1) Program Staffing—training and background of care coordinators, supervision of care coordinators; (2) Initial Assessment—use of standardized assessment instruments, the breadth of areas covered in the initial assessment; (3) Problem Identification and Care Planning—production of a care plan, the comprehensiveness of care plans; (4) Patient Education—use of a curriculum, assessment of patients' mastery of the material; (5) Improving Communication and Coordination—program response to ER visits or unplanned hospitalizations with plans to prevent reoccurrence, nature and extent of communications between care coordinators and physicians; (6) Improving Provider Practice—use of evidence-based guidelines by physicians and care coordinators, use of educational workshops or local physician opinion leaders; (7) Service and Resource Arranging—program access to social workers, program relationships with community-based agencies; (8) Information Technology and Electronic Records—presence of an electronic care management record system, use of home monitoring technology; (9) Ongoing Monitoring—presence of protocols or system for deciding frequency of monitoring, presence of protocol for content of monitoring contacts; and (10) Quality Management and Outcome Measurement—designation of a staff member responsible for performing quality assessment and improvement activities, program-wide reporting of program performance.

[a]Given the fair to poor ICC values of the scores for these two domains, less importance or weight should be attached to these scores.

ER = emergency room; ICC = intraclass correlation coefficient.

widest ranges were the Quality Management and Outcome Measurement domain, with scores ranging from 5 to 91, and the Improving Provider Practice domain, with scores between 0 and 77. Differences between the highest and lowest scores in the remaining six domains ranged from 60 to 82, with most domains having a 70-point difference.

There was also wide variability in the scores within each program across domains, with several programs ranking both in the top and the bottom across two or more domains (Exhibit II.10). Despite this variability, a few programs had high ranking scores for several domains, namely, Carle, which had six scores in the top quintile (the top three programs), and Mercy and Quality Oncology, each with four scores in the top quintile. Likewise, there were a few programs with several rating scores in the bottom quintiles, namely, the Jewish Home and Hospital, with nine, and the University of Maryland, with seven. Three programs exhibited a pattern of having high scores in a single domain and low or lower scores in all the others (the Jewish Home and Hospital, the University of Maryland, and QMed), but there appeared to be no other general categories of patterns.

EXHIBIT II.10

QUINTILES OF DEMONSTRATION PROGRAMS' RATING SCORES, BY DOMAIN

	Quintile	Program Staffing (ICC = 0.69)	Initial Assessment (ICC = 0.05)[a]	Problem Ident.& Care Plan. (ICC = 0.23)[a]	Patient Education (ICC = 0.90)	Impr. Comm. and Coord. (ICC = 0.80)	Improving Provider Practice (ICC = 0.82)	Service & Resource Arrange. (ICC = 0.72)	Info. Tech. & Elec. Records (ICC = 0.93)	Ongoing Monitoring (ICC = 0.57)	Qual. Mgt. & Outcome Meas. (ICC = 0.83)		Combined Score (ICC = 0.86)
Avera	1			■						■			
	2		■		■		■						
	3							■					
	4					■			■				■
	5	■									■		
Carle	1	■				■	■	■	■		■		■
	2		■	■	■								
	3									■			
	4												
	5												
CenVaNet	1		■							■			
	2								■		■		■
	3	■			■	■	■						
	4			■				■					
	5												
Charlestown	1					■							
	2			■					■				■
	3	■			■					■			
	4		■				■				■		
	5							■					
CorSolutions	1				■		■						■
	2	■	■	■				■	■	■			
	3										■		
	4					■							
	5												

EXHIBIT II.10 *(continued)*

	Quintile	Program Staffing (ICC = 0.69)	Initial Assessment (ICC = 0.05)[a]	Problem Ident.& Care Plan. (ICC = 0.23)[a]	Patient Education (ICC = 0.90)	Impr. Comm. and Coord. (ICC = 0.80)	Improving Provider Practice (ICC = 0.82)	Service & Resource Arrange. (ICC = 0.72)	Info. Tech. & Elec. Records (ICC = 0.93)	Ongoing Monitoring (ICC = 0.57)	Qual. Mgt. & Outcome Meas. (ICC = 0.83)		Combined Score (ICC = 0.86)
Georgetown	1		■	■									
	2	■				■	■				■		■
	3				■			■	■				
	4									■			
	5												
Health Quality Partners	1		■		■								
	2	■				■				■			
	3			■							■		■
	4						■	■					
	5								■				
Hospice of the Valley	1												
	2				■	■							
	3			■				■			■		
	4	■	■							■			■
	5						■		■				
Jewish Home and Hospital	1							■					
	2												
	3												
	4												
	5	■	■	■	■	■	■		■	■	■		■
Medical Care Development	1												
	2						■	■					
	3		■	■		■			■				■
	4	■			■						■		
	5									■			

EXHIBIT II.10 *(continued)*

	Quintile	Program Staffing (ICC = 0.69)	Initial Assessment (ICC = 0.05)[a]	Problem Ident.& Care Plan. (ICC = 0.23)[a]	Patient Education (ICC = 0.90)	Impr. Comm. and Coord. (ICC = 0.80)	Improving Provider Practice (ICC = 0.82)	Service & Resource Arrange. (ICC = 0.72)	Info. Tech. & Elec. Records (ICC = 0.93)	Ongoing Monitoring (ICC = 0.57)	Qual. Mgt. & Outcome Meas. (ICC = 0.83)		Combined Score (ICC = 0.86)
Mercy	1	■		■	■	■							■
	2							■					
	3		■										
	4								■	■	■		
	5						■						
QMed	1						■						
	2									■	■		
	3												
	4			■	■	■		■	■				
	5	■	■										■
Quality Oncology	1	■							■	■	■		
	2												
	3		■				■						
	4				■								■
	5			■		■		■					
University of Maryland	1								■				
	2												
	3												
	4	■					■						
	5		■	■	■	■		■		■	■		■
Washington University	1							■			■		
	2												
	3	■				■	■		■	■			■
	4		■	■									
	5				■								

Source: Means of two independent rating scores of each program by evaluator research staff. Raters consulted program documents, telephone and site visit interview notes, and evaluation case studies and evaluation 1st year reports to complete structured assessment forms. The forms asked a series of questions on the 10 domains listed in the column headings.

EXHIBIT II.10 *(continued)*

Notes: The ICC is a measure of the consistency or reliability of scorers' assessments of programs. The closer the values are to 1, the greater the amount of the variation in scores that is due to variation between programs, and the less it is due to variation among scorers in scoring practices. By convention, values from 0.8 to 1.0 are called "excellent," from 0.6 to 0.8 "good," from 0.4 to 0.6 "moderate," and 0.4 or less "fair to poor" (Landis and Koch 1977).

Because there were 15 programs, each quintile consists of 3 programs.

[a]Given the fair to poor ICC values of the scores for these two domains, less importance or weight should be attached to these scores.

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

III. WHO ENROLLED IN THE DEMONSTRATION?

All of the programs participating in the Medicare Coordinated Care Demonstration (MCCD) recruited and enrolled patients based on the target and exclusion criteria that they developed. This chapter describes the enrollment strategies used by the organizations that hosted the 15 programs. It then compares actual enrollments and disenrollments after 2 years of operations with program targets. It also describes the patients who enrolled in each of the 15 programs by presenting demographic data, preenrollment characteristics and costs, diagnoses, and a comparison of participants and potentially eligible nonparticipants based on cost waiver calculations conducted by Mathematica Policy Research, Inc. (MPR).[9] Finally, the chapter compares the preenrollment characteristics of treatment and control group members. The chapter draws on qualitative interviews with program staff, as well as on Medicare claims data, data collected by each of the programs, and a patient survey conducted by MPR.

A. PATIENT IDENTIFICATION

Programs generally adopted one of two primary approaches to identifying beneficiaries who would be asked to participate: (1) obtaining lists of prospective enrollees from hospitals or health care networks (the approach used by nine programs), or (2) recruiting physicians who then referred patients to the program (used by six programs; see Table III.1). The six programs that had hospitals or health care systems as their host organizations generally identified potentially

[9] In August 2001, MPR provided an estimate of expected costs and net savings to the Centers for Medicare & Medicaid Services (CMS) from the demonstration for each of the 15 programs as a part of an Office of Management and Budget (OMB) waiver package. This waiver allowed demonstration programs, within the context of a coordinated care program, to be reimbursed for providing additional items and services not usually allowed under Medicare Parts A and B. However, Section 4016(e)(1)(B) of Public Law 105-33 (42 U.S.C. 1395b-1) requires the Secretary of the Department of Health and Human Services to ensure that the aggregate payments made by Medicare do not exceed the amount that Medicare would have paid if the demonstration projects under this section had not been implemented.

TABLE III.1

TARGET ENROLLMENTS VERSUS ACTUAL ENROLLMENTS AFTER TWO YEARS

Program (Start Date)	Enrollment		Primary Method to Identify Potential Enrollees	Most Likely Reason for Success or Shortfall
	Target[a]	Actual (Percentage of Target)		
Avera (6/4/02)	Year 1: 788 Year 2: 980 Final: 1,268	318 (40) 624 (64)	Generated list from host system	Shortfall: High refusal rate, difficulty identifying patients
Carle (4/19/02)	Year 1: 2,256 Year 2: 2,568 Final: 3,036	2,283 (101) 2,642 (103)	Generated list from host and other systems	Success: Physicians actively promoted program
CenVaNet (4/8/02)	Year 1: 1,048 Year 2: 1,120 Final: 1,228	1,074 (102) 1,305 (117)	Recruited physicians from host network	Success: Prior relationships with physicians, access to electronic records
Charlestown (4/29/02)	Year 1: 684 Year 2: 720 Final: 792	402 (59) 802 (111)	Generated list from host system	Success: Expanded eligibility criteria
CorSolutions (6/18/02)	Year 1: 1,750 Year 2: 2,071 Final: 2,392	671 (38) 2,162 (104)	Generated list from other system Recruits physicians	Success: Began recruiting from hospitals in addition to physicians
Georgetown (6/5/02)	Year 1: 730 Year 2: 1,330 Final: 2,050	108 (15) 199 (15)	Generated list from host and other systems	Shortfall: High refusal rate, overestimate of eligible participants
Health Quality Partners (4/30/02)	Year 1: 738 Year 2: 1,644 Final: 2,140	499 (68) 1,140 (69)	Recruited physicians	Shortfall: Lack of staffing resources, high refusal rate
Hospice of the Valley (8/15/02)	Year 1: 624 Year 2: 1,248 Final: 2,184	460 (74) 814 (65)	Generated list from other systems Recruited physicians Direct marketing	Shortfall: Difficulty obtaining hospital support, high refusal rate
Jewish Home and Hospital (6/17/02)	Year 1: 730 Year 2: 730 Final: 730	543 (74) 766 (104)	Chart review for 2 large geriatric group practices affiliated with program	Success: Continued reviewing charts, overcame slow initial enrollment
Medical Care Development (4/17/02)	Year 1: 1,048 Year 2: 1,932 Final: 2,436	393 (38) 876 (45)	Generated lists from participating hospitals	Shortfall: Lack of resources for recruiting, lack of physician support
Mercy (4/19/02)	Year 1: 482 Year 2: 890 Final: 1,214	627 (130) 865 (97)	Generated list from host system	Success: Physician support based on previous work

TABLE III.1 *(continued)*

Program (Start Date)	Enrollment: Target[a]	Enrollment: Actual (Percentage of Target)	Primary Method to Identify Potential Enrollees	Most Likely Reason for Success or Shortfall
QMed (7/12/02)	Year 1: 782 Year 2: 926 Final: 1,372	1,404 (180) 1,454 (157)	Recruited physicians	Success: Physician support based on previous experience with host
Quality Oncology (9/18/02)	Year 1: 2,132 Year 2: 2,420 Final: 2,852	63 (6) 141 (6)	Recruited physicians	Shortfall: Lack of physician support
University of Maryland (6/28/02)	Year 1: 678 Year 2: 678 Final: 678	29 (4) 137 (20)	Generated lists from host and other systems	Shortfall: High refusal rate, high rate of ineligibility among referrals
Washington University (8/16/02)	Year 1: 2,000 Year 2: 2,000 Final: 2,000	1,425 (71) 2,038 (102)	Generated list from host systems	Success: Intensive marketing to physicians and patients

Source: Program documents, Mathematica Policy Research, Inc. enrollment files, and interviews with program staff during second year of program operations.

[a]The final enrollment is the number of beneficiaries that the program expects to have by the end of the demonstration. Programs differed in their planned timelines for reaching full enrollment.

eligible beneficiaries primarily from lists of host-system patients, using automated screening along broad program eligibility criteria, such as diagnosis and Medicare coverage. Rather than screen patients based on particular diagnoses, one of the six programs partnered with a commercial disease management vendor and used the vendor's proprietary algorithm to identify high-risk patients for the program from the program's physician network records. Three of the programs also recruited other health systems to provide lists of their patients. Two programs did so during their first year of operations; the third started recruiting outside hospitals and physician practices toward the end of the first year. Of the three other programs that recruited from lists provided by hospitals or health systems, one was a retirement community that included its own primary care physicians in the program, and one recruited from a few local hospitals and a large hospital group practice list. The third, which partnered with 17 hospitals in Maine, identified potentially eligible beneficiaries while the beneficiaries were inpatients at 1 of the hospitals by reviewing admission logs daily.

Seven of the nine programs that first identified patients from electronic lists subsequently contacted the identified patients' physicians to discuss the program. Some of the seven programs then asked the physicians for permission to contact their patients. Two programs that relied on electronic contact lists contacted patients directly, without first approaching the patients' physicians. Eight of the nine programs also welcomed direct physician referrals and hoped that the numbers of referrals would increase as the programs became better known. (The ninth program enrolled only recently discharged inpatients.) During their first year, however, the eight programs identified the majority of their potential enrollees through the automated review of patient databases.

Six programs used primarily the second recruitment method of enlisting physicians to refer patients. The five programs with care coordination service providers as hosts first recruited

physicians who wished to have their patients participate in the demonstration and then worked with each physician to generate lists of potentially eligible and appropriate patients. Of these, one program began reviewing lists of patients generated by local hospitals after the first year to supplement the physician referrals. Rather than recruit patients directly from its own care system, the sixth program developed a partnership with two large geriatric physician group practices prior to implementation.

After they identified potentially eligible patients, 12 of the 15 programs introduced themselves to patients by sending letters signed by the patients' own physicians. (One program's letters were signed by the program's medical director.) Of the remaining three programs, two had their care coordinators telephone identified patients instead of sending letters. (In one of these two programs the care coordinator also introduced the program during in-person visits with hospitalized patients who had been identified while they were still in the hospital.) The enrollment staff of the third program that did not use letters first contacted patients while the patients were at the physician practice clinics that had identified them for the program. The clinics had provided the program in advance with lists of patients scheduled for clinic visits. The program's enrollment coordinator then determined whether the patients were eligible for Medicare and examined their medical charts to verify that the patients had one of the program's target diagnoses. During the clinic visits, their physicians briefly discussed the program and asked whether the patients would like to meet with the enrollment coordinator at that time.

Program staff of most of the 15 programs reported that physicians were too busy and visits too short for the physicians to promote the program to patients directly. Program staff handled most of the "marketing" of the program that followed the mailing of the introductory letter or during the introductory telephone call or in-person encounter. The staff did report that, if

patients specifically asked physicians whether they should participate, most physicians encouraged them to do so.

B. ENROLLMENT AFTER 2 YEARS

1. Patient Enrollment

The programs had different goals for the number of beneficiaries that they planned to enroll, ranging from 678 to 3,036 enrollees, split equally between treatment and control groups (Table III.1).[10] Only three programs planned to reach full enrollment by the end of their first year, and to continue to recruit new patients only as replacements to patients lost through attrition. The remaining programs planned to continue enrolling new patients through their second year of operation, and to then gradually decrease the number of new patients recruited to include only replacements for participants who died or disenrolled.

Although only four programs met (or exceeded) their first-year enrollment targets, four additional programs were able to meet their second-year targets. Seven programs failed to reach either their first- or second-year enrollment targets. Four of the seven were able to enroll enough patients to meet about one-half of their target enrollments; the remaining three fell far short of their targets.

Three of the four programs that met both their first- and second-year enrollment targets used centralized patient databases to identify patients. They also enjoyed strong existing organizational links to and good relationships with the patients' physicians before the demonstration began, which likely led physicians to enthusiastically encourage patients to enroll.

[10] The one exception is CorSolutions. That organization's treatment group had two arms, one of which consisted of people who received prescription drugs if they did not already have coverage, and if had an income under 200 percent of the federal poverty level. The overall goal was to randomize 500 patients into the treatment group arm that received the drug benefit, 500 patients into the treatment group that did not receive the benefit, and 750 into the control group. Thus, the treatment-control ratio was 4:3.

One program also marketed the program to network physicians before the start of the demonstration; another had good relationships with area physicians and was considered a well-regarded disease management provider with long-standing ties to managed care plans in a service area with a high level of managed care penetration.

Four additional programs overcame low first-year enrollments to meet their second-year enrollment targets. Two of the four programs relied on patient lists supplied by the program host to identify patients but faced higher-than-expected patient refusal rates during their first year. One program also found that its eligibility criteria were overly restrictive; as a result, it expanded its guidelines to include patients who had chronic obstructive pulmonary disease (COPD). During its second year, that program also began recruiting patients from a third retirement community. A second program increased its outreach efforts to physicians and other providers, asking them to encourage their patients to enroll in the program. During its first year, the third program had recruited exclusively from its physician referral network, but it fell short of its enrollment target. To address this shortfall, during the second year, the program began identifying patients from lists generated by local hospitals; it met its second-year target. The fourth program met its second-year enrollment target without having to alter its patient identification and recruitment strategy; as it had done during its first year, it continued to review charts at two large geriatric group practices. The remaining programs did not meet their first- or second-year enrollment targets. Five of those programs identified patients using lists generated by health care providers or networks, whereas two relied on local physicians to provide referrals to the demonstration.

Program staff provided a variety of reasons to explain these enrollment shortfalls. Five programs cited higher-than-expected patient refusal rates. In addition, one of the five programs also indicated that it did not have enough staff resources to devote adequate time to enrollment;

another attributed its low enrollment to inadequate physician support for the program. In particular, a number of physicians with relatively larger practices in this program's service area were hostile because they had had negative experiences with the disease management vendor when it served as a managed care subcontractor; program staff reported that the vendor's focus on keeping the costs of chemotherapy agents low had irritated physicians who were used to greater "price flexibility."

The experiences of the 15 programs during their first 2 years of operations offer some lessons about patient recruitment. Physician support seemed to be the key to the programs' ability to meet their enrollment targets, rather than the particular technique used to recruit patients. Of the programs that obtained lists of prospective enrollees from hospitals or health care networks, four programs fell short of their enrollment targets and five programs met (or exceeded) theirs. Of the programs that recruited physicians to provide patient referrals, two met their enrollment targets and two fell short. Finally, two programs tried both methods of identifying and recruiting patients; one program met its second-year enrollment target and the other fell short.

Regardless of the recruitment strategy that they picked, the programs with strong, existing relationships with local health care providers were able to meet their enrollment targets. Six of the eight programs that met their first- or second-year enrollment targets had strong preexisting ties to participating physicians; only two of the seven programs that failed to reach either their first- or second year-enrollment targets had had any previous experience working with participating physicians. One of the programs that fell short of its first-year enrollment target increased its outreach efforts to physicians during its second year and was able to surpass its target. Compared with the programs that failed to meet their enrollment targets, the ones that were able to meet their targets also established relatively broad eligibility criteria and had a

larger pool of potential patients from which to recruit (Table III.2); however, all of the programs had sufficient numbers of potentially eligible participants.

Seven programs accepted patients who had only one target diagnosis. (Five programs recruited patients who had a primary diagnosis of congestive heart failure [CHF]; one program recruited patients who had a primary diagnosis of coronary artery disease [CAD]; and one program recruited only patients who had cancer.) Only two of the seven programs managed to meet their target enrollments. The eight remaining programs recruited patients with diverse primary conditions; of those, only two failed to meet their target enrollments. Staff from all 15 programs noted that recruiting patients took more staff time than expected, and that recruitment made it difficult for care coordinators to balance their workloads.

2. Patient Disenrollment

According to data that each program prepared for the evaluation, substantial numbers of patients who had enrolled during the first year of operations died or disenrolled during the first year after the month of randomization (Table III.2). The rates at which patients left each program because they died, lost their eligibility, or voluntarily chose to disenroll varied from slightly more than 2 percent to more than 45 percent. Eight programs had a combined death and disenrollment rate of between 10 and 20 percent. Three programs experienced patients leaving the program at a rate between 20 and 40 percent. Three programs reported that more than 40 percent of their patients left. The average length of enrollment during the first year after enrolling for all the programs was 10.8 months. Programs that had lower-than-average patient tenures experienced correspondingly higher rates of patient death or disenrollment.

Although the disenrollment rates were generally high, the reasons for leaving the various programs are varied, with few patients (3 percent overall) choosing to leave the

TABLE III.2

PROGRAM ENROLLMENTS AND DISENROLLMENTS
DURING THE 1ST YEAR OF OPERATIONS

Program	Treatment Group Members Enrolled (Number)	Average Length of Enrollment (Months)	Percentage Who Died or Disenrolled	Reason for Disenrollment (Percent)			
				Died	Voluntarily Disenrolled	Entered Nursing Home, Hospice, ESRD	Other[a]
Avera	157	9.9	28.7	9.6	8.3	7.0	3.9
Carle	1,151	11.4	10.2	3.5	2.4	0.1	4.2
CenVaNet	538	11.0	16.0	5.9	3.9	0.0	6.2
Charlestown	212	11.2	10.8	7.1	0.9	2.4	0.5
CorSolutions	366	8.7	43.4	12.3	1.9	7.4	21.9[b]
Georgetown	53	10.1	26.4	9.4	7.5	0.0	9.5
Health Quality Partners	243	11.9	2.5	0.8	1.2	0.0	0.4
Hospice of the Valley	236	9.0	41.9	20.8	5.5	10.2	5.5
Jewish Home and Hospital	271	10.7	19.9	6.3	9.6	2.2	1.9
Medical Care Development	196	9.8	36.7	15.8	2.0	4.1	14.7[c]
Mercy	317	11.2	12.9	5.4	0.6	6.6	0.3
QMed	698	11.1	12.5	1.1	2.9	0.7	7.8
Quality Oncology	31	8.9	45.2	25.8	3.2	3.2	12.9[d]
University of Maryland	29	11.2	20.7	17.2	3.4	0.0	0.0
Washington University	715	11.0	14.8	10.5	1.4	0.0	2.9
All Programs	**5,210**	**10.8**	**17.8**	**7.0**	**3.0**	**2.1**	**3.9**

Source: Data reported by individual programs to Mathematica Policy Research, Inc.

[a]"Other" includes patients who disenrolled because they relocated, lost program eligibility, completed the program, were uncooperative, or had physicians who left the program, or who disenrolled for "other" reasons, as reported by the programs.

[b]"Other" for CorSolutions includes disenrollments for the following reasons: 4.4 percent due to relocation, 3.6 percent due to loss of program eligibility, 1.1 percent who were uncooperative, 1.6 percent whose physicians left the program, and 11.2 percent for "other" reasons.

[c]"Other" for Medical Care Development includes disenrollments for the following reasons: 1.5 percent due to relocation, 5.6 percent because they completed the program, 6.6 percent who were uncooperative, and 1.0 percent for "other" reasons.

[d]"Other" for Quality Oncology includes disenrollments for the following reasons: 3.2 percent due to relocation, 3.2 percent who lost program eligibility, and 6.5 percent who were uncooperative.

ESRD = end-stage renal disease.

demonstration. In most programs, death accounted for about one-half of patient attrition (Table III.2). An exception was the University of Maryland, in which nearly all of the patients leaving did so because of death. At the other extreme was QMed, the only program that required permission from patients' physicians to participate and from which very few patients left due to death; rather, the most often cited reason for patient disenrollment was "physician left program" (Table III.2). The percentage of patients voluntarily disenrolling from the demonstration at each site ranged from fewer than 1 percent to slightly fewer than 10 percent. Few patients lost their eligibility because they entered hospice care, a nursing home, or developed end-stage renal disease (ESRD). Several programs also indicated that patient relocation was another contributor to the high disenrollment rates (not shown; combined with "Other" in Table III.2).

3. Patients' Characteristics

The 9,617 patients enrolled in the 15 demonstration programs through the first year of each program's operations were somewhat older than the 42 million Medicare beneficiaries nationally. This finding is not surprising, given that older beneficiaries are more likely to have chronic illnesses, and that nine programs chose to exclude beneficiaries who were younger than age 65. In 2002, 14 percent of Medicare beneficiaries were younger than age 65, and about 12 percent were aged 85 or older (Centers for Medicare & Medicaid Services 2002b). Overall, fewer than 7 percent of patients enrolled in the demonstration were younger than age 65; 14 percent were aged 85 or older (Table III.3). Although the age distribution across programs varied widely, the majority of patients enrolling in the demonstration during its first year were aged 65 to 84. However, six programs did not have any patients younger than age 65, and three had fewer than 5 percent who were that young. In contrast, more than one-fourth of the patients of one program were younger than age 65, far more than among beneficiaries nationally.

TABLE III.3

DEMONSTRATION PARTICIPANTS' AGES AND OTHER DEMOGRAPHIC CHARACTERISTICS
(Percentages Unless Otherwise Noted)

		Age				Race		Education			Income		
Program	Number	≤64	65 to 74	75 to 84	≥85	Nonwhite	Hispanic	<High School	High School/ GED	>High School	<$20,000	$20,000 to $40,000	>$40,000
Avera	315	0.0	30.8	48.6	20.6	1.0	0.3	34.4	39.4	26.3	53.8	34.5	11.8
Carle	2,042	0.9	46.9	40.1	12.1	2.9	0.6	14.4	35.9	49.7	28.0	40.8	31.3
CenVaNet	1,021	0.0	39.5	47.8	12.7	17.5	0.9	26.2	28.4	45.4	39.3	33.1	27.6
Charlestown	384	0.0	5.2	49.0	45.8	1.6	0.5	10.5	30.9	58.6	25.0	36.8	38.3
CorSolutions	619	12.6	39.6	34.1	13.7	33.1	10.5	37.2	25.3	37.5	66.2	19.0	14.8
Georgetown	104	0.0	31.7	54.8	13.5	52.9	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Health Quality Partners	438	0.0	48.4	43.4	8.2	1.4	0.9	10.7	31.5	57.9	26.7	34.5	38.8
Hospice of the Valley	432	0.0	27.3	46.8	25.9	4.2	5.2	18.4	33.9	57.7	44.8	35.0	20.2
Jewish Home and Hospital	505	0.2	21.0	42.8	36.0	46.9	22.5	33.3	25.6	41.1	65.2	20.0	14.9
Medical Care Development	384	7.3	45.3	34.4	13.0	0.8	0.2	32.2	37.3	30.5	61.4	27.2	11.3
Mercy	612	4.1	32.4	46.1	17.5	0.3	0.6	30.7	43.3	26.1	59.6	34.4	6.0
QMed	1,258	6.8	46.8	41.3	5.1	10.5	6.3	20.0	31.0	49.1	34.1	36.1	29.9
Quality Oncology	60	10.0	43.3	38.3	8.3	13.3	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
University of Maryland	55	14.5	43.6	38.2	3.6	43.6	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Washington University	1,388	28.0	34.8	27.9	9.4	37.6	1.6	25.6	27.4	47.0	54.0	26.6	19.4
Demonstration Total	**9,617**	**6.8**	**38.8**	**40.3**	**14.1**	**15.2**	**3.9**	**23.4**	**32.4**	**44.3**	**45.2**	**32.0**	**22.8**
Medicare Total	**41,887,000**	**14.1**	**43.1**	**31.2**	**11.7**	**20.8**	**7.3**	**31.0**	**30.3**	**38.7**	**50.1**	**32.1**	**17.8**

Source: Data provided by individual programs and patient survey conducted by Mathematica Policy Research, Inc.

GED = General Educational Development credential; n.a. = not available.

Twenty percent or more of the enrolled patients of four programs were aged 85 or older; in one of the four programs, nearly one-half the patients were in that age range.

Patients enrolling in the demonstration were generally more likely to be white, to be better educated, and to have slightly higher incomes than Medicare beneficiaries nationally. However, wide variations in patient demographics and patients' characteristics were observed across the programs. Overall, about 15 percent of the demonstration's participants were nonwhite, compared with 20 percent nationally; according to patient survey data, slightly fewer than 5 percent of patients identified themselves as Hispanic, whereas 7 percent of Medicare patients nationally were Hispanic (Centers for Medicare & Medicaid Services 2002b). Across the 15 programs, the ethnic composition of demonstration participants varied. One-third to one-half of the patients in five programs that targeted urban areas were non-white. In one of the five programs, 22 percent of enrollees were Hispanic, making this program the most ethnically diverse one.

Twenty-three percent of demonstration participants reported having less than a high school education, compared with slightly more than 30 percent among Medicare beneficiaries nationally (Table III.3). Forty-four percent of demonstration patients had completed some education beyond high school, compared with 39 percent of all Medicare beneficiaries. However, as with race, large variations in patients' educational attainment were observed across the 15 programs. Roughly 50 percent or more of the patients in five programs had education beyond the high school level; in three of the five programs, fewer than 15 percent of patients enrolled had less than a high school education, far less than the national Medicare average.

Demonstration participants also had slightly higher incomes than did Medicare beneficiaries, nationally. In addition, however, self-reported income varied across the 15 programs. Twenty-three percent of demonstration participants reported having annual incomes greater than $40,000,

compared with only 18 percent of Medicare beneficiaries nationwide (Table III.3). Similarly, only 45 percent of demonstration patients reported having an income of less than $20,000, compared with 50 percent of all Medicare beneficiaries. Three programs did enroll substantial numbers of participants with low incomes; the majority of patients enrolled by those programs had annual incomes of less than $20,000. In contrast, roughly one-third or more of the participants enrolled by four programs had annual incomes that were higher than $40,000. Those programs also enrolled the highest percentages of patients with education beyond the high school level, and, in general, the patients were healthier than the patients in the 11 other programs.

As expected, based on the numbers and types of medical conditions for which they were treated during the 2 years preceding randomization, patients enrolling in the 15 programs were sicker than the average Medicare beneficiary (Table III.4). Heart disease was the most commonly treated condition; more than 90 percent of the patients enrolled in four programs had been treated for CHF. In five other programs, about 50 to 75 percent of patients had been treated for CHF during the 2 years before randomization. More than three-fourths of the patients enrolling in five programs had been treated for CAD during the 2 years prior to random assignment. These figures also reflect the eligibility criteria that the programs imposed, as well as the programs' choices about which diagnoses to target. Of the 15 programs, 13 enrolled patients with CHF, and 4 focused primarily on patients with that diagnosis. In contrast, only about 45 percent of the Medicare population as a whole reports a diagnosis for heart disease, which includes both CHF and CAD (Centers for Medicare & Medicaid Services 2002a). Because most of the programs that targeted CHF also included a recent hospitalization as a part of their eligibility criteria, programs enrolling primarily patients with a primary diagnosis of CHF generally also enrolled patients with higher average numbers of annualized hospitalizations.

TABLE III.4

DEMONSTRATION PARTICIPANTS, BY MEDICARE STATUS AND DIAGNOSIS
(Percentages Unless Otherwise Noted)

Program	Number	Medicare Buy-In	Disabled or ESRD	Diagnosis[a]						
				CAD	CHF	Stroke	Diabetes	Cancer	COPD	Dementia
Avera	315	7.9	9.5	79.0	99.0	26.7	43.5	25.4	49.2	3.5
Carle	2,042	4.3	7.9	51.1	27.5	16.1	39.8	22.1	22.4	5.7
CenVaNet	1,021	5.8	8.4	70.0	54.8	28.6	55.7	28.4	31.3	5.8
Charlestown	384	0.0	2.3	61.5	49.0	38.0	27.6	31.0	38.3	8.9
CorSolutions	619	23.7	23.6	87.2	97.4	41.0	54.3	18.7	49.6	11.8
Georgetown	104	14.4	12.5	83.7	96.2	28.8	51.9	30.8	32.7	14.4
Health Quality Partners	438	1.6	2.7	33.1	9.1	16.9	23.7	24.7	8.7	2.5
Hospice of the Valley	432	16.9	13.9	63.2	59.3	37.0	33.3	29.4	54.9	25.0
Jewish Home and Hospital	505	33.7	10.1	48.9	34.7	28.1	39.0	28.5	23.8	35.4
Medical Care Development	384	21.4	20.8	90.4	77.6	23.4	48.4	20.8	40.6	4.9
Mercy	612	10.5	17.8	68.6	65.2	30.7	36.1	26.0	53.8	8.5
QMed	1,258	10.9	15.7	46.1	17.6	15.7	26.7	20.7	13.7	1.6
Quality Oncology	60	13.3	16.7	55.0	16.7	20.0	26.7	100.0	40.0	8.3
University of Maryland	55	20.0	30.9	87.3	94.5	32.7	49.1	9.1	43.6	10.9
Washington University	1,388	19.4	42.9	61.5	46.4	26.9	46.5	39.2	36.9	12.2
Demonstration Total	**9,617**	**12.0**	**16.4**	**60.5**	**46.0**	**24.8**	**40.5**	**26.8**	**9.1**	**9.1**
Medicare Total	**41,887,000**	**12.5**	**14.1**	**45.4**[b]	**45.4**[b]	**11.7**	**19.7**	**16.6**[c]	**15.0**	**5.0**[d]

Source: Medicare National Claims History File, Standard Analytic File, and Enrollment Databases, Medicare Current Beneficiary Survey 2002.

[a]Medical conditions treated during the 2 years before randomization, as reported in Medicare claims data.

[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]Does not include 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.

Most programs also enrolled substantial percentages of patients who had been treated for stroke, COPD, or diabetes; smaller percentages of patients had been treated for cancer or dementia. One-third to one-half of the patients of 11 programs had been treated for diabetes, and the pattern is similar for COPD (Table III.4). All of the patients of one program (Quality Oncology) had been treated for cancer during the 2 years before random assignment, which reflects the program's target diagnosis. Among the 15 programs, Health Quality Partners enrolled the healthiest group of enrollees.

The programs that enrolled the highest percentages of patients with self-reported annual incomes of more than $40,000 also had the fewest percentages of patients who qualified for Medicare buy-in (Table III.4). During their first year, two programs enrolled few to no patients who qualified for state assistance for Medicare buy-in. In contrast, more than one-third of the patients of another program received state assistance for their Medicare premiums, and about 20 percent or more of the patients enrolling in four other programs qualified for Medicare buy-in programs. Nationally, the percentage of Medicare beneficiaries who are dually eligible is 12.5 percent.[11] The percentage of patients in seven programs who were originally entitled to Medicare due to a disability or to ESRD was higher than the national average of 14 percent. One program enrolled the lowest percent of patients with a disability or ESRD, with only about 2 percent of its patients being eligible through those means. In contrast, nearly 40 percent of the

[11] "State buy-in" is a proxy for whether a beneficiary is also enrolled in Medicaid, as state Medicaid programs typically pay the Medicare Part B premium for their Medicaid enrollees who are also eligible for Medicare. However, some beneficiaries for whom the state buys in (those in the Qualified Medicare Beneficiary Program or the Specified Low-Income Medicare Beneficiary Program) do not have full Medicaid coverage, and some states do not buy in for some Medicare beneficiaries who have full Medicaid coverage, depending on their type of eligibility (for example, for those who are eligible for Medicaid due to spending down their assets).

patients enrolled by another program were eligible for Medicare because of a disability or a diagnosis of ESRD.[12]

C. PARTICIPATING PATIENTS AND ELIGIBLE NONPARTICIPATING BENEFICIARIES

Medicare claims and eligibility data were used to estimate both the number of Medicare beneficiaries in each program's service area who *potentially* were eligible for the program and the percentage of those beneficiaries who actually participated. Beneficiaries were identified as potentially eligible for a particular program if, for any month during the program's first 6 months of operations, they (1) lived in the program's catchment area, (2) were enrolled in Medicare Parts A and B, (3) had Medicare as their primary payer, (4) were not in a Medicare managed care plan (Medicare+Choice or Medicare Advantage plan), (5) met the program's target diagnosis and utilization requirements that could be simulated using Medicare claims data, and (6) did not have any of the program's exclusion criteria measured in the Medicare claims data (for example, ESRD or terminal cancer).

This definition of potential eligibility is inexact for many of the programs, however, as they imposed additional restrictions at intake that the evaluator could not take into account when trying to identify eligible beneficiaries by using claims data (such as having a telephone, having at least a fourth grade reading level, or exceeding a specified disease-severity threshold). Furthermore, the proportion of patients who actually participated was influenced heavily by both the scope and intensity of the programs' recruiting efforts and referral sources. For example, because many programs relied heavily on their own data systems or on those of a few affiliated

[12] Because some programs excluded beneficiaries with ESRD, and nationally very few beneficiaries are entitled to Medicare due to ESRD, most of these beneficiaries originally qualified for Medicare because of a disability. According to the elderly, disabled, and ESRD data from CMS in 2002, only 0.6 percent of the national Medicare population qualified for benefits due to ESRD (Centers for Medicare & Medicaid Services 2002a).

hospitals or physician groups to identify potentially eligible patients, they were not be able to identify or enroll many of the beneficiaries in the catchment area that the evaluator had identified from the claims data as being potentially eligible. Nonetheless, the proportions are useful as a rough gauge of program penetration among Medicare beneficiaries with specific illnesses. To lessen possible confusion, this report will refer to these potentially eligible beneficiaries as "comparable beneficiaries (or nonparticipants) by claims data" from this point on.[13]

The evaluator also used Medicare claims and enrollment data to assess whether the programs enrolled a mix of beneficiaries representative of the larger pool of comparable beneficiaries by claims data. The evaluator conducted that analysis by comparing the demographic characteristics, diagnoses, and utilization histories of comparable nonparticipants by claims data with those of participants who enrolled during the first year of each program's operations. The analysis compared service use and cost measures for the 12-month period preceding enrollment for the enrollees and the service use and cost measures for an analogous period for the comparable nonparticipants by claims data (the 12 months beginning 9 months before program startup and ending 3 months after startup). In addition, the evaluator compared the average costs for participants with its projected average costs for the target population that were presented in the OMB waiver package for the demonstration.

That simulation shows that the programs' pools of comparable nonparticipants by claims data during the first 6 months of operations ranged in size from about 1,800 to more than 125,000 (Table III.5). Participation rates (the number of beneficiaries enrolled during the first 2 years divided by the number of comparable nonparticipants by claims data; right-hand column of the table) varied from fewer than 1 percent to more than 15 percent. The overall participation

[13] This terminology differs from that used in previous evaluation reports on the demonstration, which used the term "eligible beneficiaries" or "eligible nonparticipants."

TABLE III.5

NUMBER OF PARTICIPANTS AND COMPARABLE NONPARTICIPANTS BY CLAIMS DATA

Program	Beneficiaries Enrolled	Comparable Nonparticipants	"Participation" Rate (Percent)
Avera	624	5,505	11.3
Carle	2,642	23,284	11.3
CenVaNet	1,305	55,262	2.4
Charlestown	802	38,745	2.1
CorSolutions	2,162	13,119	16.5
Georgetown	199	5,122	3.9
Health Quality Partners	1,140	60,740	1.9
Hospice of the Valley	814	85,293	1.0
Jewish Home and Hospital	766	125,821	0.6
Medical Care Development	876	10,655	8.2
Mercy	865	11,322	7.6
QMed	1,454	13,148	11.1
Quality Oncology	141	1,840	7.7
University of Maryland	137	2,398	5.7
Washington University	2,038	117,322	1.7
Total	**15,965**	**569,576**	**2.8**

Source: Enrollment data were provided by the programs. Data on the number of comparable nonparticipants are from cost waiver calculations produced by Mathematica Policy Research, Inc.

Note: "Comparable Nonparticipants by Claims Data" refers to a group of Medicare beneficiaries identified in each program's service area as *potentially* eligible for the program as measured by Medicare claims data. Comparable nonparticipants were beneficiaries who, for any month during the program's first 6 months of operation (1) lived in the program's catchment area, (2) were enrolled in Medicare Parts A and B, (3) had Medicare as their primary payer, (4) were not in a Medicare managed care plan (Medicare+Choice or Medicare Advantage plan), (5) met the program's target diagnosis and utilization requirements that could be simulated using Medicare claims data, and (6) did not have any of the program's exclusion criteria that could be measured in the Medicare claims data (for example, end-stage renal disease or terminal cancer). This definition of comparable nonparticipants is inexact, as many of the programs imposed additional restrictions at intake that the evaluator could not take into account when trying to identify such beneficiaries through claims data (such as having a telephone, having at least a fourth grade reading level, or exceeding a specified disease severity threshold). Charlestown, for example, recruited only beneficiaries residing in one of its three residential communities, whereas the comparable nonparticipants in the table were drawn from all beneficiaries in the zip codes of the three Charlestown communities. Nevertheless, the proportions in the table are useful as a rough gauge of program penetration among Medicare beneficiaries with specific illnesses. (Note that the "comparable nonparticipants by claims data" terminology differs from terminology in previous evaluation reports on the demonstration, which used the term "eligible nonparticipants.")

rate was 2.8 percent. These rates do not imply that few people were interested in the programs. Many people probably were unaware of them, others may have failed to meet additional eligibility criteria beyond those that can be simulated with claims data (such as a minimum severity of illness threshold), and others may eventually enroll during the remaining years of the demonstration. As stated, some programs already have reached their enrollment targets, and others are continuing to enroll additional patients. The estimates simply give an indication of the number of Medicare beneficiaries who live in program service areas and have the target diagnoses of each program.

Comparisons of actual program patients with the comparable nonparticipants by claims data show some differences in characteristics (Table III.6). All but three programs enrolled a smaller percentage of very elderly beneficiaries (those aged 85 or older) than were in the group of comparable nonparticipants by claims. Among the 15 programs, those 3 also enrolled the highest proportions of participants in the group of patients aged 85 or older; 1 of the 3 recruited exclusively from 3 retirement communities, and another targeted frail beneficiaries from 2 large geriatric practices, who were older on average than the general Medicare population.

As noted, most programs enrolled relatively few beneficiaries who were dually eligible for Medicaid and Medicare, with the rate of dually eligible participants in 10 of the 15 programs falling below the rate of dually eligible nonparticipants in the programs' target areas (see the column labeled "State Buy-In" in Table III.6). The proportion of demonstration participants who were dually eligible ranged from 0 to 33 percent, with more than one-half of the programs falling under 15 percent. This finding is consistent with the fact that most of the programs enrolled patients who were wealthier than the average Medicare beneficiary nationally.

In general, the patients enrolling in the demonstration were very sick, which is consistent with the programs' recruitment criteria. Only four programs enrolled a substantial percentage of

TABLE III.6

COMPARISON OF PARTICIPANTS AND COMPARABLE NONPARTICIPANTS BY CLAIMS DATA

Program	Number	Age (Percent) <65	Age (Percent) 65 to 84	Age (Percent) >85	Total Monthly Medicare Payments	Total Monthly Payment: Cost Estimate Ratio	State Buy-In (Percent)	Average Annualized Number of Hospitalizations	Average Annualized Hospitalizations in Past 2 Years: 0	≤1	1 to 2	≥2
Avera												
Participants	618	0.0	79.4	20.6	$1,615	1.09	7.9	1.8	1.9	44.4	29.8	23.8
Nonparticipants	5,505	0.0	63.1	37.0	$1,376	0.93	19.6		1.4	53.2	28.5	16.9
Carle												
Participants	2,371	0.9	87.0	12.1	$521	0.70	4.3	0.5	56.9	32.2	7.9	3.0
Nonparticipants	23,284	7.9	75.5	16.6	$625	0.84	13.6		49.4	36.0	9.9	4.7
CenVaNet												
Participants	1,256	0.0	87.3	12.7	$953	0.76	5.8	0.8	40.2	38.5	12.4	8.9
Nonparticipants	55,262	0.0	85.5	14.2	$507	0.41	10.1		61.4	29.2	7.0	2.5
Charlestown												
Participants	742	0.0	54.2	45.8	$1,159	0.78	0.0	0.8	32.3	43.7	17.7	6.2
Nonparticipants	38,745	0.0	81.3	18.7	$1,112	0.75	14.3		25.6	52.0	15.5	7.0
CorSolutions												
Participants	2,102	12.6	73.7	13.7	$2,644	1.27	23.7	2.0	7.9	35.5	25.4	31.2
Nonparticipants	13,119	11.9	71.3	16.8	$1,942	0.93	25.5		16.1	38.6	25.8	19.5
Georgetown												
Participants	197	0.0	86.5	13.5	$2,530	0.73	14.4	2.2	5.8	34.6	25.0	34.6
Nonparticipants	5,122	0.0	76.6	23.4	$2,410	0.69	13.8		2.1	53.4	26.0	18.4
Health Quality Partners												
Participants	999	0.0	91.8	8.2	$414	0.64	1.6	0.3	72.1	18.9	6.6	2.3
Nonparticipants	60,740	0.0	87.5	12.6	$355	0.55	5.5		74.2	19.7	4.6	1.6
Hospice of the Valley												
Participants	747	0.0	74.1	25.9	$2,174	2.12	16.9	1.5	15.5	45.6	19.4	19.4
Nonparticipants	85,293	0.0	82.5	17.4	$965	0.94	8.9		40.1	44.4	11.2	4.4

TABLE III.6 *(continued)*

Program	Number	Age (Percent) <65	Age (Percent) 65 to 84	Age (Percent) >85	Total Monthly Medicare Payments	Total Monthly Payment: Cost Estimate Ratio	State Buy-In (Percent)	Average Annualized Number of Hospitalizations	Average Annualized Hospitalizations in Past 2 Years: 0	≤1	1 to 2	≥2
Jewish Home and Hospital												
Participants	720	0.2	63.8	36.0	$1,450	0.92	33.7	0.7	44.0	35.2	12.3	8.5
Nonparticipants	125,821	0.0	80.3	19.7	$982	0.62	24.2		59.7	28.6	7.7	4.0
Medical Care Development												
Participants	869	7.3	79.7	13.0	$1,718	0.72	21.4	1.5	3.6	53.4	24.2	18.7
Nonparticipants	10,655	7.7	66.1	26.2	$1,274	0.53	26.7		16.1	49.2	21.9	12.9
Mercy												
Participants	848	4.1	78.4	17.5	$1,315	1.03	10.5	1.1	10.8	58.8	22.1	8.3
Nonparticipants	11,322	5.1	71.3	23.6	$610	0.48	16.7		47.6	38.5	10.0	4.0
QMed												
Participants	1,293	6.8	88.2	5.1	$507	0.45	10.9	0.3	73.4	20.0	4.2	2.3
Nonparticipants	13,148	0.0	85.4	14.7	$954	0.85	34.6		57.9	28.7	9.7	3.7
Quality Oncology												
Participants	129	10.0	81.7	8.3	$2,885	0.79	13.3	0.8	36.7	40.0	10.0	13.3
Nonparticipants	1,840	0.0	85.0	15.0	$2,463	0.67	6.4		36.3	41.9	14.0	7.9
University of Maryland												
Participants	134	14.5	81.8	3.6	$3,299	1.11	20.0	2.0	5.5	32.7	32.7	29.1
Nonparticipants	2,398	15.2	67.3	17.5	$3,156	1.06	20.2		2.0	34.9	29.7	33.5
Washington University												
Participants	1,961	28.0	62.7	9.4	$2,263	2.49	19.4	1.6	17.5	35.1	24.6	22.9
Nonparticipants	117,322	12.9	72.8	14.3	$787	0.87	10.8		56.2	30.1	8.7	4.9

Source: Medicare Enrollment Database and National Claims History File.

Notes: The intake date used in this table is the date of enrollment for participants. For comparable nonparticipants, the intake date is the midpoint of the 6-month enrollment period examined.

"Comparable Nonparticipants by Claims Data" refers to a group of Medicare beneficiaries identified in each program's service area as *potentially* eligible for the program as measured by Medicare claims data. Comparable nonparticipants were beneficiaries who, for any month during the

TABLE III.6 *(continued)*

program's first 6 months of operation (1) lived in the program's catchment area, (2) were enrolled in Medicare Parts A and B, (3) had Medicare as their primary payer, (4) were not in a Medicare managed care plan (Medicare+Choice or Medicare Advantage plan), (5) met the program's target diagnosis and utilization requirements that could be simulated using Medicare claims data, and (6) did not have any of the program's exclusion criteria that could be measured in the Medicare claims data (for example, end-stage renal disease or terminal cancer). This definition of comparable nonparticipants is inexact, as many of the programs imposed additional restrictions at intake that the evaluator could not take into account when trying to identify such beneficiaries through claims data (such as having a telephone, having at least a fourth grade reading level, or exceeding a specified disease severity threshold). Charlestown, for example, recruited only beneficiaries residing in one of its three residential communities, whereas the comparable nonparticipants in the table were drawn from all beneficiaries in the zip codes of the three Charlestown communities. Nevertheless, the proportions in the table are useful as a rough gauge of program penetration among Medicare beneficiaries with specific illnesses. (Note that the "comparable nonparticipants by claims data" terminology differs from terminology in previous evaluation reports on the demonstration, which used the term "eligible nonparticipants.")

patients who had not had any hospitalizations during the 2 years preceding randomization. Nearly three-fourths of the patients enrolled in two of the four programs had no hospitalizations during that 2-year period. Five other programs enrolled a high proportion of patients (about 30 percent or more) who had had two or more hospitalizations during the 2 years preceding randomization, indicating that they were enrolling very sick patients. The annualized number of hospitalizations was similar between the program participants and comparable nonparticipants by claims data in nearly all the programs. Notably, two programs enrolled patients who were substantially sicker than were the programs' comparable nonparticipants, as indicated by the patients' higher rate of hospitalizations and higher preenrollment Medicare expenditures.[14]

Because hospitalizations account for the bulk of Medicare expenses, it is not surprising that Medicare payments for participants during the year preceding enrollment was high. Most of the patients enrolled in the demonstration had significantly higher preenrollment monthly Medicare costs than did the average noninstitutionalized Medicare beneficiary (Table III.6). Programs that enrolled patients with the highest preenrollment total monthly Medicare costs also had among the highest rates of disenrollment due to death during the first year of operations, as well as the highest average number of hospitalizations during the 2 years preceding random assignment. Three of the four programs that lost more than 15 percent of their patients due to death within the first year enrolled patients with average monthly Medicare costs of $2,000 or higher, much higher than the Medicare average of $514 (Centers for Medicare & Medicaid Services 2002a). The fourth program enrolled patients with average monthly preenrollment costs of about $1,700, also much higher than the national average. Three programs enrolled patients whose monthly

[14] Because Washington University used a proprietary algorithm developed by StatusOne to identify patients, the evaluator did not have access to the full eligibility criteria used by that program to identify potential enrollees. Thus, the simulation of eligible nonparticipants is likely to be less accurate than those for the other programs.

expenditures were close to or lower than the national average. Those programs also enrolled fewer patients who were older than 85 years, had lower mortality rates for their patients during the first year of program operations, and enrolled patients with fewer average hospitalizations during the 2 years preceding enrollment.

The average monthly Medicare payments for program participants ranged from a low of $414 to a high of $3,299 (Table III.6). Three programs enrolled patients who had monthly payments of $600 or less (less than $7,000 per year), six programs enrolled patients with monthly payments between about $1,000 and $1,700 ($12,000 to about $20,000 per year), and six programs enrolled patients with extremely high payments of $2,000 per month ($24,000 per year) or more. Two programs enrolled patients whose monthly payments were substantially higher than the monthly payments that had been estimated by the evaluator for the waiver calculations. Six programs enrolled patients whose monthly reimbursements were about 75 percent or less than those predicted by the waiver cost estimates. Three programs enrolled patients whose monthly payments were very close to the estimates.

D. COMPARISON OF TREATMENT AND CONTROL GROUP MEMBERS

As expected under random assignment, with a few minor exceptions, the treatment group and the control group had similar preenrollment characteristics (Table III.7). Overall, the number of statistically significant differences observed between the treatment and control groups was small and what one would expect to occur by chance. In terms of medical conditions treated during the 2 years before randomization, nine scattered differences between the treatment and control groups were statistically significant, exactly what one would expect to occur by chance with 90 separate comparisons (15 sites x 6 conditions), at the 10-percent significance level. The average number of hospitalizations during the 2 years preceding random assignment was consistent between treatment and control groups for all of the programs except the University of

TABLE III.7

PREENROLLMENT CHARACTERISTICS OF TREATMENT AND CONTROL GROUPS RANDOMIZED DURING THE 1ST YEAR OF PROGRAM OPERATIONS

Program	Number	Age (Percent)			Diagnosis (Percent)						Average Annualized Number of Hospitalizations	Average Total Medicare Cost
		<65	65 to 84	>85	CAD	CHF	Stroke	Diabetes	Cancer	COPD		
Avera												
Treatment	157	0.0	77.1	22.9	82.8	100.0	27.4	49.0	26.1	45.9	1.75	$1,537
Control	158	0.0	81.6	18.4	75.3	98.1	25.9	38.0	24.7	52.5	1.85	$1,693
Difference		0.0	-4.5	4.6	7.5	1.9*	1.4	11.1**	1.4	-6.7	-0.10	-$156
Carle												
Treatment	1,024	1.0	87.3	11.7	51.2	29.0	16.6	39.5	24.0	23.3	0.50	$536
Control	1,018	0.9	86.6	12.5	51.0	26.0	15.5	40.1	20.1	21.4	0.43	$505
Difference		0.1	0.7	-0.8	0.2	3.0	1.1	-0.6	3.9**	1.9	0.06	$30
CenVaNet												
Treatment	512	0.0	87.3	12.7	70.5	56.1	27.5	54.5	29.1	32.0	0.79	$1,018
Control	509	0.0	87.2	12.8	69.5	53.4	29.7	57.0	27.7	30.6	0.74	$888
Difference		0.0	0.1	-0.1	1.0	2.6	-2.1	-2.5	1.4	1.4	0.05	$131
Charlestown												
Treatment	195	0.0	50.8	49.2	60.5	52.8	36.9	26.2	31.8	35.9	0.74	$1,134
Control	189	0.0	57.7	42.3	62.4	45.0	39.2	29.1	30.2	40.7	0.86	$1,184
Difference		0.0	6.9	6.9	-1.9	7.8	-2.2	-2.9	1.6	-4.8	-0.12	-$51
CorSolutions[a]												
Treatment	354	13.8	70.6	15.5	88.7	98.6	38.7	53.1	19.5	50.8	1.91	$2,468
Control	265	10.9	77.7	11.3	85.3	95.8	44.2	55.8	17.7	47.9	2.15	$2,879
Difference		2.9	-7.1	4.2	3.4	2.7**	-5.5	-2.7	1.8	2.9	-0.24	-$411*
Georgetown												
Treatment	51	0.0	94.1	5.9	84.3	94.1	23.5	56.9	27.5	29.4	1.82	$1,872
Control	53	0.0	79.2	20.8	83.0	98.1	34.0	47.2	34.0	35.8	2.52	$3,164
Difference		0.0	14.9	-14.9	1.3	-4.0	-10.4	9.7	-6.5	-6.4	-0.70	-$1,292***
Health Quality Partners												
Treatment	219	0.0	91.3	8.7	31.1	9.6	20.1	23.7	24.7	10.0	0.29	$373
Control	219	0.0	92.2	7.8	35.2	8.7	13.7	23.7	24.7	7.3	0.30	$455
Difference		0.0	0.9	0.9	-4.1	0.9	6.4*	0.0	0.0	2.7	-0.01	-$81

TABLE III.7 *(continued)*

Program	Number	Age (Percent) <65	65 to 84	>85	Diagnosis (Percent) CAD	CHF	Stroke	Diabetes	Cancer	COPD	Average Annualized Number of Hospitalizations	Average Total Medicare Cost
Hospice of the Valley												
Treatment	222	0.0	71.6	28.4	62.6	57.7	40.1	35.1	27.5	52.3	1.43	$2,276
Control	210	0.0	76.7	23.3	63.8	61.0	33.8	31.4	31.4	57.6	1.52	$2,066
Difference		0.0	-5.0	5.0	-1.2	-3.3	6.3	3.7	-4.0	-5.4	-0.09	$210
Jewish Home and Hospital												
Treatment	253	0.4	64.0	35.6	51.4	38.3	31.2	42.7	27.7	29.6	0.72	$1,523
Control	252	0.0	63.5	36.5	46.4	31.0	25.0	35.3	29.4	17.9	0.76	$1,377
Difference		0.4	0.5	-0.9	5.0	7.4*	6.2	7.4*	-1.7	11.8***	-0.04	$146
Medical Care Development												
Treatment	192	10.4	75.5	14.1	89.6	78.1	22.9	46.9	18.2	45.3	1.51	$1,766
Control	192	4.2	83.9	12.0	91.1	77.1	24.0	50.0	23.4	35.9	1.46	$1,671
Difference		6.3**	-8.3	2.1	-1.6	1.0	-1.0	-3.1	-5.2	9.4	0.05	$95
Mercy												
Treatment	304	3.9	80.6	15.5	68.4	64.8	29.3	34.2	27.3	56.9	1.11	$1,376
Control	308	4.2	76.3	19.5	68.8	65.6	32.1	38.0	24.7	50.6	1.01	$1,256
Difference		-0.3	4.3	-4.0	-0.4	-0.8	-2.9	-3.8	2.6	6.3	0.10	$120
QMed												
Treatment	633	6.6	88.8	4.6	46.6	16.9	14.8	26.5	22.1	14.8	0.29	$544
Control	625	6.9	87.5	5.6	45.6	18.2	16.6	26.9	19.4	12.5	0.26	$470
Difference		-0.2	1.3	-1.0	1.0	-1.3	-1.8	-0.3	2.8	2.4	0.03	$74
Quality Oncology												
Treatment	29	10.3	79.3	10.3	62.1	17.2	24.1	27.6	100.0	41.4	0.86	$3,072
Control	31	9.7	83.9	6.5	48.4	16.1	16.1	25.8	100.0	38.7	0.82	$2,710
Difference		0.7	-4.6	3.9	13.7	1.1	8.0	1.8	0.0	2.7	0.04	$362
University of Maryland												
Treatment	29	20.7	75.9	3.4	96.6	96.6	34.5	55.2	10.3	41.4	2.51	$3,711
Control	26	7.7	88.5	3.8	76.9	92.3	30.8	42.3	7.7	46.2	1.52	$2,840
Difference		13.0	-12.6	-0.4	19.6**	4.2	3.7	12.9	2.7	-4.8	0.98*	$871
Washington University												
Treatment	698	27.8	61.9	10.3	62.2	47.4	28.1	46.3	39.3	37.2	1.57	$2,267
Control	690	28.1	63.5	8.4	60.7	45.4	25.7	46.8	39.1	36.5	1.54	$2,259
Difference		-0.3	1.6	1.9	1.5	2.1	2.4	-0.5	0.1	0.7	0.03	$8

TABLE III.7 *(continued)*

Source: Medicare claims data.

[a]CorSolutions had two treatment group arms, one that received prescription drugs if the group's members did not already have coverage, and one that received the regular care coordination intervention without the supplemental prescription drug benefit. Its overall goal was to randomize into the treatment group 500 patients who did not receive the prescription drug benefit, and to randomize into the control group 750 patients who did not receive the benefit. Thus, the treatment-control ratio was 10:7.5. The treatment-control ratio for the 14 other programs was 1:1.

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

Maryland, one of the two smallest sites. Finally, for 13 out of the 15 programs, average Medicare expenditures for the treatment group during the 2 years preceding enrollment were statistically indistinguishable from those for the control group. In the third smallest program, which had only 51 treatment group members, the treatment group's average preenrollment expenditures were 40 percent below those of the control group. The evaluator will control for preenrollment differences, using regression analyses for key outcomes, to ensure that differences that have arisen by chance do not distort inferences about whether treatment-control differences are evidence of true program effects.

E. SUMMARY

Most of the 15 programs found enrollment to be a challenge. In general, although each program had unique features, all of them took one of two basic approaches to enrolling patients: (1) obtaining lists of prospective enrollees from hospitals or health care networks, or (2) recruiting physicians who then referred patients to the program. However, programs with strong preexisting ties to participating physicians and health care providers had greater success in recruiting and enrolling patients than did programs without such ties. Notably, many program staff frequently reported high patient refusal rates as the primary reason for enrollment shortfalls. The patients who were enrolled were, for the most part, older and sicker than was the average Medicare beneficiary. On average, however, patients enrolling in the demonstration also had higher annual incomes and were wealthier and better educated than were Medicare beneficiaries nationally. Based on Medicare claims data, patients enrolling in the demonstration suffered from a number of chronic conditions, including CHF, COPD, diabetes, and cancer. A substantial percentage of patients in many of the programs died during the first year of program operations.

The evaluator also compared the patients who enrolled in the demonstration with groups of comparable nonparticipants simulated from Medicare claims data, based on each program's

target diagnoses and utilization requirements. In general, the programs enrolled patients with costs similar to or lower than the costs estimated by the evaluator's waiver calculation. Only two programs enrolled patients whose monthly payments were substantially higher than those estimated for the waiver calculations. Finally, as expected under random assignment, the treatment and control group members had similar characteristics before random assignment. The number of statistically significant differences between the two groups was small, and what one would expect to occur by chance.

IV. PATIENTS' AND PHYSICIANS' PERCEPTIONS ABOUT CARE COORDINATION AND ITS EFFECTS

Beneficiaries with chronic illnesses often face fragmented health care that is poorly coordinated across multiple provider types and settings, and that fails to devote sufficient time to education about the beneficiaries' conditions, appropriate self-care, or assistance with access to support services. The combination of these factors, along with the frequency, intensity, and mix of health care services that chronically ill beneficiaries require, may lead to poor clinical outcomes, increased health care expenditures, and both patients' and providers' dissatisfaction with the care received.

The improvement in care coordination that the Medicare Coordinated Care Demonstration (MCCD) seeks to foster requires patient and physician buy-in. Participating programs are unlikely to interest and obtain cooperation from either patients or physicians unless the programs are able to convince them that their active participation in the program will benefit them in some way.

For example, patients must be satisfied with their relationship with their care coordinators, with the information and assistance they receive, and with their communication with their providers. Physicians must have confidence in the professional competence of the care coordinators and must believe that cooperating with the program will benefit their patients, without adding burden to themselves, their office staff, or their patients. They also are likely to be concerned about any possible program effects on their relationship with patients and on practice income. If patients and providers are not satisfied in these ways, the programs probably will not affect patients' and physicians' behaviors, the quality of care probably will not improve, and the use and cost of Medicare services probably will not decrease.

The analysis of patient satisfaction in this chapter is based on data from telephone interviews that Mathematica Policy Research, Inc. (MPR) conducted with patients at 12 of the 15 MCCD programs. To assess the extent to which patients were aware of the intervention they were supposed to be receiving, the chapter first examines the proportion of the treatment group that reported receiving help from care coordinators. To assess potential contamination, it also examines the proportion of the control group that reported receiving such assistance. It then shows the proportion of treatment group participants who gave their care coordinators favorable ratings on various aspects of the services provided, such as the care coordinators' knowledge, ability, and attitudes. Finally, the chapter assesses treatment-control differences in ratings of treatment choice and overall access to information and health care (for example, providers remaining in contact with each other and explaining treatment).

This chapter also examines the reactions of physicians to various aspects of the programs and assesses the physicians' perceptions about the programs' effects on their practices and patients. The evaluation measured physicians' impressions by surveying a small sample of the physicians who were providing care to treatment group members in the 15 MCCD programs.

A. PATIENTS' SATISFACTION

The patient survey for each program was conducted in two waves: one wave approximately 12 months after program startup (May through September 2003), and the second about 6 months after the first wave (October 2003 through June 2004). By drawing from beneficiaries who enrolled during the first 6 months of program operations, the sample for the first wave consisted of those who would have had 7 to 12 months of experience with the program by the time they were interviewed. Similarly, the second wave surveyed beneficiaries who enrolled during the 7th through 12th months after program startup, again yielding a follow-up period of 7 to 12 months after enrollment; that sample was pooled with the first cohort. The interviews were

conducted by telephone, using computer-assisted software. The patient survey instruments contain a core set of questions that were asked of all interviewees, regardless of diagnosis or condition, as well as a series of condition- or disease-specific modules. Each patient completed the one disease-specific module that best matched his or her primary health problem, as assessed by the program's intake staff at the time of enrollment. (A "generic" module was administered to patients who had no dominant chronic illness reported.) The survey collected data on patient demographics, primary language, well-being, health status, satisfaction with care, health-related behaviors, adherence to medication regimens, and knowledge of condition. Patients spent an average of about 35 minutes to complete the survey.

The combined target sample size in each demonstration program for the 2 survey waves was 618 completed interviews (309 each for the treatment and control groups), which would yield 80 percent power to detect a difference of 10 percentage points in a binary variable with a mean of 0.50. However, most of the programs did not have enough enrollees to generate that number of completed interviews. Thus, MPR attempted to interview all of the patients of the 8 programs that had at least 400 but fewer than 700 enrollees, (allowing for some survey nonresponse). For the 4 programs with more than 700 enrollees, MPR drew a random sample of the patients to interview, under the assumption that interviews would be completed with about 90 percent. Three programs enrolled about 100 or fewer patients during their first year and were not surveyed. The total number of patients surveyed was 7,526, with an overall response rate of nearly 95 percent and response rates of 90 percent or higher for 11 of the 12 programs whose patients were surveyed (Table IV.1).

TABLE IV.1

SAMPLE SIZES AND RESPONSE RATES FOR THE PATIENT SURVEY

Program	Sample Size	Completes	Response Rate (Percent)
Avera (AVE)	424	395	96.8
Carle (CAR)	704	684	97.6
CenVaNet (CEN)	704	652	94.4
Charlestown (CCI)	607	562	94.6
CorSolutions (COR)	686	620	93.4
Hospice of the Valley (HOS)	483	414	92.0
Health Quality Partners (HQP)	689	675	98.4
Jewish Home and Hospital (JHH)	575	483	84.9
Medical Care Development (MCD)	589	532	95.5
Mercy (MER)	686	649	97.0
QMed (QMD)	691	658	95.8
Washington University (WSH)	688	623	94.1
Total	**7,526**	**6,947**	**94.7**

Source: Mathematica Policy Research, Inc. (MPR) Patient Survey, conducted between May 2003 and June 2004.

Note: The response rate is slightly higher than the ratio of completed interviews to sample size because sample members who were ineligible for the survey have been excluded from the denominator in calculating the rate. Sample members were ineligible for the survey if they died within 3 months after enrolling in the study. In the case of other sample members who were deceased at the time that the survey was fielded, MPR attempted to locate a family member to complete the survey on the decedent's behalf.

1. Receipt of Care Coordination and Satisfaction with Care Coordinator

Across the 12 demonstration programs included in the survey, 65 percent of treatment group patients reported receiving help from a nurse, care coordinator, or social worker in arranging care (Table IV.2). However, the percentages varied widely across the programs, from a low of only 30 percent of patients being aware of the intervention (QMed) to a high of 81 percent (Mercy). Not surprisingly, relatively few control group members (15 percent overall) reported receiving help from a care coordinator, although this proportion may be higher than one might have expected; furthermore, for four programs, the proportion of patients who reported receiving such help exceeded 20 percent. Because a sizeable proportion of the treatment group was unaware that the assistance of a care coordinator was available, and because a nontrivial proportion of the control group received some services that potentially were similar to the services offered by the programs, the effects of the intervention are likely to be considerably lower than if all treatment group members and no control group members received care coordinator assistance.

Care coordinators were rated on 4 dimensions—support and monitoring, help arranging services, ability to provide education to patients, and ability to assist patients in adhering to treatment recommendations—each of which had 3 or 4 specific indicators (for a total of 14 indicators). Treatment group patients generally were very satisfied with the care coordination that they received, with about one-third to one-half of the patients surveyed rating

TABLE IV.2

PERCENTAGE REPORTING THAT A CARE COORDINATOR HELPED TO ARRANGE CARE

	AVE	CAR	CEN	CCI	COR	HQP	HOS	JHH	MCD	MER	QMD	WSH	Average
Treatment	59	77	66	74	55	76	74	66	57	81	30	62	65
Control	17	8	9	9	19	4	21	28	23	22	3	20	15

Note: See Table IV.1 for the full names of the programs.

their coordinators as excellent on the 14 indicators examined (Tables IV.3 and IV.4), and most of the rest rating them as "very good." Very few patients (fewer than 10 percent in each program) rated the programs as only fair or poor on any of the measures (not shown).

Care coordinators received especially high marks on indicators of the emotional support and monitoring they offered, especially on their "caring attitude," with more than 60 percent of the patients, on average, giving ratings of "excellent" (Table IV.3). Patients also rated programs highly on staying in touch (more than one-half rating their programs' care coordinators as excellent on this measure, on average). The programs' inclusion of the patients and their families in decisions and the programs' ability to help patients to cope with their illnesses and to avoid complications received somewhat lower but still quite positive ratings, on average.

Two programs (Avera and Health Quality Partners) received markedly higher care coordinator ratings from their patients on each of the four indicators related to support and monitoring than did the other programs (see the shaded sections of Table IV.3). For example, more than two-thirds of the patients of those programs reported that their care coordinators did an excellent job of staying in touch with them. On every measure, two other programs (Jewish Home and Hospital and QMed) fared markedly worse than the other programs.

Patients were somewhat less satisfied with the help that their programs' care coordinators gave them in arranging appointments or services. On average across the 12 programs, about 35 to 40 percent of the patients gave their care coordinators an excellent rating on this indicator. Carle received consistently high ratings. Jewish Home and Hospital and QMed received markedly lower ratings than did the other programs on the service arrangement indicators. Substantial minorities of patients (10 to 24 percent) gave their programs' care coordinators a fair or poor rating (not shown). Those less favorable ratings are likely due to the fact that most programs focused more on monitoring and education than on service arrangements.

TABLE IV.3

TREATMENT GROUP PATIENTS' RATINGS OF CARE COORDINATORS ON SUPPORT AND SERVICE ARRANGEMENT
(Percentage Rating Excellent)

	AVE	CAR	CEN	CCI	COR	HQP	HOS	JHH	MCD	MER	QMD	WSH	Average
Support/Monitoring													
Staying in Touch	**70.2**	55.5	50.9	56.4	43.6	**68.4**	50.3	**26.8**	52.0	61.0	**28.3**	58.5	51.8
Having a Caring Attitude	71.6	66.9	62.6	66.0	53.1	**76.2**	63.6	**39.9**	66.1	69.4	**36.7**	64.3	61.4
Including Patient and Family in Decisions	**56.6**	50.2	47.0	47.8	32.7	**57.2**	43.0	**29.1**	45.8	50.7	**21.4**	43.8	43.8
Helping Patient to Cope and to Avoid Complications	56.1	50.7	46.6	49.5	38.2	**59.9**	47.1	**24.1**	50.3	49.0	**21.9**	45.8	44.9
Service Arrangement													
Helping to Make Appointments	50.0	**58.9**	42.5	37.1	43.5	43.5	41.7	**12.8**	47.8	36.8	**16.7**	42.4	39.5
Helping to Arrange Payment for Noncovered Services	42.2	**44.8**	41.9	35.8	32.6	36.8	40.0	**23.2**	30.6	38.2	**24.0**	35.9	35.5
Recommending Community Services	35.2	**48.5**	39.5	36.0	30.5	**54.2**	**51.2**	**24.1**	**26.7**	41.0	**25.0**	33.0	37.1

Source: Mathematica Policy Research, Inc. Patient Survey, conducted between May 2003 and June 2004.

Notes: See Table IV.1 for the full names of the programs.

Estimates in boldface are ones that are more than 1 standard deviation above or below the mean across all sites for that measure. Shaded estimates are 1 standard deviation *above* the mean; unshaded boldface estimates are 1 standard deviation *below* the mean.

TABLE IV.4

TREATMENT GROUP PATIENTS' RATINGS OF CARE COORDINATORS ON EDUCATION AND ABILITY TO IMPROVE ADHERENCE
(Percentage Rating Excellent)

	AVE	CAR	CEN	CCI	COR	HQP	HOS	JHH	MCD	MER	QMD	WSH	Average
Education Skills													
Knowledge	61.6	57.2	57.0	56.3	43.9	**67.1**	51.2	**25.2**	54.5	56.7	**26.5**	48.7	50.5
Ability to Obtain Answers from Physician	44.4	**52.4**	40.7	42.5	40.0	49.2	42.7	**24.2**	48.3	42.1	**23.0**	36.9	40.5
Ability to Explain Terms	51.6	49.4	46.9	49.5	36.2	**56.2**	45.2	**19.7**	48.2	45.7	**26.5**	45.0	43.3
Ability to Explain Warning Signs	**52.9**	45.2	45.1	40.3	39.0	**53.7**	41.6	**24.4**	46.2	43.4	**23.2**	42.4	41.5
Adherence Assistance													
Ability to Explain Diet	38.2	39.5	44.8	39.0	40.9	**57.9**	42.6	**17.3**	49.4	33.2	**19.5**	36.4	38.2
Ability to Explain Exercise	35.6	37.6	40.9	30.1	34.9	**53.5**	41.2	**17.9**	44.6	32.1	**20.3**	36.0	35.4
Ability to Explain Medication	42.4	45.6	45.4	51.1	42.4	**54.1**	46.7	**17.7**	51.5	46.7	**27.4**	43.8	42.9

Source: Mathematica Policy Research, Inc. Patient Survey, conducted between May 2003 and June 2004.

Notes: See Table IV.1 for the full names of the programs.

Estimates in boldface are ones that are more than 1 standard deviation above or below the mean across all sites for that measure. Shaded estimates are 1 standard deviation *above* the mean; unshaded boldface estimates are 1 standard deviation *below* the mean.

Patients had high praise for the care coordinators' ability to educate them. On average, more than one-half the patients gave a rating of excellent on the care coordinators' knowledge, and for only two programs did fewer than 43 percent of patients give such a rating (Table IV.4). On average, about 40 to 43 percent of patients rated their programs' care coordinators as excellent on their ability to explain symptoms or get physicians to answer questions or help them to identify warning symptoms of their medical conditions; these rates varied only moderately across most programs.

A similarly modest proportion of patients gave excellent ratings on the ability of their programs' care coordinators to explain recommended diet, medication, and exercise regimens (Table IV.4). Of all the measures, patients were least likely to give coordinators very high marks on their ability to explain exercise regimens. Again, however, few patients in most programs rated the programs as fair or poor on these indicators (not shown). The somewhat lower ratings on these measures may reflect the difficulty of helping patients to adhere to treatment recommendations, or that education on exercise was a less intense focus for the care coordinators. Health Quality Partners received markedly higher ratings than did the other programs on all four measures of the coordinators' ability to explain recommendations. These high patient ratings were consistent with the scoring results described in Chapter II, in which Health Quality Partners had the highest ranking of the 15 programs on patient education and was one of the most highly ranked on monitoring as well.

Overall, a consistent pattern emerges from these numerous measures of patient ratings of their care coordination intervention, with Health Quality Partners consistently receiving notably higher marks than other programs. Carle and Avera also were rated highly on some measures, especially those related to support and service arrangement. Avera's high ranking on explaining early warning signs is consistent with the scoring algorithm's strong rating of Avera on

monitoring (the third highest). Carle's high rating from patients on the ability of care coordinators to obtain answers from physicians is consistent with its top score among all programs on improving communications and coordination among providers. Carle also received one of the three highest scores on service arrangement in the scoring algorithm, consistent with its patients' high opinion of this aspect of Carle's intervention.

2. Satisfaction with Health Care

If the demonstration programs succeed in improving communications and in coordinating care, then one would expect that treatment group members would rate various aspects of their care more highly than would control group members. MPR compared the treatment and control groups' satisfaction levels (excellent, very good, good, fair, or poor) on seven indicators of the perceived quality of the health care that the groups' members received from their providers. The measures include satisfaction with the following: patients' perceived degree of choice in treatment, the extent to which providers maintained contact with patients, explanations received from specialists, explanations about side effects, explanations about treatments, explanations about tests, and the speed with which test results were provided.

Choice of Treatment. Only 1 of the 12 programs (Avera) had a statistically significant treatment-control difference in patients' belief that they had a choice in the treatment of their condition (Table IV.5). The vast majority (98 percent) of Avera's treatment group members reported believing that they had a choice in the treatment they received, compared with 82 percent of the control group ($p = 0.01$) (not shown).

Participants' Satisfaction with Health Care. Treatment group members in only a few programs were more likely than their control counterparts to rate as excellent the information they received from their providers and the providers' ability to communicate with each other (Table IV.5). Differences favoring the treatment group occurred most often on the measure for

TABLE IV.5

SUMMARY OF ESTIMATED EFFECTS ON PATIENTS' SATISFACTION-WITH-CARE MEASURES

	AVE	CAR	CEN	CCI	COR	HOS	HQP	JHH	MCD	MER	QMD	WSH
Access to Care Coordinator and Treatment Choice												
Believed a choice in treatment of condition was available	++											
Patients' Ratings (Percentage Excellent)												
Providers keep in touch with each other	++	++		+			+			+		
Explanations from specialists	++								++			
Explanations of possible side effects		+				+				+		
Explanations of treatment						+	+			+	+	
How quickly received test results												
Explanations of test results	++											

Source: Mathematica Policy Research, Inc. Patient Survey, conducted between May 2003 and June 2004.

Note: See Table IV.1 for the full names of the programs.

+Signifies statistically significant treatment-control difference ($p < 0.10$) that favors the treatment group and is modest (less than 10 percentage points and less than one-half the size of the control group proportion [*pc*] or its complement [1-*pc*]).

++Signifies statistically significant treatment-control difference ($p < 0.10$) that favors the treatment group and is large (more than 10 percentage points or at least one-half the size of the control group proportion [*pc*] or its complement [1-*pc*]).

providers keeping in touch with each other; in 5 of the 12 programs, treatment group members gave significantly higher ratings than did control group members. Table IV.6 illustrates the magnitude of these treatment-control differences: 25 to 50 percent of treatment group members gave excellent ratings on how well their providers kept in touch with each other, with the treatment-control differences for Carle and Avera being particularly large. CenVaNet's treatment group gave providers significantly *lower* ratings than did the control group on this measure.

Treatment group members in four programs also were more likely than control group members to rate as excellent explanations of treatments (Table IV.5). Treatment-control differences for explanations of side effects and for explanations from specialists favored treatment groups for only three programs and for only two programs, respectively. Differences for explanations of tests favored the treatment group for only one program. None of the programs improved the timeliness with which test results were delivered.

No consistent pattern emerges across programs for these care ratings. Avera's and Mercy's treatment groups gave significantly higher ratings than did their control groups on three of the six measures examined. Three other programs had significant differences on two of the six measures, and another three programs had significant effects on just one of the measures.

B. PHYSICIANS' SATISFACTION

On the whole, physicians appeared to be satisfied with the program and its effects on their practice (although with some concerns about increased paperwork and reduced revenue), service arrangement, care coordination, physician-patient relations, and patients' health and satisfaction. The majority of physicians also were comfortable with care coordinators' clinical skills, valued their input, and felt the program improved patients' quality of care. Most would also recommend

TABLE IV.6

ILLUSTRATIVE OUTCOMES FOR PATIENTS' SATISFACTION WITH HEALTH CARE
(Percentage Rating Excellent)

	AVE	CAR	CEN	CCI	COR	HQP	HOS	JHH	MCD	MER	QMD	WSH
Providers Keep in Touch with Each Other												
Treatment	46.3	44.8	31.9	33.1	33.0	43.9	30.2	25.6	47.1	35.1	29.6	34.0
Control	31.1	27.6	37.1	30.7	32.1	36.4	24.7	19.9	41.8	27.4	25.9	36.0
Difference	15.2***	17.2***	-5.3*	2.4*	0.9	7.5*	5.4	5.7	5.2	7.7**	3.7	-2.0
p-Value	0.009	0.000	0.063	0.068	0.986	0.053	0.142	0.228	0.380	0.015	0.837	0.808
Explanations of Possible Side Effects												
Treatment	29.3	29.9	23.1	20.8	25.2	27.7	23.1	12.4	30.0	23.7	29.8	26.4
Control	19.1	22.7	25.1	19.0	25.3	27.7	17.0	15.4	25.4	18.6	25.5	24.4
Difference	10.2	7.2*	-2.0	1.7	-0.1	0.0	6.1*	-3.0	4.6	5.1*	4.4	2.1
p-Value	0.115	0.089	0.859	0.303	0.233	0.836	0.080	0.626	0.176	0.074	0.594	0.658

Source: Mathematica Policy Research, Inc. Patient Survey, conducted between May 2003 and June 2004.

Note: See Table IV.1 for the full names of the programs.

*Significantly different from 0 at the 0.10 level, 2-tailed test.
**Significantly different from 0 at the 0.05 level, 2-tailed test.
***Significantly different from 0 at the 0.01 level, 2-tailed test.

the program to other physicians. Opinions were more mixed about the program's effect on patients' self-management behavior.

To assess physicians' satisfaction, the evaluation examined physicians' responses to survey questions in eight areas. Six areas focused on physicians' opinions about the programs' perceived effects, including effects on (1) their medical practice, (2) patients' self-management, (3) service arrangements for patients, (4) care coordination, (5) their own relationships with their patients, and (6) patient outcomes. For the seventh area, physicians were asked to rate care coordinators' clinical competence. For the eighth area, they were asked to provide an overall assessment of the care coordination program.

The physician survey was conducted at all 15 programs and in 2 waves, each intended to yield completed surveys on a sample of 25 physicians from each program (or on the number that could be obtained, if the patients in a particular program identified fewer than 25 physicians). The first wave of the survey began in June 2003; the sample for that wave was drawn from the physicians identified by treatment group patients who enrolled during the first 9 months after program startup. (At the time of enrollment, patients were asked to name the physician whom they saw most often for care for their targeted health problems.) The second-wave sample, which began in May 2004, was drawn from physicians identified by patients enrolling between 10 and 20 months after program startup. In order to obtain 25 completed interviews, a sample of 37 physicians for each program was selected (assuming a 70-percent completion rate). Physicians were selected for the survey with probability proportional to the number of treatment group members in their practices. If treatment group members enrolled in a program identified fewer than 37 different physicians, all of the physicians of that program were interviewed.

The interviews were conducted by telephone, using computer-assisted software, and took an average of 12 minutes to complete. The total number of physicians surveyed was 1,018, and the overall response rate was about 64 percent (Table IV.7). Physicians who were contacted but

TABLE IV.7

SAMPLE SIZES AND RESPONSE RATES FOR THE PHYSICIAN SURVEY

Program	Sample Size	Physicians Contacted	Response Rate (Percent)	Familiar with Program[a]	Percentage Familiar with Program[a]
Avera	78	52	67.5	48	92.3
Carle	89	49	56.3	42	85.7
CenVaNet	75	20	31.3	12	60.0
Charlestown	28	23	85.2	21	91.3
CorSolutions	78	46	59.0	39	84.8
Georgetown	66	51	78.5	40	78.4
Hospice of the Valley	78	39	50.7	21	53.9
Health Quality Partners	77	57	76.0	54	94.7
Jewish Home and Hospital	78	51	65.4	26	51.0
Medical Care Development	89	68	76.4	40	58.8
Mercy	89	72	80.9	66	91.7
QMed	78	29	37.7	23	79.3
Quality Oncology	19	14	73.7	10	71.4
University of Maryland	18	16	88.9	14	87.5
Washington University	78	51	66.2	16	31.4
Total	**1,018**	**638**	**64.0**	**472**	**74.0**

Source: Mathematica Policy Research, Inc. Physician Survey.

[a]Among physicians contacted.

who did not recognize the program name were not interviewed any further, as they would not have been able to answer the questions about their reactions to the demonstration programs. About 25 percent of the physicians who were contacted were not familiar with the program. With respect to the findings, physicians in Charlestown appeared to be the most satisfied compared with physicians in other programs on the majority of the measures. Physicians in QMed appeared to be the least satisfied.

1. Effects on Physicians' Practice

Physicians generally believed that the programs had favorable effects on their practices, with the average percentage of physicians across programs who gave favorable ratings ranging from 4 to 75 percent for the 10 measures (Table IV.8). Physicians were especially appreciative of the programs' help in making it easier overall to care for patients; 75 percent of physicians across all the programs reported that the programs made overall care at least a little easier. However, they did not typically give high ratings to the usefulness of the reports generated by the programs. Only 42 percent said the reports were very useful.

Physicians' perceptions varied across measures and across programs. On the one hand, for example, 9 of 10 physicians at the University of Maryland found the program's reports to be very useful, and 95 percent of Carle's physicians believed that the program helped to make their care more evidence-based. On the other hand, QMed was given low ratings by physicians on reducing paperwork and telephone burden; more than one-half its patients' physicians reported that the program increased their paperwork burden (not shown).

Z-scores were calculated to compare the programs across the various measures of physicians' practice. The *Z*-score yields a standardized measure for each program indicating how far, and in what direction, that program's score deviates from the mean for all of the programs on a particular outcome.

TABLE IV.8

PHYSICIANS' PERCEPTIONS ABOUT PROGRAM EFFECTS ON MEDICAL PRACTICE
(Percentages)

	Mean	Standard Deviation	AVE	CAR	CEN	CCI	COR	GEO	HOS	HQP	JHH	MER	MCD	QMD	QUA	UMD	WSH
Usefulness of Program Reports on Patients																	
Very useful	42	23.7	52	49	0	76	42	49	40	13	23	51	32	14	67	91	33
Reduces Problems with Polypharmacy																	
A little or a lot better	56	18.8	69	66	44	81	66	58	80	50	52	79	62	41	11	54	29
Reduces Physicians' and Staffs' Telephone Time																	
A little or a lot better	55	23.8	69	88	27	86	68	70	55	52	52	84	67	4	22	54	33
Makes Things Easier for Staff																	
A little or a lot better	56	20.1	63	78	40	86	66	60	57	32	40	73	74	22	22	77	47
Reduces Physicians' Paperwork																	
A little or a lot better	26	11.5	26	32	10	48	29	38	32	15	13	37	38	9	22	14	20
Reduces Malpractice Risk																	
A little or a lot better	24	11.3	13	47	22	50	29	20	18	16	17	28	31	16	11	23	14
Makes Care More Evidence-Based																	
A little or a lot better	49	20.1	56	95	30	67	53	41	48	45	33	47	78	61	20	21	33
Overall Effects of Program on Practice Revenue																	
Increased	4	5.6	0	19	0	12	11	0	0	0	5	4	6	5	0	0	0
Overall Monitoring and Followup of Patients																	
Very good or excellent	71	19.2	95	83	38	100	63	79	67	69	67	76	70	40	38	92	83
Makes It Easier Overall to Care for Patients																	
A little or a lot better	75	16.8	81	90	55	100	79	85	81	71	75	83	77	65	33	93	53
Average *Z*-Value			0.24	**1.24**	**-1.04**	**1.51**	0.36	0.19	0.07	-0.56	-0.44	0.57	0.49	-0.93	**-1.22**	0.20	-0.67

Source: Mathematica Policy Research, Inc. Physician Survey.

Notes: See Table IV.1 for the full names of the programs.

Estimates in boldface are ones that are more than 1 standard deviation above or below the mean across all sites for that measure. Shaded boldfaced estimates are 1 standard deviation *above* the mean; unshaded boldface estimates are 1 standard deviation *below* the mean.

To calculate the average Z-value, the program's Z-score for the percentage rating a particular measure favorably was first calculated for each of the individual outcome measures. The score for each measure was calculated by subtracting the mean of the percentage rating the program as favorable across the 15 programs from the percentage rating the measure as favorable in the program, and dividing by the standard deviation across the 15 programs. Next, an average of the Z-scores for all of the measures for that program was calculated. Average Z-values greater than 1 or less than -1 indicate that a program was, on average, more than one standard deviation from the mean across all measures.

The physicians of two programs (Carle and Charlestown) consistently rated their programs more highly in terms on their effect on medical practice than did the physicians of the other programs. Quality Oncology's and CenVaNet's physicians consistently rated their programs lower than did the other programs' physicians, by more than one standard deviation, on average, across the 10 measures.

2. Effects on Patients' Self-Management

Overall, physicians found that the MCCD programs did not have much effect on improving patients' self-management behaviors (Table IV.9). More than one-half of all physicians believed that their programs improved patients' medication adherence (59 percent), ability to monitor themselves (59 percent), and ability to make and keep appointments (54 percent). These percentages imply, however, that more than 40 percent of the physicians believed that the programs did not improve those behaviors. Fewer physicians believed that the MCCD was as strong in improving their patients' exercise habits or diets; only 22 percent and 36 percent of physicians felt that the programs improved their patients' exercise habits and diets, respectively. About one-half of physicians (51 percent) thought that the MCCD did a very good or excellent job overall in improving patients' self-management behaviors.

There was wide variation across the programs on individual measures. On the one hand, for example, 9 of 10 physicians associated with the University of Maryland believed that the program improved patients' ability to monitor themselves. Medical Care Development's effect on improving its patients' exercise habits was rated relatively highly, with its physicians rating the program higher compared with physicians of other programs by almost 2 standard deviations. Moreover, physicians in two programs (Avera and Charlestown) consistently rated their programs' effects on all six patient self-management behaviors higher than did physicians in the other programs, by more than 1 standard deviation, on average. On the other hand, physicians in

TABLE IV.9

PHYSICIANS' PERCEPTIONS ABOUT PROGRAM EFFECTS ON PATIENTS' SELF-MANAGEMENT
(Percentages)

	Mean	Standard Deviation	AVE	CAR	CEN	CCI	COR	GEO	HOS	HQP	JHH	MER	MCD	QMD	QUA	UMD	WSH
Increases Patients' Medication Adherence																	
Very good or excellent	59	18.3	78	71	44	76	58	72	60	50	52	70	58	28	14	83	64
Improves Patients' Diets																	
Very good or excellent	36	13.4	62	36	25	38	47	50	30	38	27	38	54	13	14	27	44
Improves Patients' Exercise Habits																	
Very good or excellent	22	12.6	29	27	14	42	17	16	29	23	27	11	46	6	0	9	33
Improves Patients' Self-Monitoring																	
Very good or excellent	59	23.1	85	76	44	86	43	61	70	53	50	76	55	32	0	91	67
Improves Patients' Making And Keeping of Appointments																	
Very good or excellent	54	20.7	74	73	25	86	47	50	61	41	55	71	39	29	14	64	80
Overall Improvement in Patients' Self-Management																	
Very good or excellent	51	16.7	75	63	38	76	41	66	55	41	48	63	53	30	14	42	64
Average *Z*-Value			**1.16**	0.58	-0.85	**1.15**	-0.21	0.34	0.20	-0.29	-0.20	0.36	0.40	**-1.39**	**-2.09**	0.16	0.68

Source: Mathematica Policy Research, Inc. Physician Survey.

Notes: See Table IV.1 for the full names of the programs.

Estimates in boldface are ones that are more than 1 standard deviation above or below the mean across all sites for that measure. Shaded boldfaced estimates are 1 standard deviation *above* the mean; unshaded boldface estimates are 1 standard deviation *below* the mean.

To calculate the average *Z*-value, the program's *Z*-score for the percentage rating a particular measure favorably was first calculated for each of the individual outcome measures. The score for each measure was calculated by subtracting the mean of the percentage rating the program as favorable across the 15 programs from the percentage rating the measure as favorable in the program, and dividing by the standard deviation across the 15 programs. Next, an average of the *Z*-scores for all of the measures for that program was calculated. Average *Z*-values greater than 1 or less than -1 indicate that a program was, on average, more than 1 standard deviation from the mean across all measures.

two other programs (QMed and Quality Oncology) consistently rated their programs less favorably than did physicians in the other programs.

3. Effects on Service Arrangement

Overall, physicians were impressed with the programs' help in arranging other care for patients (65 percent believed that their programs were helpful in this area) and in arranging transportation or meals (66 percent) (Table IV.10). Physicians were much less impressed with the programs' ability to help patients to obtain specialist appointments (41 percent) or expensive prescriptions (46 percent).

Notable differences across programs and specific measures were reported. For example, 95 percent of its physicians believed that Charlestown's program helped patients to arrange for care, but only 11 percent of physicians in QMed believed that their program was helpful in this area.[1] In summarizing findings across the four measures of service arrangement, physicians in two programs (Charlestown and Mercy) consistently rated their programs higher than did physicians in the other programs, by more than 1 standard deviation, on average. Conversely, the physicians of patients enrolled in QMed and Quality Oncology rated those programs substantially lower.

4. Effects on Care Coordination

Even though care coordination was the focus of the MCCD, physicians' opinions about how helpful the programs had been with various aspects of care coordination were decidedly mixed. Across all programs, about one-half to two-thirds of physicians found that their programs were

[1] In fact, helping patients to arrange for care was not part of QMed's intervention. Again, as with the rating scores in Chapter II, some of these survey results serve more of a descriptive purpose, as some programs intentionally designed interventions that did not focus on certain areas.

TABLE IV.10

PHYSICIANS' PERCEPTIONS ABOUT PROGRAM EFFECTS ON SERVICE ARRANGING
(Percentages)

	Mean	Standard Deviation	AVE	CAR	CEN	CCI	COR	GEO	HOS	HQP	JHH	MER	MCD	QMD	QUA	UMD	WSH
Helps Patients to Obtain Specialist Appointments																	
A little or a lot better	41	19.8	38	41	11	85	42	34	70	25	59	50	18	14	30	50	42
Helps Patients to Obtain Expensive Prescriptions																	
A little or a lot better	46	15.9	24	56	50	85	48	47	47	34	44	75	39	29	30	38	45
Helps Patients to Arrange Other Care (for Example, Therapy)																	
A little or a lot better	65	22.1	55	82	56	95	57	63	85	72	87	82	73	11	33	42	75
Helps Patients to Arrange Transport or Meals																	
A little or a lot better	66	21.7	58	94	56	90	52	61	85	86	80	92	68	15	38	50	67
Average *Z*-Value			-0.59	0.68	-0.53	**1.80**	-0.20	-0.14	0.84	-0.07	0.62	**1.07**	-0.29	**-1.80**	**-1.07**	-0.45	0.13

Source: Mathematica Policy Research, Inc. Physician Survey.

Notes: See Table IV.1 for the full names of the programs.

Estimates in boldface are ones that are more than 1 standard deviation above or below the mean across all sites for that measure. Shaded boldfaced estimates are 1 standard deviation *above* the mean; unshaded boldface estimates are 1 standard deviation *below* the mean.

To calculate the average *Z*-value, the program's *Z*-score for the percentage rating a particular measure favorably was first calculated for each of the individual outcome measures. The score for each measure was calculated by subtracting the mean of the percentage rating the program as favorable across the 15 programs from the percentage rating the measure as favorable in the program, and dividing by the standard deviation across the 15 programs. Next, an average of the *Z*-scores for all of the measures for that program was calculated. Average *Z*-values greater than 1 or less than -1 indicate that a program was, on average, more than 1 standard deviation from the mean across all measures.

helpful in coordinating care with family members (69 percent), in resolving family conflicts (50 percent), and in maintaining or improving continuity of care (66 percent) (Table IV.11). Fewer than half of the physicians found their programs to be helpful in coordinating with other physicians (41 percent), reducing the frequency of duplicate testing (41 percent), or in helping their patients to deal with contradictory information (48 percent).

Some differences across the programs were particularly noteworthy. For example, 95 percent of Charlestown's physicians believed that the program helped to maintain or improve continuity of care, and 91 percent of Mercy's physicians believed that the Mercy program helped physicians to coordinate care with family members. In summarizing findings across the six measures of care coordination, physicians in two programs (Charlestown and Mercy) consistently rated their programs substantially higher than did physicians in the other programs. Physicians in CenVaNet and QMed consistently rated their programs substantially lower, by an average of more than 1 standard deviation.

5. Effects on Physician-Patient Relations

Physicians found that the programs had only a limited positive effect on improving physician-patient relations, but they also believed that the programs produced few negative effects. Most physicians believed that their patients accepted the programs: 80 percent of physicians overall felt that their patients did not mind having the program involved (Table IV.12). Nearly all physicians (95 percent) felt that the programs did not undermine their patients' confidence in the physician care they received. However, only 40 percent of physicians felt that the programs improved their relations with patients.

Analysis across the programs revealed some substantial differences in physicians' beliefs about effects on physician-patient relations. For example, 57 percent of Charlestown's physicians believed that the program did improve relations with patients, a rating that was higher

TABLE IV.11

PHYSICIANS' PERCEPTIONS ABOUT PROGRAM EFFECTS ON CARE COORDINATION
(Percentages)

	Mean	Standard Deviation	AVE	CAR	CEN	CCI	COR	GEO	HOS	HQP	JHH	MER	MCD	QMD	QUA	UMD	WSH
Helps Physicians to Coordinate Care with Other Physicians																	
A little or a lot better	41	16.1	55	51	25	76	32	51	67	27	32	47	21	30	22	36	40
Helps Patients to Deal with Contradictory Information from Other Health Care Providers																	
A little or a lot better	48	14.0	59	58	40	83	32	48	53	36	27	68	50	38	44	42	45
Helps to Reduce Duplicated Tests																	
A little or a lot better	41	14.4	47	55	13	60	41	44	58	33	30	61	50	23	33	23	38
Helps Physicians to Coordinate Care with Family Members																	
A little or a lot better	69	20.9	78	90	36	86	81	76	81	72	68	91	79	24	33	85	54
Helps to Resolve Family Conflicts or Difficult Family Situations																	
A little or a lot better	50	21.8	48	65	11	84	55	58	71	59	43	85	41	16	38	20	55
Maintains or Improves Continuity of Care																	
A little or a lot better	66	17.4	71	83	30	95	64	74	76	67	56	84	74	43	38	77	64
Average *Z*-Value			0.44	0.83	**-1.49**	**1.68**	-0.17	0.34	0.88	-0.27	-0.63	**1.15**	0.00	**-1.28**	-0.98	-0.33	-0.18

Source: Mathematica Policy Research, Inc. Physician Survey.

Notes: See Table IV.1 for the full names of the programs.

Estimates in boldface are ones that are more than 1 standard deviation above or below the mean across all sites for that measure. Shaded boldfaced estimates are 1 standard deviation *above* the mean; unshaded boldface estimates are 1 standard deviation *below* the mean.

To calculate the average *Z*-value, the program's *Z*-score for the percentage rating a particular measure favorably was first calculated for each of the individual outcome measures. The score for each measure was calculated by subtracting the mean of the percentage rating the program as favorable across the 15 programs from the percentage rating the measure as favorable in the program, and dividing by the standard deviation across the 15 programs. Next, an average of the *Z*-scores for all of the measures for that program was calculated. Average *Z*-values greater than 1 or less than -1 indicate that a program was, on average, more than 1 standard deviation from the mean across all measures.

TABLE IV.12

PHYSICIANS' PERCEPTIONS ABOUT PROGRAM EFFECTS ON PHYSICIAN-PATIENT RELATIONS
(Percentages)

	Mean	Standard Deviation	AVE	CAR	CEN	CCI	COR	GEO	HOS	HQP	JHH	MER	MCD	QMD	QUA	UMD	WSH
Patients Do Not Mind Having Program Involved																	
Somewhat or strongly agree	80	11.0	89	93	64	95	67	79	81	72	87	82	82	59	67	92	85
Program Does Not Undermine Patients' Confidence in Physicians' Care																	
Somewhat or strongly agree	95	6.2	100	95	92	100	97	95	100	96	96	97	100	76	87	100	92
Effect on Physicians' Relations with Patients																	
Improved	40	11.1	54	48	36	57	47	50	48	34	40	41	35	30	11	43	33
Average Z-Value			0.97	0.64	-0.78	**1.25**	-0.07	0.27	0.55	-0.36	0.27	0.20	0.17	**-1.95**	**-1.69**	0.73	-0.21

Source: Mathematica Policy Research, Inc. Physician Survey.

Notes: See Table IV.1 for the full names of the programs.

Estimates in boldface are ones that are more than 1 standard deviation above or below the mean across all sites for that measure. Shaded boldfaced estimates are 1 standard deviation *above* the mean; unshaded boldface estimates are 1 standard deviation *below* the mean.

To calculate the average Z-value, the program's Z-score for the percentage rating a particular measure favorably was first calculated for each of the individual outcome measures. The score for each measure was calculated by subtracting the mean of the percentage rating the program as favorable across the 15 programs from the percentage rating the measure as favorable in the program, and dividing by the standard deviation across the 15 programs. Next, an average of the Z-scores for all of the measures for that program was calculated. Average Z-values greater than 1 or less than -1 indicate that a program was, on average, more than 1 standard deviation from the mean across all measures.

than the mean across all the programs by 1½ standard deviations. However, about one-fourth of QMed's physicians felt that QMed undermined patients' confidence in their physicians' care, a very negative rating relative to that of other programs (more than 3 standard deviations from the mean).

In summarizing findings across the three measures of physician-patient relations, the rating given by Charlestown's physicians was consistently higher than was the average rating across all programs, by more than 1 standard deviation. Avera was a close second, with physicians rating that program higher than the ratings given by physicians of other programs, by almost 1 (0.97) standard deviation on average across the measures. In contrast, physicians of QMed and Quality Oncology consistently rated their programs substantially lower than the average across all programs.

6. Effects on Patient Outcomes and Service Use

Overall, physicians found that the programs increased patients' satisfaction and health, with few effects on office visits in either direction (Table IV.13). The majority of physicians found that the programs were beneficial to their patients' health (88 percent), and that they increased patients' satisfaction with health care (65 percent). On average, physicians did not believe the programs had much effect on increasing office visits (9 percent of physicians reported such an increase), but, in the few instances in which a change in office visits was reported, the great majority of physicians (89 percent) believed that the change was medically appropriate. Only a few physicians (2 percent) felt that the programs increased nursing home admissions.

In some cases, large differences across the programs on some of these patient outcome measures were reported. For example, 95 percent of physicians in Charlestown felt that the program increased patients' satisfaction with health care, a rating higher than the average across all programs by more than 2 standard deviations. By contrast, only 29 percent of physicians in

TABLE IV.13

PHYSICIANS' PERCEPTIONS ABOUT PROGRAM EFFECTS ON PATIENTS' OUTCOMES AND SERVICE USE
(Percentages)

	Mean	Standard Deviation	AVE	CAR	CEN	CCI	COR	GEO	HOS	HQP	JHH	MER	MCD	QMD	QUA	UMD	WSH
How Beneficial Program Is to Patients' Health																	
Very or somewhat beneficial	88	12.4	96	95	70	100	89	95	90	86	96	97	97	68	63	100	71
Patients' Satisfaction with Health Care																	
Increased	65	18.4	76	74	38	95	66	74	71	62	56	84	76	29	38	85	57
Office Visits																	
Increased	9	8.4	2	10	0	29	0	15	5	2	0	6	3	18	11	7	21
Increase or Decrease in Visits Medically Appropriate																	
Yes	89	19.4	100	93	67	100	100	100	100	88	100	100	100	40	100	100	50
Nursing Home Admissions																	
Increased	2	3.5	0	0	0	0	0	0	6	0	0	3	4	0	13	0	0
Average *Z*-Value			0.11	0.20	**-1.12**	**1.02**	-0.16	0.38	0.37	-0.34	-0.16	0.48	0.35	**-1.09**	0.12	0.39	-0.56

Source: Mathematica Policy Research, Inc. Physician Survey.

Notes: See Table IV.1 for the full names of the programs.

Estimates in boldface are ones that are more than 1 standard deviation above or below the mean across all sites for that measure. Shaded boldfaced estimates are 1 standard deviation *above* the mean; unshaded boldface estimates are 1 standard deviation *below* the mean.

To calculate the average *Z*-value, the program's *Z*-score for the percentage rating a particular measure favorably was first calculated for each of the individual outcome measures. The score for each measure was calculated by subtracting the mean of the percentage rating the program as favorable across the 15 programs from the percentage rating the measure as favorable in the program, and dividing by the standard deviation across the 15 programs. Next, an average of the *Z*-scores for all of the measures for that program was calculated. Average *Z*-values greater than 1 or less than -1 indicate that a program was, on average, more than 1 standard deviation from the mean across all measures.

QMed reported the program had increased patients' satisfaction, a rating lower than the average across programs by almost 2 standard deviations. Across the five measures of patient outcomes and service use, only Charlestown had higher ratings than the other programs, on average, with ratings 1 standard deviation higher than those of the other programs. Physicians in QMed rated their program lower than others across the five measures, by 1 standard deviation, on average.

7. Care Coordinators' Clinical Judgment and Competence

Overall, physicians across the 15 programs were comfortable with their care coordinators' clinical skills, and they valued the input of these staff. The majority of physicians found their care coordinators' information or feedback to be useful (85 percent), and they rated their care coordinators' overall clinical judgment and competence as very good or excellent (65 percent) (Table IV.14). About one-half of all physicians (48 percent) also reported that their care coordinators identified acute problems in patients or influenced their decisions. A great majority of physicians (94 percent) found that they rarely disagreed with their care coordinators, and they rated their care coordinators' ability to deal with specific issues as very good or excellent (87 percent). Physicians also were favorably impressed with their care coordinators' ability to identify patients' functional problems, and to assess patients' home situations (56 percent and 66 percent of physicians, respectively). Physicians were slightly less impressed with the care coordinators' ability to identify emotional problems (with only 43 percent reporting that care coordinators identified such problems at least sometimes).

Analysis across the programs revealed substantial differences. For example, although all of Charlestown's, Medical Care Development's, and the University of Maryland's physicians found their care coordinators' information or feedback to be useful, only 33 percent of physicians in Quality Oncology held the same view (a rating lower than the average across the programs by almost 3 standard deviations). Likewise, 90 percent of physicians in Charlestown reported that

TABLE IV.14

PHYSICIANS' PERCEPTIONS ABOUT CARE COORDINATORS' CLINICAL COMPETENCE
(Percentages)

	Mean	Standard Deviation	AVE	CAR	CEN	CCI	COR	GEO	HOS	HQP	JHH	MER	MCD	QMD	QUA	UMD	WSH
How Well Care Coordinators Dealt with Specific Issues Physicians Asked Them to Address																	
Very good or excellent	87	12.4	94	83	100	100	74	82	100	92	58	94	96	67	***NA***	89	88
Usefulness of Coordinators' Information or Feedback																	
Somewhat or very useful	85	17.7	88	95	58	100	91	94	94	80	76	97	100	81	33	100	89
Care Coordinators Ever Influenced Physicians' Decisions																	
Yes	48	20.0	58	68	27	90	37	48	52	41	48	65	60	36	10	64	20
How Often Care Coordinators Identified Patients' Acute Problems																	
Sometimes or frequently	48	22.1	73	60	8	90	38	58	57	42	38	76	50	24	11	54	43
How Often Care Coordinators Identified Patients' Emotional Problems																	
Sometimes or frequently	43	20.0	37	63	18	76	34	38	67	45	36	66	67	14	10	38	36
How Often Care Coordinators Identified Patients' Physical or Functioning Problems																	
Sometimes or frequently	56	23.7	74	80	30	95	42	65	57	62	33	77	68	25	10	83	47
How Often Care Coordinators Disagreed with Physicians on Approach to Patients' Problems																	
Sometimes or frequently	6	4.4	6	7	0	0	14	3	5	6	12	6	5	5	0	8	14
Rating of Care Coordinators' Ability to Assess Patients' Home Situations																	
Very good or excellent	66	16.9	74	76	50	95	56	73	68	76	58	75	82	36	29	73	62
Overall Rating of Care Coordinators' Clinical Judgment and Competence																	
Very good or excellent	65	20.3	77	76	43	95	53	73	68	64	45	71	89	28	40	100	57
Average Z-Value			0.44	0.58	**-1.01**	**1.23**	-0.24	0.07	0.38	0.03	-0.53	0.69	0.64	**-1.13**	**-1.87**	0.61	-0.10

Source: Mathematica Policy Research, Inc. Physician Survey.

Notes: See Table IV.1 for the full names of the programs.

Estimates in boldface are ones that are more than 1 standard deviation above or below the mean across all sites for that measure. Shaded boldfaced estimates are 1 standard deviation *above* the mean; unshaded boldface estimates are 1 standard deviation *below* the mean.

To calculate the average Z-value, the program's Z-score for the percentage rating a particular measure favorably was first calculated for each of the individual outcome measures. The score for each measure was calculated by subtracting the mean of the percentage rating the program as favorable across the 15 programs from the percentage rating the measure as favorable in the program, and dividing by the standard deviation across the 15 programs. Next, an average of the Z-scores for all of the measures for that program was calculated. Average Z-values greater than 1 or less than -1 indicate that a program was, on average, more than 1 standard deviation from the mean across all measures.

their care coordinators had influenced their decisions on occasion (a much higher proportion than those of other programs), whereas only 10 percent of physicians in Quality Oncology felt that way (a far lower rating than those of other programs).

Only Charlestown had noticeably higher physician ratings of coordinators' clinical competence across all nine measures than the cross-program average. QMed and Quality Oncology had markedly lower ratings than the other programs.

8. Overall Rating

The patients' primary physicians were pleased with the program overall. Across the 15 programs, on average, a majority (67 percent) of physicians found that the program increased patients' overall quality of care, and 80 percent reported that they would recommend the program to their patients and colleagues (Table IV.15). (About 60 percent would "definitely" recommend the program, and the remainder would "probably" do so.)

Some variations across programs in physicians' ratings were observed. For example, 100 percent of the University of Maryland physicians and 95 percent of Charlestown physicians believed that the program improved their patients' quality of care. By contrast, only 11 percent of physicians in Quality Oncology believed this (representing a rating lower than those given to other programs by more than 2 standard deviations, on average). In fact, Charlestown's physicians both rated their program more highly on quality and were more likely to recommend the program relative to physicians of the other programs, whereas three programs (CenVaNet, QMed, and Quality Oncology) consistently received lower ratings on these two measures.

TABLE IV.15

PHYSICIANS' OVERALL PERCEPTIONS ABOUT PROGRAMS
(Percentages)

	Mean	Standard Deviation	AVE	CAR	CEN	CCI	COR	GEO	HOS	HQP	JHH	MER	MCD	QMD	QUA	UMD	WSH
Overall Impact of Program on Patients' Quality of Care																	
Increased	67	23.8	82	78	30	95	55	80	84	65	72	78	79	41	11	100	53
Would Recommend Program to Patients or Colleagues																	
Definitely yes	59	22.1	70	67	27	95	59	68	86	48	43	70	70	33	11	79	60
Average *Z*-Value			0.57	0.41	**-1.49**	**1.41**	-0.25	0.47	0.97	-0.29	-0.26	0.48	0.51	**-1.13**	-2.26	**1.15**	-0.27

Source: Mathematica Policy Research, Inc. Physician Survey.

Notes: See Table IV.1 for the full names of the programs.

Estimates in boldface are ones that are more than 1 standard deviation above or below the mean across all sites for that measure. Shaded boldfaced estimates are 1 standard deviation *above* the mean; unshaded boldface estimates are 1 standard deviation *below* the mean.

To calculate the average *Z*-value, the program's *Z*-score for the percentage rating a particular measure favorably was first calculated for each of the individual outcome measures. The score for each measure was calculated by subtracting the mean of the percentage rating the program as favorable across the 15 programs from the percentage rating the measure as favorable in the program, and dividing by the standard deviation across the 15 programs. Next, an average of the *Z*-scores for all of the measures for that program was calculated. Average *Z*-values greater than 1 or less than -1 indicate that a program was, on average, more than 1 standard deviation from the mean across all measures.

V. PROGRAM EFFECTS ON PATIENTS' KNOWLEDGE AND BEHAVIOR, UNMET NEEDS, AND QUALITY OF CARE

Among the objectives of the care coordination programs were increasing patients' knowledge regarding their chronic conditions and improving their self-care behavior, service arrangements to meet functional deficits, and quality of care. Accomplishing these objectives should lead eventually to improvements in health and well-being. This chapter describes the effects of the care coordination programs on measures of patients' education and knowledge, self-care behavior, unmet needs, quality of care, health status, and well-being. The analysis is based on comparisons of outcomes for the treatment and control groups, using patient survey and Medicare claims data.

This chapter does not include the Quality Oncology MCCD program, one of three whose patients were not surveyed due to low enrollment (described further below). In addition, the claims-based measures in this chapter gauge general preventive care; preventive care for diabetes, congestive heart failure (CHF), and coronary artery disease (CAD); and potentially preventable hospitalizations for those conditions. Thus, because Quality Oncology focuses on beneficiaries with cancer, none of those measures apply to the program.

Because of the large number of outcomes examined across the 14 other MCCD programs, the tables in the chapter use plus and minus symbols to summarize treatment-control differences that are significant at the 10-percent level. Appendix C contains the actual numerical estimates of the treatment-control differences for the outcomes discussed.

Overall, the analyses indicate that significantly more treatment than control group members reported receiving education, but that this education did not translate into the hoped-for changes in their diet, exercise, or self-care, or into favorable effects on functional ability. There was also increased recognition of having had help in service arrangement, and there were suggestions of

favorable (that is, favoring the treatment group) effects on measures of quality of care and health status and well-being in a few programs.

A. DATA SOURCES

1. Patient Survey

Measures of patients' receipt of education, knowledge, behavior, adherence, receipt of care, and functioning were obtained from a telephone survey of samples of patients enrolled during the first year of each program's operations. A full description of the patient survey is contained in Chapter IV. As described in Chapter IV, the survey was conducted on only 12 of the 15 MCCD programs, because 3 enrolled too few patients to permit an analysis of impacts based on survey data. Therefore, only 12 programs are included in the analyses of survey data.

2. Medicare Claims Data

Additional measures of quality of care were constructed from Medicare claims data. Some of these measures, such as colon cancer screening or mammography, are general screening tests that are not specific to the targeted population in the program.[2] Receipt of colon cancer screening by participants was actually assessed in two ways, through self-report in the patient survey and through claims data.

Other measures of the quality of care from claims data were specific to certain diseases. For example, it is recommended that patients with diabetes or CAD have periodic blood tests for lipid levels, and that patients with CHF have a measure of left ventricular ejection at least once.[3]

[2] Such general preventive screening may not be appropriate for enrollees with severe comorbidities or shortened life expectancy, but, on average, across all enrollees, one might expect to see increased attention to prevention among the intervention group of a care coordination program, compared with usual care.

[3] Again, measurement of left ventricular ejection may not have been indicated for individual enrollees with CHF during the study period (for example, if they had had a recent test before the study period). On average,

Finally, adverse event outcomes were constructed from the claims data. These measure events that are presumably preventable with high-quality outpatient care. Some were measured in all participants regardless of their diagnoses (for example, hospitalizations for pneumonia or urinary tract infections); others were measured only among beneficiaries with one of the main targeted conditions (for example, hospitalizations for uncontrolled diabetes among beneficiaries with diabetes, or fluid and electrolyte problems among those with CHF).[4,5]

In the analyses of Medicare claims data, first-year enrollees were studied over the first year after the month of their random assignment, excluding any months during which they did not meet demonstration-wide eligibility requirements. Enrollees in all 14 MCCD programs discussed in this chapter were included. An enrollee who had a claim in the Medicare data involving diabetes, CHF, or CAD in the 2 years prior to enrollment was defined as having the condition. The diagnosis categories are not mutually exclusive (that is, the same enrollee could be included in more than one category).

B. RECEIPT OF EDUCATION

It is thought that much morbidity and health care use among people with chronic illness could be reduced if these people were better able to make difficult changes in lifestyle, adhere to complex medication regimens, recognize early signs and symptoms of worsening of their

(continued)
however, across all enrollees with CHF, one might expect to see increased assessment of left ventricular function among the intervention group of a care coordination program, compared with usual care.

[4] The potentially preventable hospitalizations examined among all participants, regardless of diagnosis, were those for CHF, chronic obstructive pulmonary disease, dehydration, and urinary tract infection. The reason for including hospitalizations for CHF is that new-onset CHF occurs frequently among the elderly, even among those without a history of the condition. Good ambulatory care should be able to detect and treat such occurrences before they deteriorate to the point of requiring hospitalization.

[5] The potentially preventable hospitalizations examined among participants with diabetes include a broad range of cardiac problems, as CAD and CHF cause substantial morbidity among people with diabetes. High-quality ambulatory care should be able to decrease that morbidity.

condition, and seek medical care effectively when they need it (Institute of Medicine 2001; Wagner et al. 2001). The programs sought, to varying degrees, to impart to their patients the knowledge and skills necessary for making these behavioral changes.

In 11 of the 12 programs with patient survey data, the treatment group was significantly more likely than the control group to report having been taught about diet, exercise, taking medication, or recognizing warning signs for their conditions, or to report having received materials explaining their conditions or treatment (Table V.1). These findings were consistent with the emphasis of these programs' interventions on patient education. In particular, CenVaNet, Health Quality Partners, and Mercy had treatment-control differences favoring the treatment group in all five measures of receipt of education. In the Carle program, more treatment group members than control group members reported receiving education on four of the five education measures. Jewish Home and Hospital, which focused on reducing social isolation, was the only program in which participants did not report any significant treatment-control differences in health education.

The health topic with the most widespread impacts was diet. In 9 of the 12 programs, a higher proportion of treatment group members than control group members reported receiving education on how to follow a healthy diet (Table V.1). Treatment-control group differences favoring the treatment group ranged from a minimum of 10 percentage points (CorSolutions) to a high of 52 percentage points (Health Quality Partners). Hospice of the Valley, Jewish Home and Hospital, and Washington University were the only programs in which the treatment groups did not report receiving more education than the control groups about diet.

The health topic with the next highest number of impacts was exercise. In 7 of the 12 programs, more treatment group members reported receiving education on exercise. Treatment group means exceeded the control group means by a minimum of 8 percentage points

TABLE V.1

SUMMARY OF TREATMENT-CONTROL DIFFERENCES IN PARTICIPANTS' SELF-REPORTED RECEIPT OF HEALTH EDUCATION

Program	Reported Being Taught How to: Follow a Healthful Diet	Exercise	Take Medication	Recognize Warning Signs to Seek Urgent Care	Reported Receiving Materials About Condition or Treatment
Avera	++				++
Carle	++	++		++	++
CenVaNet	++	++	+	++	++
Charlestown	++	+			
CorSolutions	++				++
Hospice of the Valley				+	
Health Quality Partners	++	++	++	++	++
Jewish Home and Hospital					—
Medical Care Development	++	+			
Mercy	++	+	+	++	++
QMed	++				
Washington University		+			—

Source: Patient survey of treatment and control group beneficiaries participating in the programs, conducted by Mathematica Policy Research, Inc. from May 2003 through June 2004. Sample sizes for each program ranged from 395 to 684. The mean duration of participation at the time of interview was 301 days. Response rates ranged from 85 to 98 percent.

Note: Only 12 programs are shown. The patient survey was not conducted in three programs whose enrollments were too small to allow survey-based treatment-control comparisons with acceptable power.

+Denotes statistically significant treatment-control difference ($p \leq 0.10$) that favors the treatment group and is modest (less than 10 percentage points and less than one-half the size of the control group proportion [*pc*] or its complement [1-*pc*]).

++Denotes statistically significant treatment-control difference ($p \leq 0.10$) that favors the treatment group and is large (more than 10 percentage points or at least one-half the size of the control group proportion [*pc*] or its complement [1-*pc*]).

—Denotes statistically significant treatment-control difference ($p \leq 0.10$) of any size that favors the control group.

(in the Charlestown program) and by a maximum of 34 percentage points (for the Health Quality Partners program) (Table V.1).

Treatment group members in several programs were also more likely to report having been taught how to recognize when to seek urgent care and how to take medications (Table V.1). Of the 12 programs, 5 had large treatment-control differences in the proportion who reported being taught how to recognize warning signs to seek urgent care, with 4 of these differences substantial—between 11 and 21 percentage points. The treatment groups in three programs were more likely to report being taught how and when to take medications (with treatment-control differences of 6, 8, and 14 percentage points).

Treatment group members in six programs were much more likely to have received materials explaining their conditions or treatment. In two programs, however, treatment-control differences actually favored the control group in this measure (Table V.1).

C. HEALTH KNOWLEDGE AND HEALTH BEHAVIORS

If chronically ill people are to make significant changes in their health-related behavior, mere receipt of education is generally not enough. They must also understand the information they receive and must apply it to their daily lives. The evaluation survey also had respondents self-assess their knowledge of healthy eating, exercising, medication adherence, and when to seek care for specific symptoms, as well as their adherence to regimens of diet, exercise, and medication.

Despite a higher proportion of the treatment groups reporting receiving health-related education, there were few detectable treatment-control differences in self-reported knowledge or health-related behavior. For the four outcomes of (1) understanding how to follow a healthy diet, (2) following a healthy diet, (3) understanding how to exercise, and (4) exercising regularly, there was only one favorable treatment-control difference for each outcome, with the differences

scattered across four different programs (Table V.2). The absence of significant differences may reflect an already high baseline level of knowledge among participants. In all programs, 76 to 97 percent of control group members said they knew how to follow a healthy diet, and 66 to 85 percent said they knew the proper way to exercise. It may thus have been difficult for programs to effect large improvements for these outcomes among the treatment groups.

There were also a few treatment-control differences in trying to quit smoking and trying to cut down on drinking (Table V.2). However, sample sizes for these questions were very small, as the questions were asked only of the very few beneficiaries who had reported smoking or drinking. The treatment-control differences for these measures were mixed, with two favoring the treatment group and two favoring the control group.

D. UNMET NEEDS

Chronically ill beneficiaries commonly face physical and financial barriers that can lead to adverse health outcomes and increased use of health care. Some beneficiaries have functional limitations that complicate such essential activities as eating, dressing, bathing, using the telephone, shopping, and traveling to medical appointments. In some cases, financial constraints make it hard for beneficiaries to afford critical medications. Personal care or financial assistance services can help to lower these barriers.

As described in Chapter II, 9 of the 12 programs sought to improve the provision of Medicare- or non-Medicare-covered services. Some programs paid for non-Medicare-covered services, such as home health or prescription drugs, whereas others helped patients to gain access to community resources, such as meal delivery services or transportation to their physicians' offices.

Treatment group members' recognition that they had care coordinators was very high across all programs, as evidenced by large treatment-control differences across programs in participants

TABLE V.2

SUMMARY OF TREATMENT-CONTROL DIFFERENCES IN PARTICIPANTS' SELF-REPORTED KNOWLEDGE AND BEHAVIOR

Program	Understands Diet	Follows Healthful Diet	Understands Exercise	Exercises Regularly	Misses Doses of Medications	Visits Physician with List of Questions	Tried to Quit Smoking	Tried to Cut Down on Drinking
Avera						—		
Carle								
CenVaNet	+							
Charlestown								
CorSolutions								
Hospice of the Valley								
Health Quality Partners				+				++
Jewish Home and Hospital							—	
Medical Care Development			+					
Mercy								
QMed							++	—
Washington University		++						

Source: Patient survey of treatment and control group beneficiaries participating in the programs, conducted by Mathematica Policy Research, Inc. from May 2003 through June 2004. Sample sizes for each program ranged from 395 to 684. The mean duration of participation at the time of interview was 301 days. Response rates ranged from 85 to 98 percent.

Note: Only 12 programs are shown. The patient survey was not conducted in three programs whose enrollments were too small to allow survey-based treatment-control comparisons with acceptable power.

+Denotes statistically significant treatment-control difference ($p \leq 0.10$) that favors the treatment group and is modest (less than 10 percentage points and less than one-half the size of the control group proportion [*pc*] or its complement [1-*pc*]).

++Denotes statistically significant treatment-control difference ($p \leq 0.10$) that favors the treatment group and is large (more than 10 percentage points or at least one-half the size of the control group proportion [*pc*] or its complement [1-*pc*]).

—Denotes statistically significant treatment-control difference ($p \leq 0.10$) of any size that favors the control group.

Understands diet = understands how to follow a healthful diet.
Follows healthful diet = follows healthful eating plan most or all of the time.
Understands exercise = understands proper way to exercise.
Exercises regularly = gets regular exercise.
Misses doses of medications = misses dose of medication twice or more a week.
Visits physician with list of questions = visits physician with list of questions most or all of the time.
Tried to quit smoking = if beneficiary reported smoking in the past 6 months, tried to quit.
Tried to cut down on drinking = if beneficiary reported drinking in the past year, tried to cut down.

reporting having received help in arranging care (Table V.3). Effects on participants' unmet needs were less dramatic, with a few differences favoring the treatment group across the programs (Table V.3).

The number of people who needed assistance with any of these activities was very small, however, and even then, the great majority in the control group reported being able to get the help they needed. It would have been difficult for the programs to have improved matters much.

E. QUALITY OF CARE

Many of the programs sought to improve the quality of their patients' care, most by teaching patients to assume more responsibility for keeping track of their own care, and to advocate for themselves. For example, a program might instruct a patient on the target values for cholesterol levels or blood pressure and might encourage the patient to monitor results, to remind the physician when to retest, and to ask the physician to address unsatisfactory numbers. Program care coordinators might also call physicians on a case-by-case basis to suggest medication adjustments or monitoring tests in keeping with national evidence-based guidelines. Few programs attempted to improve physician behavior directly by giving physicians feedback on their practice patterns compared with those of peers or with those recommended by care guidelines.

The summary of general and disease-specific preventive care measures in Table V.4 suggests that the Carle and Health Quality Partners programs had some favorable effects, as each program had four favorable treatment-control differences. The Carle program had moderate to large treatment-control differences in disease-specific preventive measures (testing for cholesterol, hemoglobin A1c, and urine protein in beneficiaries with diabetes, and testing for cholesterol in beneficiaries with CAD). The Health Quality Partners program had two favorable

TABLE V.3

SUMMARY OF TREATMENT-CONTROL DIFFERENCES IN PARTICIPANTS' SELF-REPORTED SERVICE ARRANGEMENT AND UNMET NEEDS

Program	Received Help in Arranging Care[a]	Among Beneficiaries Unable to Do by Themselves, Reported Being Able to Get Help with:						
		Telephone	Transportation	Shopping	Preparing Meals	Housework	Taking Medication	Handling Money
Avera	++							
Carle	++							
CenVaNet	++							
Charlestown	++	++						
CorSolutions	++	++		++				
Hospice of the Valley	++					++		
Health Quality Partners	++							
Jewish Home and Hospital	++							
Medical Care Development	++							
Mercy	++		—					
QMed	++							—
Washington University	++							

Source: Patient survey of treatment and control group beneficiaries participating in the programs, conducted by Mathematica Policy Research, Inc. from May 2003 through June 2004. Sample sizes for each program ranged from 395 to 684. The mean duration of participation at the time of interview was 301 days. Response rates ranged from 85 to 98 percent.

Note: Only 12 programs are shown. The patient survey was not conducted in three programs whose enrollments were too small to allow survey-based treatment-control comparisons with acceptable power.

[a]Whether a nurse, care coordinator, or social worker helped arrange care.

+Denotes statistically significant treatment-control difference ($p \leq 0.10$) that favors the treatment group and is modest (less than 10 percentage points and less than one-half the size of the control group proportion [*pc*] or its complement [1-*pc*]).

++Denotes statistically significant treatment-control difference ($p \leq 0.10$) that favors the treatment group and is large (more than 10 percentage points or at least one-half the size of the control group proportion [*pc*] or its complement [1-*pc*]).

—Denotes statistically significant treatment-control difference ($p \leq 0.10$) of any size that favors the control group.

TABLE V.4

SUMMARY OF TREATMENT-CONTROL DIFFERENCES IN GENERAL AND DISEASE-SPECIFIC PREVENTIVE CARE FOR PARTICIPANTS

	Non-Disease-Specific or General Preventive Measures					Disease-Specific Preventive Measures for Beneficiaries with:						
						Diabetes					CHF	CAD
	Survey-Based Measures			Claims-Based Measures		Claims-Based Measures[a]						
Program	Flu Vaccine	Pneu-monia Vaccine	Colon Cancer Screening[b]	Colon Cancer Screening[c]	Screening Mammo-graphy[d]	Diabetes Education[e]	Eye Exami-nation	Choles-terol or Lipid Test	Hemo-globin A1c Test	Urine Test for Protein	LV Function Test	Choles-terol or Lipid Test
Avera										—	—	
Carle								+	+	++		++
CenVaNet	+				—							
Charlestown												
CorSolutions												
Georgetown												
Hospice of the Valley		—			—							
Health Quality Partners		+			++			++				+
Jewish Home and Hospital												
Medical Care Development												
Mercy												
QMed												
University of Maryland[f]												
Washington University							—					

Source: For survey-based measures, patient survey of treatment and control group beneficiaries participating in the programs, conducted by Mathematica Policy Research, Inc. from May 2003 through June 2004. Sample sizes for each program ranged from 395 to 684. The mean duration of participation at the time of interview was 301 days. Response rates ranged from 85 to 98 percent. For claims-based measures, Medicare National Claims History File. First-year enrollees over the 1st year after random assignment in 14 programs were included, excluding any months during which an enrollee did not meet demonstration-wide eligibility requirements.

Note: Quality Oncology is not included. The patient survey was not conducted for Quality Oncology because its enrollment was too small to allow survey-based treatment-control comparisons with acceptable power. The claims-based measures in this table were not analyzed for Quality Oncology because it focuses on beneficiaries with cancer, and the claims-based measures of general preventive care and preventive care for diabetes, CHF, and CAD in this table were not appropriate for the program.

[a]Enrollees were defined as having diabetes, CHF, or CAD through Medicare claims data. An enrollee who had a claim with one of these diagnoses in the 2 years prior to enrollment was defined as having the condition. The diagnosis categories are not mutually exclusive (that is, the same enrollee could be included in more than one category).

[b]Reported having had a stool blood test, sigmoidoscopy, or colonoscopy in the past year.

[c]Fecal occult blood testing, screening colonoscopy, sigmoidoscopy, or barium enema.

[d]Females only.

[e]Any claims for individual or group diabetes outpatient self-management training services, or for education/training services, including diabetes diet training.

[f]The patient survey was not conducted for Georgetown and the University of Maryland because their enrollments were too small to allow survey-based treatment-control comparisons with acceptable power.

TABLE V.4 *(continued)*

+Denotes statistically significant treatment-control difference ($p \leq 0.10$) that favors the treatment group and is modest (less than 10 percentage points and less than one-half the size of the control group proportion [*pc*] or its complement [1-*pc*]).
++Denotes statistically significant treatment-control difference ($p \leq 0.10$) that favors the treatment group and is large (more than 10 percentage points or at least one-half the size of the control group proportion [*pc*] or its complement [1-*pc*]).
—Denotes statistically significant treatment-control difference ($p \leq 0.10$) of any size that favors the control group.

CAD = coronary artery disease; CHF = congestive heart failure; LV Function Test = contrast left ventriculographic, nuclear medicine, or echocardiographic tests for assessment of left ventricular function.

differences in general preventive care (pneumonia vaccination and screening mammography) and two in disease-specific care (cholesterol testing in diabetes and in CAD).[6]

Some of the programs may have had some effects on the rates of general and disease-specific potentially preventable hospitalizations (Table V.5).[7] On the one hand, the CenVaNet program had a modest favorable difference of 4 percentage points on the rate of all potentially preventable hospitalizations (compared with a control group mean of 16 percent) and a large difference of 4 percentage points on the rate of potentially preventable CHF hospitalizations among patients with diabetes (relative to a control group mean of 10 percent). The Georgetown program had a large treatment-control difference of 0.48 hospitalizations per beneficiary for all potentially preventable hospitalizations (with a control group mean of 0.93 hospitalizations per beneficiary) and also a large difference of 26 percentage points for the rate of potentially preventable CHF hospitalizations among patients with diabetes (with a control group mean of 48 percent). For all potentially preventable hospitalizations, the Hospice of the Valley program had a small to moderate treatment-control difference of 0.1 hospitalizations per beneficiary against a control group mean of 0.4 hospitalizations per beneficiary. Finally, in the Health Quality Partners program, the treatment group had a potentially preventable CAD hospitalization rate among patients with diabetes of 0 percent, whereas the control group had a rate of 8 percent.

[6] Again, as noted previously in the discussion in Chapter II of the scoring ratings, and in the presentation of the physician survey results in Chapter IV, some of the programs' interventions intentionally did not include certain features. Many of the programs did not focus on general, non-disease-specific preventive care.

[7] Table V.5 summarizes results for treatment-control differences for both rates of potentially preventable hospitalizations (that is, rates of whether any such hospitalization occurred in each group) and rates of average number of potentially preventable hospitalizations (that is, the total number of such hospitalizations divided by the number of beneficiaries in the treatment or control group). A difference in rates of hospitalization is considered modest if it is less than 10 percentage points and less than one-half the size of the control group proportion [*pc*] or its complement [1-*pc*]), and large if it is more than 10 percentage points or at least one-half the size of the control group proportion [*pc*] or its complement [1-*pc*]). A difference in average number of hospitalizations is considered modest if it is less than one-half of the average number for the control group, and large if at least one-half that of the control group.

TABLE V.5

SUMMARY OF TREATMENT-CONTROL DIFFERENCES IN GENERAL AND DISEASE-SPECIFIC POTENTIALLY PREVENTABLE HOSPITALIZATIONS AMONG PARTICIPANTS

		Among Participants with:						
	All Participants	Diabetes				CHF		CAD
Program	Hosp. for Potentially Preventable Conditions[a]	Hosp. for CAD[b]	Hosp. for Diabetes[c]	Hosp. for CHF	Micro-vascular Comp.[d]	Hosp. for Fluid/Electro-lyte Problem[e]	Hosp. for CHF	Hosp. for CAD[b]
Avera						—		
Carle								
CenVaNet	+			++	—			
Charlestown								
CorSolutions								
Georgetown	++			++				
Hospice of the Valley	+	—						
Health Quality Partners		++						
Jewish Home and Hospital								
Medical Care Development						—		
Mercy								
QMed								
University of Maryland								
Washington University	—		—					

Source: Medicare National Claims History File, Standard Analytic File, and Enrollment Database. First-year enrollees over the 1st year after random assignment in the 14 programs were included, excluding any months during which an enrollee did not meet demonstration-wide eligibility requirements.

Notes: Quality Oncology is not included. Quality Oncology focuses on beneficiaries with cancer, and the claims-based measures this table for general potentially preventable hospitalizations and potentially preventable hospitalizations for diabetes, CHF, and CAD were not appropriate for the program.

Enrollees were defined as having diabetes, CHF , or CAD through Medicare claims data. An enrollee who had a claim with one of these diagnoses in the 2 years prior to enrollment was defined as having the condition. The diagnosis categories are not mutually exclusive (that is, the same enrollee could be included in more than one category). This table summarizes results for treatment-control differences for both rates of potentially preventable hospitalizations (that is, rates of whether any such hospitalization occurred in each group) and average number of potentially preventable hospitalizations (that is, the total number of such hospitalizations divided by the number of beneficiaries in the treatment or control group). The average number of hospitalizations per beneficiary may be considered the more important outcome as it is a stronger determinant of hospital costs and is a more informative indicator of adverse outcomes. Restricting attention to the average number of hospitalizations per beneficiary, only the Georgetown and Hospice of the Valley programs had favorable treatment-control differences.

[a]Any hospitalizations for CHF, chronic obstructive pulmonary disease, dehydration, or urinary tract infection.

[b]Any hospitalizations for acute myocardial infarction, coronary artery bypass graft surgery, percutaneous transluminal angioplasty, or coronary artery stenting.

[c]Any hospitalizations for diabetes with hyperosmolarity, diabetes with ketoacidosis, diabetes with other (non-hyperosmolar and non-ketotic) complications, diabetes with other (non-hyperosmolar and non-ketotic) coma, or diabetes without mention of complications.

[d]Any hospitalizations, claims, or change in enrollment status for diabetic eye disease, laser treatment for diabetic eye disease, nephropathy, or new end-stage renal disease.

[e]Any hospitalizations for hyperkalemia, hypernatremia, hypokalemia, hyponatremia, or other fluid/electrolyte problems.

+Denotes statistically significant treatment-control difference ($p \leq 0.10$) that favors the treatment group and is modest (for rates of hospitalization, less than 10 percentage points and less than one-half the size of the control group proportion [*pc*] or its complement [1-*pc*], or for average number of hospitalizations is less than one-half the average number for the control group).

++Denotes statistically significant treatment-control difference ($p \leq 0.10$) that favors the treatment group and is large (for rates of hospitalization, more than 10 percentage points or at least one-half the size of the control group proportion [*pc*] or its complement [1-*pc*], or for average number of hospitalizations is less than one-half the average number for the control group).

—Denotes statistically significant treatment-control difference ($p \leq 0.10$) of any size that favors the control group.

CAD = coronary artery disease; CHF = congestive heart disease; Hosp. For = hospitalization for; Microvascular Comp. = microvascular complications.

On the other hand, there were also four treatment-control differences favoring the control group in four programs (one of which was also one of the programs with a difference favoring the treatment group).

The average number of hospitalizations per beneficiary may be considered the more important outcome, since it is a stronger determinant of hospital costs and is a more informative indicator of adverse outcomes. Restricting attention to the average number of hospitalizations per beneficiary, only the Georgetown and Hospice of the Valley programs had favorable treatment-control differences.

F. FUNCTIONING, HEALTH STATUS, AND WELL-BEING

Finally, one of the ultimate goals of the demonstration programs is the improvement of patients' functioning, health status, and well-being. As noted, it was hoped that the programs would improve patients' self-management of chronic illnesses, reduce unmet needs, and improve the quality of care, and that these changes would lead to increased functioning and health.

Many of the treatment-control differences in functioning actually favored the control group (Table V.6), but they were scattered across several different outcomes and programs and of small magnitude (with all treatment-control differences less than 9 percentage points, all but one less than 7 percentage points, and control group means in the 70- to 90-percent range). This evidence suggests no true underlying program effects. Furthermore, there are no plausible mechanisms for why program interventions would have led to isolated lower rates among the treatment group in ability to eat or prepare meals independently, for example, without affecting any other functioning outcomes.

A few of the programs may have led to some positive effects on self-reported measures of health status and well-being (Table V.7). Two in particular, the CorSolutions program and the

TABLE V.6

SUMMARY OF TREATMENT-CONTROL DIFFERENCES IN MEASURES OF PARTICIPANTS' FUNCTIONING: SELF-REPORTED ABILITY TO PERFORM SELECTED ACTIVITIES OF DAILY LIVING INDEPENDENTLY

Program	Eat	Dress	Bathe	Use Telephone	Prepare Meals	Do Housework	Shop	Take Medications	Handle Money
Avera		—							
Carle				—				—	—
CenVaNet									
Charlestown					—				
CorSolutions	—								
Hospice of the Valley									
Health Quality Partners									
Jewish Home and Hospital			—						
Medical Care Development				+					
Mercy			+						
QMed			+						
Washington University					—				

Source: Patient survey of treatment and control group beneficiaries participating in the programs, conducted from May 2003 through June 2004. Sample sizes for each program ranged from 395 to 684. The mean duration of participation at the time of interview was 301 days. Response rates ranged from 85 to 98 percent.

Note: Only 12 programs are shown. The patient survey was not conducted in three programs whose enrollments were too small to allow survey-based treatment-control comparisons with acceptable power.

+Denotes statistically significant treatment-control difference ($p \leq 0.10$) that favors the treatment group and is modest (less than 10 percentage points and less than one-half the size of the control group proportion [*pc*] or its complement [1-*pc*].

++Denotes statistically significant treatment-control difference ($p \leq 0.10$) that favors the treatment group and is large (more than 10 percentage points or at least one-half the size of the control group proportion [*pc*] or its complement [1-*pc*].

—Denotes statistically significant treatment-control difference of any size ($p \leq 0.10$) that favors the control group.

TABLE V.7

SUMMARY OF TREATMENT-CONTROL DIFFERENCES IN PARTICIPANTS' SELF-REPORTED HEALTH STATUS AND WELL-BEING

	Emotional Distress[a]	Depression[b]	Poor Sleep[c]	Pain[d]	Effect of Primary Condition on Life[e]	Effect of Primary Condition on Family[f]	Physical Health Summary Score[g]	Mental Health Summary Score[g]
Avera			+					
Carle								
CenVaNet					+			
Charlestown						—		
CorSolutions	+					+	+	
Hospice of the Valley				+		++		
Health Quality Partners	+							
Jewish Home and Hospital			+					
Medical Care Development						—		
Mercy	+							
QMed				+				
Washington University								

Source: Patient survey of treatment and control group beneficiaries participating in the programs, conducted from May 2003 through June 2004. Sample sizes for each program ranged from 395 to 684. The mean duration of participation at the time of interview was 301 days. Response rates ranged from 85 to 98 percent.

Note: Only 12 programs are shown. The patient survey was not conducted in three programs whose enrollments were too small to allow survey-based treatment-control comparisons with acceptable power.

+Denotes statistically significant treatment-control difference ($p \leq 0.10$) that favors the treatment group and is modest (less than 10 percentage points and less than one-half the size of the control group proportion [*pc*] or its complement [1-*pc*]).

++Denotes statistically significant treatment-control difference ($p \leq 0.10$) that favors the treatment group and is large (more than 10 percentage points or at least one-half the size of the control group proportion [*pc*] or its complement [1-*pc*]).

—Denotes statistically significant treatment-control difference of any size ($p \leq 0.10$) that favors the control group.

[a]Feeling calm and peaceful most or all of the time in the past 4 weeks.

[b]Feeling downhearted and blue most or all of the time in the past 4 weeks.

[c]Bothered by poor sleep most or all of the time in the past 4 weeks.

[d]Pain interfered with usual activities in the past 4 weeks.

[e]Primary condition interfered a lot or somewhat with enjoyment of life in the past 4 weeks.

[f]Feeling that primary condition placed a burden on family in the past 4 weeks.

[g]Physical and mental health summary scores from the 12-item Short Form Health Status questionnaire (SF-12).

Hospice of the Valley program, had more than one favorable effect. Treatment group members in the CorSolutions program had better results on measures of emotional distress, burden of primary condition on family members, and a summary score of physical health. Treatment group members in the Hospice of the Valley program had better results on measures of pain and family burden. Six other programs also had isolated favorable effects on a single outcome.

G. SUMMARY

The strongest program effect was on patients' recognition of having received health education. Most programs showed a clear pattern of treatment group members reporting having received instruction on health behaviors, particularly diet and exercise, as well as educational materials. These results were generally consistent with the quantitative scores of the programs' patient education efforts described in Chapter II. Two of the programs that had favorable treatment-control differences across all five measures of patient education in Table V.1 (the Health Quality Partners and Mercy programs) also received the highest and second-highest scores for patient education (Table II.6). The sole program with no treatment-control differences in any of the patient education measures in Table V.1 (Jewish Home and Hospital) was the lowest-scoring program in Table II.6. The correspondence between program scores and survey results on patient education was inexact, however, as some of the high-scoring programs (CorSolutions, Hospice of the Valley, and Avera, for example) were among those with few differences on the survey-based patient education measures.

Another obvious treatment effect across programs was in patients' recognition of care coordinator help in arranging services. Treatment groups for 11 of the 12 programs with patient survey data were much more likely to report having received help in arranging care. These results are largely consistent with the rating scores for Service and Resource Arranging assigned to the programs, as the 11 programs are ranked 2nd through 12th on these scores (Table II.6).

The Washington University program is an exception, as it was the top-scoring program for Service and Resource Arranging but had no significant treatment-control differences in any of the service arrangement and unmet needs outcomes summarized in Table V.3.

Unfortunately, there were no clear-cut effects across the programs for the key outcomes of health knowledge and behavior, potentially preventable hospitalizations, or physical functioning. The CenVaNet, Georgetown, and Health Quality Partners programs may have had some favorable effects in the potentially preventable hospitalization outcomes. However, the pattern of scattered treatment-control differences distributed across many outcomes and many programs does not provide strong evidence that either the programs as a group or any single or few programs had widespread effects in any of these areas.

There are suggestions that the Carle and Health Quality Partners programs had some effects on general and disease-specific preventive care, such as vaccination, screening mammography, and recommended blood and urine tests for beneficiaries with diabetes or CAD (Table V.4). The correlation between the results for these outcomes and the Improving Provider Practice scores was poor, however. The Carle program had the highest score for Improving Provider Practice, but the Health Quality Partners program had one of the lowest.

Finally, there are also suggestions that some of the programs had favorable effects on measures of health status and well-being, particularly the CorSolutions and Hospice of the Valley programs. These two programs had high scores for Problem Identification and Care Planning and for Patient Education; strong performance in these two areas might lead to increased detection and help for patients with psychosocial distress, which is what many of the health status and well-being questions measure.

The outcomes discussed in this chapter are subject to measurement "noise," or biases, that could make detecting true program effects difficult. For example, for the survey-based

measures, it is possible that the treatment group members increased their health knowledge through the programs and, as a result, became more stringent in their self-assessments—they might have reported their health knowledge and adherence as low even though it was superior to the control group's. Furthermore, the measures of quality of care in claims-based processes of care are not appropriate for all beneficiaries, and many factors besides the intervention can affect potentially preventable hospitalizations.

With longer followup of the demonstration, the enhanced patient recognition of having received teaching may translate into detectable behavioral changes and measurable effects on health care use and costs. Improvements in provider quality of care may persist and become broader-based across programs, and thus contribute to positive effects on health care use. Finally, the programs may adjust their interventions and refocus their efforts to increase program effectiveness.

VI. SHORT-TERM EFFECTS ON MEDICARE EXPENDITURES AND SERVICE USE

A key goal of the demonstration is to determine whether the programs reduce Medicare expenditures and service use and, if so, whether such reductions are enough to offset the increased costs of providing care coordination. The evaluation estimates the effects of the interventions on these outcomes in two ways: (1) for beneficiaries who enrolled over the first 12 months of program operations during the 12 months following their month of random assignment; and (2) each month, over the first 25 calendar months of program operations for beneficiaries randomized during that time. The evaluation compares the treatment and control groups in each program separately and for the programs together on mean outcomes, but only statistically significant differences are taken as evidence that the intervention caused the difference. Treatment-control differences in the use of hospital, physician, emergency room (ER), and imaging and procedure services, as well as in Medicare expenditures and mortality, are described. Differences in the use of skilled nursing facilities, hospice, home health services, and durable medical equipment were also examined, but they did not indicate any evidence of major effects. The report therefore presents only those key outcomes most likely to be influenced by the intervention.

Because of the high variability of these measures in the Medicare population, detecting differences in service use and expenditures is difficult for the samples enrolled in each program. Many programs had unexpected difficulties with enrollment and fell short of their targets, which also leads to less precise estimates than desired. For 10 of the programs, there is at least 70 percent power (or probability) to detect a 20-percent reduction (or increase) in Medicare expenditures in the 25-month sample. Thus, smaller, but real, program effects on expenditures may not be detected. For the three smallest programs, the evaluation has only about a 30-percent chance of detecting a 20-percent reduction in expenditures.

The evidence presented here indicates that, for the average number of hospitalizations and Medicare expenditures over the year after the randomization month and over the first 25 months of program operations, only one of the programs had a statistically significant effect that favored the treatment group. None of the programs had significant treatment-control differences in either the number of hospitalizations or the average monthly Medicare expenditures in the first year after randomization. The treatment group in the program with a favorable effect—Mercy—had 27 percent fewer hospitalizations per beneficiary per year than the control group over the first 25 months (0.73 versus 1.01, $p = 0.003$). Average Medicare expenditures per month were 13 percent lower, or $154 less, than those of the control group ($p = 0.105$). The treatment groups in four other programs—Quality Oncology, Hospice of the Valley, Georgetown University, and Health Quality Partners—had 12 to 18 percent fewer hospitalizations over the 25 months, but these differences were not statistically significant. Only Georgetown had a concomitant treatment-control difference in Medicare expenditures (of 12 percent), but this, too, was not statistically significant. Across all 15 programs, hospitalizations were 4 percent lower for the treatment than for the control group ($p = 0.145$) and Medicare expenditures were 2 percent less ($p = 0.368$). Turning to whether the programs are cost neutral, the 15 programs combined are not. There is some evidence that five of the programs *might* be cost neutral over the first 25 months of operations. However, because of large variations in Medicare expenditures, more follow-up time is needed to determine this conclusively.

A. TREATMENT-CONTROL DIFFERENCES IN MEDICARE EXPENDITURES AND SERVICE USE

Treatment-control differences were estimated in two ways: (1) through comparison of outcomes during the 12 months following the month of random assignment for all beneficiaries who were randomized during the program's first year of operations, and (2) through examination

of how cost-effectiveness might vary over a longer follow-up period by estimating cumulative monthly impacts over the first 25 calendar months of program operations.[8] For the 25-month analysis, the patients who were enrolled in the program through the first 25 months were identified and the average monthly Medicare-covered expenditures and hospitalizations for all patient-months through that month were analyzed. For example, a beneficiary who was randomized in August 2002 and died on October 31, 2002, would contribute 3 patient-months to the cumulative total (for August, September, and October of 2002), provided he or she had been eligible in each month.[9]

Because these measures, especially Medicare expenditures, are highly variable, it is important to note that there can be sizable differences due to chance between the treatment and

[8] The evaluation began measuring Medicare expenditures and service use for this analysis in the first full month after random assignment. For example, for a beneficiary randomized on September 15, 2002, 1-year outcomes are calculated from October 1, 2002, through September 30, 2003. The evaluation examines expenditures and service use over 12 *calendar* months because basic eligibility is assessed on a calendar-month basis. The month of randomization is omitted because programs are not expected to alter service use during the first month. As a sensitivity test, two key outcomes, expenditures and hospital use, were calculated over the year starting from the date of random assignment. The findings were nearly identical.

[9] The estimates for both analyses exclude Medicare expenditures and service use during months when the beneficiary did not meet basic insurance and coverage requirements for the demonstration specified by the Centers for Medicare & Medicaid Services (CMS) (has both Part A and B coverage, has Medicare as the primary payer, is not in a health maintenance organization, and is alive for some part of the month), because we could not fully observe the outcomes during those months.

For all outcome measures except mortality, observations are weighted to reflect the length of time the patient was eligible for the study. For binary outcomes covering a 1-year follow-up period (such as had any hospitalizations, used physician services), a separate weight was constructed for each outcome. A person who is observed (that is, eligible) for the full follow-up time receives a weight of 1. Similarly, a person who is ineligible for some months but experienced the outcome during the observed period (for example, was hospitalized) receives a weight of 1. A person who is ineligible for some months and did not experience the outcome during the observed follow-up period receives a weight equal to the number of follow-up months observed, divided by 12. For example, someone observed for 3 months of a 12-month follow-up period receives a weight of 0.25. For continuous outcomes, such as costs and the number of visits, the weight is calculated as the proportion of the follow-up period that is observed. The last step in weighting is to normalize the weights so they sum to the number of observations.

The payment for the intervention was calculated as the amount CMS paid to the program for treatment group patients, using claims from the physician claims file with "G" codes.

Because of rounding, the column in all tables reporting treatment-control differences may differ slightly from the result when the control column is subtracted from the treatment column.

control groups, both before and after enrollment. Whether the difference is likely to be due to the program is determined by a test of its statistical significance, which takes into account the size of the estimate, the sample size, and how varied the values are. A difference that is not statistically significant indicates that the treatment and control groups are comparable—that is, that the observed difference is well within the range that might be expected simply as a result of chance and should not be attributed to the program.

Any chance differences observed between the treatment and control groups on preenrollment characteristics were accounted for through use of regression models to adjust the estimates of three key outcomes—the proportion with a hospitalization, the average number of hospitalizations, and total Part A and Part B expenditures. The regressions controlled for age; gender; whether the beneficiary had been treated for congestive heart failure (CHF) during the 2 years before randomization (in programs that did not exclusively target CHF); the number of the following conditions the patient had been treated for during the two years before randomization: coronary artery disease, CHF, stroke, diabetes, cancer, chronic obstructive pulmonary disease, dementia, peripheral vascular disease, end-stage renal disease, depression, and asthma; the annualized number of hospital admissions in the previous year; and total Medicare Part A and Part B expenditures per month in the prior year.

The findings show that the programs treated patients with very different levels of risk for hospital utilization and expenditures. This is consistent with the differences in populations served, described in Chapter III. About 43 percent of the control group was hospitalized over the year after enrollment, ranging from a low of about 20 percent in Health Quality Partners and QMed to a high of nearly two-thirds in CorSolutions and the University of Maryland (Table VI.1). During the first 25 months, control group members in Health Quality Partners and QMed had experienced an annualized average of 0.4 hospitalizations per year, whereas those in

TABLE VI.1

HOSPITAL USE FOR 1ST-YEAR ENROLLEES DURING THE YEAR AFTER RANDOM ASSIGNMENT
(Regression Adjusted)

	Sample Sizes		Any Admission (Percent)[a]					Average Annualized Number of Admissions per Year[a]				
	Treatment Group	Control Group	Treatment Group	Control Group	Treatment-Control Difference	Percent Change	*p*-Value	Treatment Group	Control Group	Treatment-Control Difference	Percent Change	*p*-Value
Avera	157	158	56.6	63.8	-7.2	-11	0.184	1.55	1.47	0.07	5	0.765
Carle	1,024	1,018	29.5	31.5	-2.0	-6	0.310	0.51	0.56	-0.05	-9	0.294
CenVaNet	512	509	35.7	38.9	-3.2	-8	0.266	0.77	0.74	0.03	4	0.696
Charlestown	195	189	47.7	46.6	1.2	2	0.805	0.87	0.78	0.08	11	0.538
CorSolutions	354	265	64.0	66.4	-2.4	-4	0.517	1.86	2.03	-0.17	-9	0.361
Georgetown	51	53	65.7	61.5	4.2	7	0.633	2.02	1.59	0.43	27	0.336
Health Quality Partners	219	219	19.3	20.1	-0.8	-4	0.821	0.34	0.39	-0.05	-12	0.610
Hospice of the Valley	222	210	58.0	61.8	-3.8	-6	0.407	1.34	1.51	-0.17	-12	0.392
Jewish Home and Hospital	253	252	44.0	35.8	8.2*	23	0.052	0.88	0.76	0.13	17	0.335
Medical Care Development	192	192	56.3	63.3	-6.9	-11	0.150	1.48	1.51	-0.04	-2	0.878
Mercy	304	308	43.1	50.2	-7.1*	-14	0.075	0.80	0.95	-0.15	-16	0.251
QMed	633	625	21.1	23.3	-2.2	-9	0.377	0.35	0.37	-0.01	-4	0.785
Quality Oncology	29	31	42.1	61.4	-19.3	-31	0.110	1.17	1.76	-0.59	-33	0.349
University of Maryland	29	26	61.8	65.5	-3.8	-6	0.751	2.65	2.48	0.17	7	0.855
Washington University	698	690	55.3	55.2	0.1	0	0.975	1.49	1.39	0.10	7	0.368
Overall	**4,872**	**4,745**	**40.6**	**42.8**	**-2.2****	**-5**	**0.017**	**0.91**	**0.93**	**-0.03**	**-3**	**0.404**

Source: 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 1-year follow-up period, or who had an invalid Health Insurance Claim number on Mathematica Policy Research, Inc.'s enrollment file, are excluded

TABLE VI.1 *(continued)*

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 12-month 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. See footnote 9 in this chapter for a detailed explanation of the weighting algorithm.

[a]A statistically significant and negative treatment-control difference and percent change value indicate that the treatment group is using fewer services than the 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.

the University of Maryland experienced more than 2 per year (Table VI.2). Similarly, average monthly Medicare expenditures for the control group over the first 25 months of operations varied substantially. The lowest-cost members were in Health Quality Partners, where the monthly average was $608, only slightly above the 2002 national average of $509 for all beneficiaries in fee-for-service (Cubanski et al. 2005). The highest-cost patients were served by Quality Oncology, where Medicare expenditures per patient averaged $4,280 a month.

The treatment group was slightly less likely than the control group overall to experience a hospitalization in the first year after randomization. The difference (2.2 percentage points, about 5 percent of the control group mean) is statistically significant ($p = 0.017$). The treatment group had a lower proportion hospitalized than the control group in 11 of the 15 programs, but Mercy was the only program to have a statistically significant difference favoring the treatment group. In Mercy, 43 three percent of the treatment group and 50 percent of the control group had a hospitalization, a 14-percent difference ($p = 0.075$). The proportion with a hospitalization was 23 percent *higher* for the treatment than the control group in Jewish Home and Hospital, and the difference was statistically significant ($p = 0.052$).

Overall, combining the 15 programs, the number of hospitalizations per patient was 3 percent lower during the year after intake and 4 percent lower during the first 25 months of operations. Neither modest difference was statistically significant ($p = 0.404$ and 0.145, respectively), but given the statistically significant effect on the proportion with an admission in their first year after enrollment, it probably reflects a true (though small) effect of the program. The magnitude of treatment-control differences in the year after the randomization month ranged across the 15 programs from the treatment group having 27 percent *more* hospitalizations to it having 33 percent *fewer* hospitalizations. Over the first 25 months of operations, the differences ranged from 14 percent more to 27 percent fewer hospitalizations (Table VI.2).

TABLE VI.2

AVERAGE MONTHLY MEDICARE EXPENDITURES AND NUMBER OF HOSPITALIZATIONS, CUMULATIVE THROUGH MONTH 25 OF PROGRAM OPERATIONS
(Regression Adjusted)

	Sample Sizes		Average Medicare Expenditures (per Patient-Month)[a]					Average Number of Hospitalizations per Year[b]				
	Treatment Group	Control Group	Treatment Group	Control Group	Treatment-Control Difference	Percent Change	*p*-Value	Treatment Group	Control Group	Treatment-Control Difference	Percent Change	*p*-Value
Avera	292	291	$1,401	$1,470	-$70	-5	0.641	1.51	1.45	0.06	4	0.728
Carle	1,178	1,161	$691	$699	-$7	-1	0.861	0.52	0.54	-0.02	-4	0.538
CenVaNet	616	611	$895	$847	$48	6	0.477	0.74	0.70	0.03	4	0.636
Charlestown	370	369	$1,216	$1,004	$212*	21	0.058	0.79	0.69	0.09	14	0.236
CorSolutions	1,159	869	$2,494	$2,700	-$206	-8	0.229	1.80	1.89	-0.09	-5	0.395
Georgetown	95	95	$2,082	$2,358	-$276	-12	0.534	1.64	1.86	-0.22	-12	0.487
Health Quality Partners	499	493	$609	$608	$1	0	0.989	0.37	0.41	-0.04	-10	0.505
Hospice of the Valley	370	358	$2,058	$2,061	-$2	0	0.990	1.25	1.46	-0.21	-14	0.127
Jewish Home and Hospital	352	347	$1,707	$1,815	-$108	-6	0.606	0.88	0.88	0.00	0	0.992
Medical Care Development	411	407	$1,531	$1,569	-$39	-2	0.820	1.39	1.38	0.01	1	0.959
Mercy	420	422	$1,039	$1,193	-$154	-13	0.105	0.73	1.01	-0.27***	-27	0.003
QMed	651	642	$606	$686	-$80	-12	0.349	0.37	0.39	-0.02	-4	0.740
Quality Oncology	65	63	$4,178	$4,280	-$101	-2	0.882	1.18	1.43	-0.25	-18	0.510
University of Maryland	66	59	$3,178	$3,178	$0	0	1.000	2.33	2.36	-0.03	-1	0.950
Washington University	968	964	$1,962	$1,893	$68	4	0.558	1.42	1.34	0.08	6	0.381
Overall	**7,512**	**7,151**	**$1,283**	**$1,314**	**-$31**	**-2**	**0.368**	**0.91**	**0.95**	**-0.04**	**-4**	**0.145**

Source: Medicare Enrollment Database, National Claims History File, and Standard Analytic File.

TABLE VI.2 *(continued)*

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 25-month follow-up period, or who had an invalid Health Insurance Claim number on Mathematica Policy Research, Inc.'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 25-month follow-up period each sample member is enrolled, 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. All other months are included, including those when the program had disenrolled the patient as a result of other reasons not described here. See footnote 9 in this chapter for a detailed explanation of the weighting algorithm.

[a]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.

[b]A statistically significant and negative treatment-control difference and percent change value indicate that the treatment group has fewer hospitalizations than the 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.

In five of the programs, the treatment group had at least 10 percent fewer hospitalizations than the control group during the first 25 months of operations. Mercy, the program with the largest difference (27 percent), is the only program for which the difference was statistically significant. This difference is somewhat larger than the 16-percent difference (p = 0.251) observed in the year after intake and is consistent with the statistically significant difference in the proportion with a hospitalization observed in year 1 (noted above). Quality Oncology, Hospice of the Valley, Georgetown University, and Health Quality Partners are the four other programs with sizable treatment-control differences in hospitalizations over the first 25 months, but these differences were not statistically significant. The four programs' treatment groups experienced 18, 14, 12, and 10 percent fewer hospitalizations, respectively, than their control groups. Two of these programs (Georgetown University and Quality Oncology) had very small sample sizes, making it difficult to attribute the differences to the programs, rather than to chance.

The treatment groups in three programs had higher hospitalizations rates than the control groups. Although none of these differences were statistically significant, it is possible that care coordination increases utilization (Congressional Budget Office 2004; Schore et al. 1999). This might occur if care coordinators uncover unmet needs. Over the 25-month period, Charlestown's treatment group had 14 percent more hospitalizations per year than the control group (0.79 versus 0.69, p = 0.236). Given the already rich medical services in that environment, the difference seems likely to be due to chance, rather than to the effects of the program, as do the smaller differences in Washington University and Avera.

Although some programs said that they expected the number of physician visits and tests to increase in the first year, as patients were encouraged to obtain more preventive care, the treatment groups had about the same rate of use as the control groups. There were very few

statistically significant treatment-control differences in the patients' number of physician visits or the proportion who had tests or imaging (Table VI.3) or in hospital outpatient department services or the numbers of those services they received (Table VI.4). Furthermore, even some of the statistically significant differences were very small in magnitude. For example, the treatment group in Carle received an average of one more test or imaging procedure (Table VI.3). The observed difference in the number of hospital outpatient services (4) was somewhat larger in Georgetown ($p = 0.010$). The one instance of significantly *lower* use in the treatment group was for use of hospital outpatient department services in the small University of Maryland program (68 percent versus 88 percent, $p = 0.095$) (Table VI.4).

The treatment groups in two programs had lower use of ER services than the control groups. The proportion of the treatment groups in CorSolutions and Hospice of the Valley using ER services were 8.4 and 10 percentage points lower than their respective control groups ($p = 0.038$ and 0.045, respectively) (Table VI.5). These differences did not, however, translate to a difference in the average number of ER visits.

The treatment and control groups in 14 of the 15 programs had a statistically comparable mortality rate over the year after the month of random assignment. The only exception was in Health Quality Partners, where the mortality rate of the treatment group was 2 percentage points lower than that of the control group ($p = 0.093$) (Table VI.6). The mortality rate of the programs' enrollees varied substantially. Enrollees in Carle, Health Quality Partners, and QMed had 1-year mortality rates well below the national average for all Medicare beneficiaries of 5 percent, which suggests that the patients enrolled are healthier than average despite having some chronic illnesses. In contrast, five programs had high mortality rates, ranging from 15 to 38 percent, which indicates that they enrolled very sick beneficiaries.

TABLE VI.3

PHYSICIAN SERVICES AND PROCEDURES FOR 1ST-YEAR ENROLLEES
DURING THE YEAR AFTER RANDOM ASSIGNMENT

	Sample Sizes		Average Annualized Number of Claims[a,b]					
			Physician Services and Procedures (Not in Hospital or ER)			Tests and Imaging		
	Treatment Group	Control Group	Treatment Group	Treatment-Control Difference	*p*-Value	Treatment Group	Treatment-Control Difference	*p*-Value
Avera	157	158	14.3	0.5	0.697	20.1	0.7	0.689
Carle	1,024	1,018	10.9	0.6	0.125	12.9	1.0**	0.048
CenVaNet	512	509	11.0	0.0	0.945	13.6	0.5	0.484
Charlestown	195	189	20.2	1.4	0.229	16.6	2.0	0.144
CorSolutions	354	265	17.1	-0.3	0.795	22.3	-0.9	0.607
Georgetown	51	53	16.2	0.3	0.873	19.3	-1.3	0.713
Health Quality Partners	219	219	13.1	0.2	0.785	8.6	0.6	0.508
Hospice of the Valley	222	210	15.9	-1.4	0.308	17.6	-1.5	0.405
Jewish Home and Hospital	253	252	19.1	1.3	0.364	11.8	1.8	0.151
Medical Care Development	192	192	13.9	-0.8	0.458	16.6	-3.0	0.153
Mercy	304	308	14.7	0.2	0.779	8.8	-0.9	0.220
QMed	633	625	13.9	0.7	0.266	8.8	0.5	0.331
Quality Oncology	29	31	36.0	2.2	0.690	37.5	1.5	0.819
University of Maryland	29	26	11.1	-0.8	0.761	22.8	3.6	0.550
Washington University	698	690	14.6	-0.3	0.637	19.8	0.4	0.741
Overall	**4,872**	**4,745**	**14.0**	**0.3**	**0.200**	**14.5**	**0.5**	**0.119**

Source: Medicare Enrollment Database and National Claims History 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 1-year follow-up period, had an invalid Health Insurance Claim number on Mathematica Policy Research, Inc.'s enrollment file, or were identified as a member of the same household as a research sample member are excluded from this table.

The outcomes are weighted according to the proportion of the 12-month 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. See footnote 9 in this chapter for a detailed explanation of the weighting algorithm.

[a]This count is limited to 1 per day for each provider for each patient.

[b]The direction of the treatment-control difference does not by itself signify whether the program is "effective." A positive difference does not necessarily mean the program is ineffective or having adverse effects, because the program may encourage patients to see their physicians more regularly for preventative care or to obtain more recommended laboratory tests for their target conditions than they would have had in the absence of the demonstration.

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

ER = emergency room.

TABLE VI.4

USE OF HOSPITAL OUTPATIENT DEPARTMENT SERVICES FOR 1ST-YEAR ENROLLEES DURING THE YEAR AFTER RANDOM ASSIGNMENT

	Hospital Outpatient Department Services[a]					
	Any Use (Percent)			Average Annualized Number of Claims		
	Treatment Group	Treatment-Control Difference	*p*-Value	Treatment Group	Treatment-Control Difference	*p*-Value
Avera	90.6	0.8	0.822	10.8	-0.4	0.797
Carle	77.0	2.8	0.141	5.0	0.5*	0.055
CenVaNet	82.7	-0.3	0.885	5.5	0.5	0.229
Charlestown	87.1	5.9	0.115	5.6	0.7	0.265
CorSolutions	83.3	-0.8	0.802	5.4	0.6	0.258
Georgetown	95.6	5.7	0.272	10.1	4.0*	0.010
Health Quality Partners	88.8	3.3	0.307	5.2	0.7	0.200
Hospice of the Valley	79.0	0.3	0.942	5.0	0.3	0.641
Jewish Home and Hospital	93.8	-1.3	0.511	9.4	-0.3	0.733
Medical Care Development	96.9	-2.0	0.167	17.5	1.0	0.514
Mercy	100.0	1.0*	0.086	18.7	1.0	0.358
QMed	75.6	2.5	0.314	4.1	0.4	0.154
Quality Oncology	89.2	-5.0	0.497	6.7	0.5	0.815
University of Maryland	67.8	-19.5*	0.095	6.4	-0.8	0.758
Washington University	96.6	0.1	0.922	11.8	0.5	0.353
Overall	**85.5**	**1.1**	**0.155**	**7.8**	**0.5****	**0.013**

Source: Medicare Enrollment Database, National Claims History File, 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 1-year follow-up period, had an invalid Health Insurance Claim number on Mathematica Policy Research, Inc.'s enrollment file, or were identified as a member of the same household as a research sample member are excluded from this table.

The outcomes are weighted according to the proportion of the 12-month 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. See footnote 9 in this chapter for a detailed explanation of the weighting algorithm.

[a]The direction of the treatment-control difference does not by itself signify whether the program is "effective." A positive difference does not necessarily mean the program is ineffective or having adverse effects, because the program may encourage patients to obtain outpatient services to identify a health problem and prevent a more expensive exacerbation, which might in turn lead to lower costs than they would have had in the absence of the demonstration.

*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 VI.5

OUTPATIENT ER USE FOR 1ST-YEAR ENROLLEES DURING THE YEAR AFTER RANDOM ASSIGNMENT

	Sample Sizes		Any Use (Percent)			Average Number of ER Visits		
	Treatment Group	Control Group	Treatment Group	Treatment-Control Difference[a]	*p*-Value	Treatment Group	Treatment-Control Difference[a]	*p*-Value
Avera	157	158	37.0	-7.9	0.164	0.83	-0.04	0.843
Carle	1,024	1,018	29.9	3.8*	0.056	0.52	0.07	0.314
CenVaNet	512	509	25.1	-3.9	0.161	0.37	-0.12*	0.054
Charlestown	195	189	19.5	4.4	0.264	0.25	0.10	0.276
CorSolutions	354	265	35.5	-8.4**	0.038	0.70	-0.08	0.564
Georgetown	51	53	40.0	-13.5	0.181	1.30	0.28	0.526
Health Quality Partners	219	219	24.3	0.3	0.948	0.30	-0.02	0.803
Hospice of the Valley	222	210	36.8	-10.0**	0.045	0.61	-0.14	0.257
Jewish Home and Hospital	253	252	29.7	-0.7	0.868	0.53	0.01	0.921
Medical Care Development	192	192	48.7	-7.7	0.139	1.26	-0.28	0.276
Mercy	304	308	40.9	-1.9	0.632	0.83	0.06	0.654
QMed	633	625	19.2	-2.5	0.276	0.32	-0.02	0.781
Quality Oncology	29	31	37.1	8.8	0.499	0.44	-0.03	0.924
University of Maryland	29	26	46.4	-1.4	0.918	0.76	-0.52	0.263
Washington University	698	690	39.4	-1.8	0.510	0.73	-0.15	0.108
Overall	**4,872**	**4,745**	**31.2**	**-1.7***	**0.081**	**0.57**	**-0.04**	**0.244**

Source: Medicare Enrollment Database, National Claims History File, 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 1-year follow-up period, had an invalid Health Insurance Claim number on Mathematica Policy Research, Inc.'s enrollment file, or were identified as a member of the same household as a research sample member are excluded from this table.

The outcomes are weighted according to the proportion of the 12-month 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. See footnote 9 in this chapter for a detailed explanation of the weighting algorithm.

[a]The direction of the treatment-control difference does not by itself signify whether the program is "effective." A positive difference does not necessarily mean the program is ineffective or having adverse effects, because the program may encourage patients to obtain ER services to identify a health problem and prevent a more expensive exacerbation, which might in turn lead to lower costs than they would have had in the absence of the demonstration.

*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.

ER = emergency room.

TABLE VI.6

MORTALITY RATE FOR 1ST-YEAR ENROLLEES DURING THE YEAR AFTER RANDOM ASSIGNMENT

	Sample Sizes				
	Treatment Group	Control Group	Treatment Group (Percent)	Treatment-Control Difference[a] (Percent)	*p*-Value
Avera	157	158	17.8	-0.5	0.905
Carle	1,024	1,018	3.9	-0.6	0.491
CenVaNet	512	509	8.4	1.7	0.299
Charlestown	195	189	11.3	-0.4	0.912
CorSolutions	354	265	16.1	-0.1	0.967
Georgetown	51	53	13.7	-1.4	0.843
Health Quality Partners	219	219	0.9	-2.3*	0.093
Hospice of the Valley	222	210	30.6	2.5	0.563
Jewish Home and Hospital	253	252	8.7	3.5	0.118
Medical Care Development	192	192	18.2	1.0	0.789
Mercy	304	308	9.5	-0.9	0.726
QMed	633	625	2.4	0.3	0.728
Quality Oncology	29	31	37.9	-0.8	0.951
University of Maryland	29	26	13.8	-9.3	0.377
Washington University	698	690	11.7	-0.7	0.683
Overall	**4,872**	**4,745**	**9.7**	**0.3**	**0.628**

Source: Medicare Enrollment Database.

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 1-year follow-up period, had an invalid Health Insurance Claim number on Mathematica Policy Research, Inc.'s enrollment file, or were identified as a member of the same household as a research sample member are excluded from this table.

[a]A statistically significant and negative treatment-control difference indicates that a lower proportion of the treatment group than the control group died during the year after random assignment. This signifies that the program is working as intended.

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

With two exceptions, the treatment groups' Medicare expenditures, excluding care coordination fees, were statistically comparable to those of the control groups over both the first 25 months of program operations (Table VI.2) and the first year after the randomization month (Table VI.7). The only program that had a statistically significant treatment-control difference in expenditures (excluding care coordination fees) was Charlestown, where Medicare expenditures for the treatment group were actually *higher.* After regression adjustment, the average monthly Medicare expenditures for Charlestown's treatment group were $393, or 40 percent, more than the expenditures for the control group over the year after the randomization month ($p = 0.044$) and $212, or 21 percent, higher per month over the first 25 months of program operations ($p = 0.058$). The other exception to the pattern of no differences was Mercy, whose treatment group used fewer inpatient hospital services. Mercy's treatment group had 13 percent ($154) lower monthly Medicare expenditures over the first 25 calendar months, and the *p*-value (0.105) was very close to the 10-percent significance level. This result is promising, but the difference is not enough to offset Mercy's care coordination fees of $245 per month over this time period. While not statistically significant, the treatment groups served by Georgetown University and QMed each had expenditures 12 percent lower than those of their control groups. The treatment groups in three other programs had expenditures 8, 6, and 5 percent lower over the first 25 months. (These were not statistically significant.)

Across the 15 programs, monthly Medicare expenditures were 2 percent lower ($p = 0.368$) during the year after randomization, and 1 percent lower ($p = 0.724$) over the 25-month period—not enough to offset care coordination fees during either period. This does not rule out the possibility that larger savings will accrue with a longer followup.

TABLE VI.7

MEDICARE EXPENDITURES FOR 1ST-YEAR ENROLLEES
DURING THE YEAR AFTER RANDOM ASSIGNMENT
(Regression Adjusted)

	Sample Sizes		Average Medicare Expenditures per Month in Fee-for-Service (Part A and Part B Combined)[a]				
	Treatment Group	Control Group	Treatment Group	Control Group	Treatment-Control Difference	Percent Change	*p*-Value
Avera	157	158	$1,299	$1,522	-$223	-15	0.264
Carle	1,024	1,018	$646	$692	-$45	-7	0.384
CenVaNet	512	509	$905	$800	$105	13	0.218
Charlestown	195	189	$1,387	$993	$393**	40	0.044
CorSolutions	354	265	$2,465	$2,912	-$447	-15	0.219
Georgetown	51	53	$2,184	$2,141	$44	2	0.939
Health Quality Partners	219	219	$554	$504	$50	10	0.611
Hospice of the Valley	222	210	$2,181	$2,026	$154	8	0.555
Jewish Home and Hospital	253	252	$1,756	$1,686	$70	4	0.811
Medical Care Development	192	192	$1,620	$1,789	-$169	-9	0.597
Mercy	304	308	$1,090	$1,099	-$8	-1	0.946
QMed	633	625	$588	$640	-$52	-8	0.621
Quality Oncology	29	31	$4,333	$4,709	-$376	-8	0.714
University of Maryland	29	26	$3,029	$2,767	$262	9	0.829
Washington University	698	690	$2,019	$1,940	$78	4	0.607
Overall	**4,872**	**4,745**	**$1,246**	**$1,262**	**-$16**	**-1**	**0.724**

Source: Medicare Enrollment Database, National Claims History File, 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 1-year follow-up period, had an invalid Health Insurance Claim number on Mathematica Policy Research, Inc.'s enrollment file, or were identified as a member of the same household as a research sample member are excluded from this table.

The outcomes are weighted according to the proportion of the 12-month 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. See footnote 9 in this chapter for a detailed explanation of the weighting algorithm.

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

B. COST NEUTRALITY

Interpreting the results on program effects on Medicare expenditures is complicated, because it may not be possible to reject either of two hypotheses that have diametrically opposed implications for the program. The first hypothesis is that the program had no effect on Part A and Part B expenditures (without considering the program fees). Because the sample sizes are relatively small and the variance of expenditures is large, the estimate of program effects on Part A and B expenditures has a relatively large confidence interval around it. In other words, the true effect may range around the particular estimate generated by this sample. As a result, the hypothesis that the treatment-control difference in total Medicare expenditures on traditional Part A and Part B services is different from zero—that the program reduced or increased Part A and B expenditures—is likely to be rejected unless the observed difference is reasonably large (discussed above). This provides an appropriately conservative test of program impacts.

To be cost neutral, programs must not only reduce Medicare expenditures, but reduce them by enough to offset program fees. Statistically, this second hypothesis is assessed by testing whether the estimated treatment-control difference is significantly different from the average care coordination fee paid per month to the program (or equivalently, whether the combined total of Part A and B expenditures and program fees are different for the treatment and control groups). Because the variances are large, it may not be possible to reject this hypothesis either—that is, the confidence interval around the estimated treatment-control difference in Part A and B expenditures encompasses both zero and the average fee paid per month. It is therefore possible to fail to reject the first hypothesis, acknowledging that the program did not reduce Medicare expenditures for traditional services, and also fail to reject the second, that is, to conclude that the program may have been cost neutral, despite not clearly reducing Part A and B expenditures.

The problem is particularly acute for programs with few enrollees, because the variances of the estimates for these programs are markedly larger.

This ambiguity, while confusing, properly represents the uncertainty about the results. The most conservative approach would be simply to conclude that, unless the treatment-control difference in traditional Medicare expenditures was significantly different from zero, there would be no need even to test 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.

One partial solution, implemented below, is to test 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 more traditional 5- or 10-percent level. This approach increases the statistical power of the analysis (that is, it increases the likelihood of concluding the program increased or decreased costs when it really had no effect).[10] However, this approach helps only marginally, because the differences to be detected are small. For example, Quality Oncology's monthly program fee was only 2 percent of the large control group mean of $4,178 per month for traditional Medicare services. If the program had actually reduced Part A and B expenditures by 2 percent, detecting with 80-percent power a difference this small in total payments including the program fee would require a huge sample size (nearly 70,000 patients in each group), far greater than the very small sample that this or any other program has.

[10] By accepting a higher-than-usual probability (20 percent) of falsely concluding there is a positive or negative effect when there really is no effect (the program is cost neutral), it is more likely than it would have been to properly reject the hypothesis of cost neutrality when it is false. Thus, this is a more conservative approach for assessing whether the program is cost neutral. Whereas the conservative approach to examining savings on Part A and B costs is to have a low probability of concluding there are savings when in fact there are not, for assessing cost neutrality, the conservative approach is to have a low probability of rejecting the null hypothesis of cost neutrality when in fact there are net cost *increases.*

Only four programs have received fees that exceed 20 percent of the control group mean, and these are the only ones for which there is at least 70-percent power to detect a difference in Medicare expenditures large enough to cover the cost of the fees (Table VI.8). For seven of the other programs, the power is less than 50 percent. That is, failure to reject the hypothesis of cost neutrality when it is false is more likely than not. Clearly, it would be inappropriate to conclude on such weak evidence that a program is cost neutral. Conversely, holding the programs to the rigorous standard of having to demonstrate statistically significant savings means running a sizable risk of failing to detect true impacts large enough to cover the cost of the intervention.

Given this uncertainty, to determine whether program effects appear to be large enough to cover the cost of the fees, the evaluation relies instead on a combination of findings about Part A and B expenditures, total expenditures including program fees, and effects on hospitalizations. While 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 the most likely place where programs could generate Medicare savings. The results for all three outcomes are regression adjusted to account for any chance differences in preenrollment characteristics of the two research groups.

Two examples illustrate how the evaluation determines whether a program is likely to be cost neutral. During the year after random assignment, among all 15 programs, the treatment group's expenditures (excluding care coordination fees) are $16 per month lower than the control group's, but the 90-percent confidence interval indicates that, if this were the true effect of the program, the treatment-control difference in any given sample of this size could be expected to fall somewhere between a reduction of $93 and an increase of $60 per month. Because the confidence interval includes $0, the hypothesis of no effect on expenditures cannot be rejected. In other words, the programs may not have reduced Medicare expenditures relative to what they

TABLE VI.8

PRECISION OF ESTIMATES OF PROGRAM EFFECTS ON MEDICARE EXPENDITURES
(2-Year Sample)

	Sample Sizes		Power to Detect 20 Percent Effect on Cost	Average Fee Received per Month in Evaluation Sample	Average Control Group Cost	Percent Savings Needed to Cover Fee	Power to Detect Impact Needed to Cover Fee
	Treatment Group	Control Group					
Avera	292	291	0.63	$271	$1,470	18	0.58
Carle	1,178	1,161	0.97	$152	$699	22	0.99
CenVaNet	616	611	0.85	$72	$1,004	7	0.33
Charlestown	370	369	0.70	$233	$847	28	0.89
CorSolutions	1,159	869	0.95	$315	$2,700	12	0.67
Georgetown	95	95	0.36	$296	$2,358	13	0.24
Health Quality Partners	499	493	0.79	$105	$2,061	5	0.23
Hospice of the Valley	370	358	0.70	$190	$608	31	0.94
Jewish Home and Hospital	352	347	0.68	$260	$1,815	14	0.49
Medical Care Development	411	407	0.73	$180	$1,569	11	0.43
Mercy	420	422	0.74	$250	$1,193	21	0.77
QMed	651	642	0.87	$88	$686	13	0.60
Quality Oncology	65	63	0.30	$81	$4,280	2	0.11
University of Maryland	66	59	0.30	$321	$3,178	10	0.18
Washington University	968	964	0.95	$166	$1,893	9	0.50
Overall	**7,512**	**7,151**	**1.00**	**$196**	**$1,314**	**15**	**1.00**

Source: Medicare Enrollment Database, National Claims History File, and Standard Analytic File.

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

would have been, absent the demonstration. In this case, even the most optimistic estimate of a reduction of Part A and B expenditures of $93 (the lower end of the confidence interval) is not enough to offset the average monthly care coordination fee of $196. As a result, the evaluation definitively demonstrates that the overall demonstration was not cost neutral during this period.

QMed, in contrast, is an example of an ambiguous case where the program might be cost neutral. The estimate of the monthly treatment-control difference in Part A and B expenditures is -$52, and the 90-percent confidence interval ranges from -$227 to $122. Because this interval encompasses $0, the evaluation cannot reject the hypothesis of no effect on Part A and B expenditures. Despite being unable to conclude that the program reduces Part A and B expenditures, the evaluation also cannot reject the possibility that the program is cost neutral, because the change in Part A and B expenditures (-$96) needed to cover the program's average monthly fee also falls within this interval.

1. Tests of Cost Neutrality for the Year Following Enrollment in the Study

Over the first year after the randomization month, the demonstration as a whole was *not* cost neutral. Total expenditures, including program fees, were $157 (about 12 percent higher) for the treatment than the control groups (p = 0.001). This result reflects the finding that, overall, the treatment group's Part A and B expenditures were virtually identical to the control group's.

Up to five of the programs may be cost neutral over the first year after randomization, but because of the large variation in the measures of cost and, in some cases, small samples, this evidence is weak. The treatment-control differences in total Medicare expenditures, including program fees, for 7 of the 15 programs are significantly greater than 0 at a 20-percent significance level (see Table VI.9, top panel)—that is, these ***seven programs are clearly not cost neutral during this period***. Each program significantly *increases* average expenditures by $109 to $631 per month. These estimates are fairly imprecise, given the sample size and sizable

TABLE VI.9

COST NEUTRALITY DURING THE 1ST YEAR AFTER RANDOMIZATION, AMONG 1ST-YEAR ENROLLEES
(Regression Adjusted)

Program	Average Monthly Care Coordination Fee[a]	Treatment-Control Expenditure Difference Without Care Coordination Fee[b]	Treatment-Control Difference in Medicare Expenditures per Month with Care Coordination Fee				
				80-Percent Confidence Interval			
			Difference[c]	Lower Bound	Upper Bound	Percent of Control Mean	*p*-Value
Not Cost Neutral							
Charlestown	$244	$393	$631	$381	$880	63.5	0.001
Hospice of the Valley	$224	$154	$352	$18	$687	17.4	0.178
Washington University	$173	$78	$246	$51	$441	12.7	0.107
Mercy	$257	-$8	$242	$82	$401	22.0	0.052
CenVaNet	$80	$105	$179	$70	$288	22.4	0.035
Health Quality Partners	$108	$50	$154	$28	$280	30.5	0.118
Carle	$159	-$45	$109	$42	$176	15.7	0.037
Probably Not Cost Neutral							
Jewish Home and Hospital	$317	$70	$336	-$40	$713	19.9	0.253
Georgetown	$320	$44	$330	-$391	$1,052	15.4	0.559
University of Maryland	$350	$262	$598	-$948	$2,143	21.6	0.623
Possibly Cost Neutral							
QMed	$96	-$52	$35	-$101	$171	5.5	0.740
Avera	$316	-$223	$53	-$200	$306	3.5	0.790
Medical Care Development	$297	-$169	$44	-$365	$453	2.5	0.890
CorSolutions	$444	-$447	-$104	-$569	$360	-3.6	0.774
Quality Oncology	$140	-$376	-$269	-$1,578	$1,040	-5.7	0.793
Overall	**$196**	**-$16**	**$157**	**$97**	**$217**	**12.4**	**0.001**

TABLE VI.9 *(continued)*

Source: Medicare Enrollment Database, National Claims History File, and Standard Analytic File.

[a]Note that the amount presented here for 12 programs is the approved 1st-year rate for months excluding the 1st month. Rates were changed slightly for Avera, Georgetown University, Mercy, and Medical Care Development on April 1, 2003. In 3 programs (CorSolutions, Health Quality Partners, and Jewish Home and Hospital), the amount presented is a blended rate that averages the actual nonzero payments, because the programs charged different fees for patients depending on the severity of their condition or their time enrolled. CorSolutions was approved to charge $437 for the first 9 months, $287 for subsequent months until patients are designated as being in maintenance, and $187 in maintenance. They also charge $366 per month for patients who receive prescription drug coverage. Health Quality Partners charges $130, $110, or $50, and Jewish Home and Hospital $379, $259, or $74, based on each patient's acuity level. The amount paid to each of the 15 programs as recorded in the Medicare claims data and the programs' approved per member per month differ as a result of payment adjustments for patients who disenrolled from the programs but remain in the research sample, as well as any billing errors.

[b]A statistically significant and negative treatment-control difference would indicate that expenditures were lower for the treatment than control group. This would signify that the program is working as intended.

[c]A statistically significant and positive treatment-control difference indicates that expenditures including the care coordination fee were higher for the treatment than control group. This signifies that the program is increasing total costs and is therefore not cost neutral. When the difference is not statistically significant, it is possible that the program is cost neutral. In those cases, information about the program's effects on hospitalizations was also considered when classifying the program as "Probably Not Cost Neutral" or "Possibly Cost Neutral."

variations in costs. The 80-percent confidence intervals illustrate the range within which the true effect is expected to fall. For example, while the net monthly costs in Mercy are $242 higher for the treatment group than for the control group, it can be stated with 80-percent confidence that the true effect of the program on total Medicare expenditures lies somewhere between $82 and $401.

Three other programs are highly unlikely to be cost neutral, even though the treatment-control differences in total expenditures were not significantly different from zero (*p*-values were greater than 0.200), for two reasons. First, a program would probably not be cost neutral unless the treatment group had fewer hospitalizations than the control group. The treatment groups in Jewish Home and Hospital, Georgetown University, and the University of Maryland all experienced *more* hospitalizations than the control groups during this period (Table VI.1), which makes it unlikely that the programs were reducing Medicare expenditures enough to offset their program fees. Second, small sample sizes and large variations in Medicare expenditures limit the power to detect the size of the differences the programs are generating. This makes it especially difficult to reject the hypothesis that the effect is statistically different from 0 in Georgetown and the University of Maryland, whose sample sizes of 104 and 55, respectively, were very small.

Among the remaining ***five programs that may have been cost neutral,*** the findings from different analyses suggest that only three are really likely to have been cost neutral during the year after the randomization month. The treatment group in Medical Care Development experienced only slightly fewer hospitalizations on average (2 percent), and this difference was not statistically significant; even if it were, a program effect would be unlikely to generate enough savings to offset the program fee. Avera's treatment group had 5-percent more hospitalizations than the control group. Thus, it is unlikely that these two programs generated

enough savings in Part A and B expenditures to offset their fees. However, the evidence suggests that the three other programs (QMed, CorSolutions, and Quality Oncology) may be cost neutral, as they are the only ones for which the treatment group had lower Medicare expenditures than the control group (8, 15, and 8 percent, respectively). (The treatment groups also had 4-, 9-, and 33-percent fewer hospitalizations, respectively, than the control groups.) However, the hypothesis that the savings in Part A and B expenditures are zero also cannot be rejected, so these results must be interpreted cautiously.

2. Tests of Cost Neutrality Over the 25 Months Since Program Startup

The results for the first 25 months after startup for all programs combined are similar to those for enrollees' first 12 months after enrollment, but their pattern across programs is somewhat different from that of the 12-month followup (Table VI.10). Overall, total Medicare expenditures, including program fees, are $144, or 11 percent, higher per month for the treatment than for the control groups ($p < 0.001$), about the same as the difference estimated above for the year after enrollment.

During the 25 months since startup, ***six programs are definitely not cost neutral, four might be but are probably not, and five might be cost neutral.*** Five of the seven programs that were not cost neutral in the 1-year results are also definitely not cost neutral during the 25 months since startup (see top panel of Table VI.10 and Figure VI.1). In addition, Avera definitively moves into this category.

Four programs are probably not cost neutral (middle panel of Table VI.10). CorSolutions had the largest treatment-control difference in Medicare Part A and B expenditures (without program fees) of -$206 (not significant). While this is larger than the treatment-control difference in all but 1 of the 14 other programs, it is too small to offset its large monthly fee (which averaged $444). In addition, CorSolutions's treatment group had only about 5 percent

TABLE VI.10

COST NEUTRALITY DURING THE 1ST 25 CALENDAR MONTHS OF PROGRAM OPERATIONS
(Regression Adjusted)

Program	Average Monthly Care Coordination Fee[a]	Treatment-Control Expenditure Difference Without Care Coordination Fee[b]	Treatment-Control Difference in Medicare Expenditures per Month with Care Coordination Fee				
			Difference[c]	80-Percent Confidence Interval		Percent of Control Mean	*p*-Value
				Lower Bound	Upper Bound		
Not Cost Neutral							
Charlestown	$244	$212	$445	$303	$588	44.4	0.000
Washington University	$173	$68	$234	$84	$383	12.3	0.045
Avera	$316	-$70	$201	$11	$391	13.7	0.175
Carle	$159	-$7	$145	$90	$199	20.7	0.001
CenVaNet	$80	$48	$120	$34	$207	14.2	0.076
Health Quality Partners	$108	$1	$106	$5	$206	17.4	0.179
Probably Not Cost Neutral							
University of Maryland	$350	$0	$321	-$692	$1,334	10.1	0.685
Jewish Home and Hospital	$317	-$108	$152	-$116	$420	8.4	0.468
Medical Care Development	$297	-$39	$141	-$76	$359	9.0	0.406
CorSolutions	$444	-$206	$109	-$110	$328	4.0	0.525
Possibly Cost Neutral							
QMed	$96	-$80	$8	-$102	$118	1.2	0.924
Mercy	$257	-$154	$96	-$26	$217	8.0	0.312
Hospice of the Valley	$224	-$2	$188	-$35	$412	9.1	0.280
Georgetown	$320	-$276	$20	-$546	$587	0.9	0.963
Quality Oncology	$140	-$101	-$20	-$894	$854	-0.5	0.976
Overall	**$196**	**-$31**	**$144**	**$99**	**$188**	**11.3**	**0.000**

TABLE VI.10 *(continued)*

Source: Medicare Enrollment Database, National Claims History File, and Standard Analytic File.

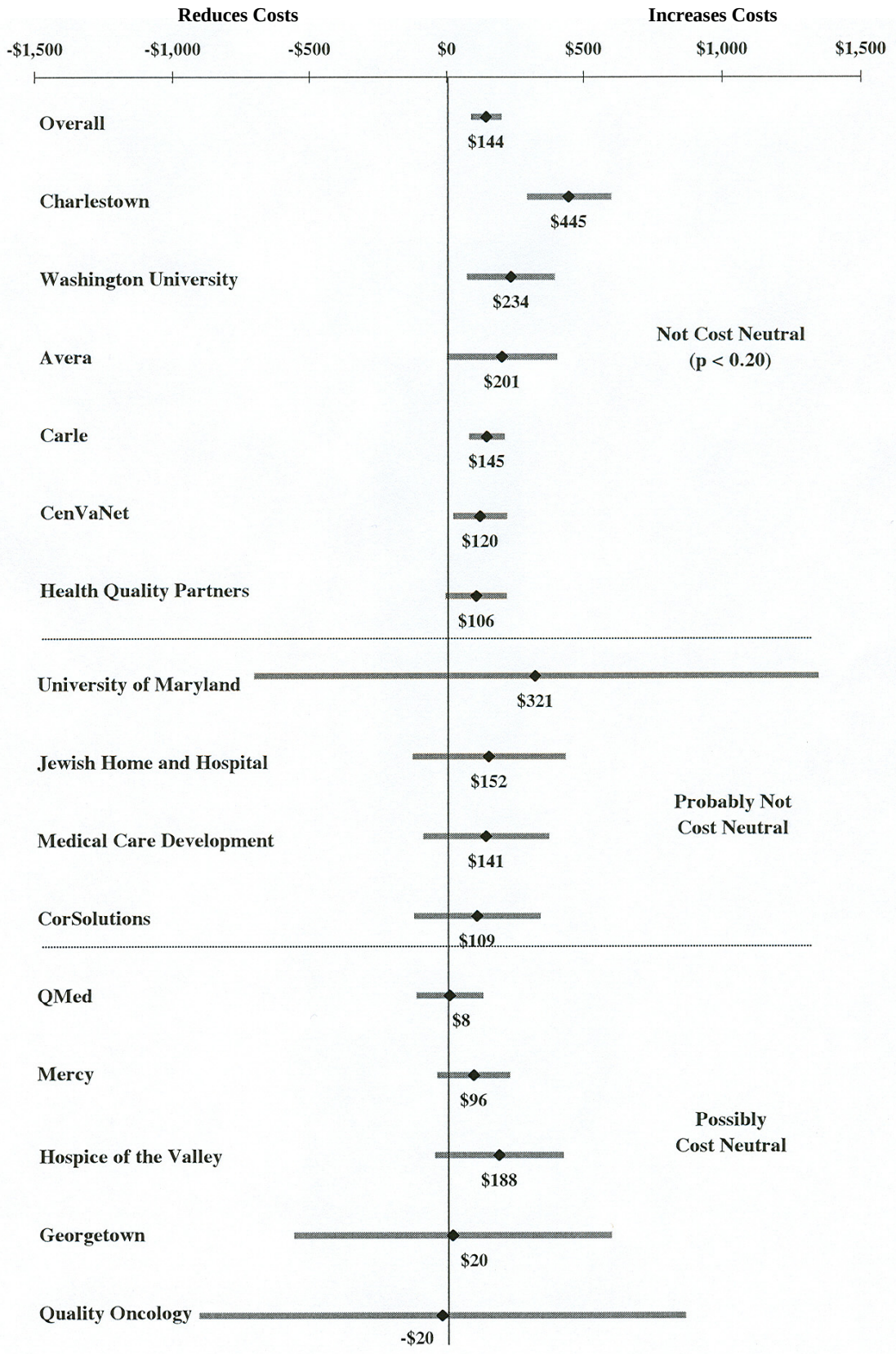
[a]Note that the amount presented here for 12 programs is the approved 1st-year rate for months excluding the 1st month. Rates were changed slightly for Avera, Georgetown University, Mercy, and Medical Care Development on April 1, 2003. In 3 programs (CorSolutions, Health Quality Partners, and Jewish Home and Hospital), the amount presented is a blended rate that averages the actual nonzero payments, because the programs charged different fees for patients depending on the severity of their condition or their time enrolled. CorSolutions was approved to charge $437 for the first 9 months, $287 for subsequent months until patients are designated as being in maintenance, and $187 in maintenance. They also charge $366 per month for patients who receive prescription drug coverage. Health Quality Partners charges $130, $110, or $50, and Jewish Home and Hospital $379, $259, or $74, based on each patient's acuity level. The amount paid to each of the 15 programs as recorded in the Medicare claims data and the programs' approved per member per month differ as a result of payment adjustments for patients who disenrolled from the programs but remain in the research sample, as well as any billing errors.

[b]A statistically significant and negative treatment-control difference would indicate that expenditures were lower for the treatment than control group. This would signify that the program is working as intended.

[c]A statistically significant and positive treatment-control difference indicates that expenditures including the care coordination fee were higher for the treatment than control group. This signifies that the program is increasing total costs and is therefore not cost neutral. When the difference is not statistically significant, and is either positive or negative, it is possible that the program is cost neutral. In those cases, information about the program's effects on hospitalizations was also considered when classifying the program as "Probably Not Cost Neutral" or "Possibly Cost Neutral."

FIGURE VI.1

COST NEUTRALITY DURING THE 1ST 25 CALENDAR MONTHS OF PROGRAM OPERATIONS



fewer hospital admissions than the control group during the 25 months since startup. Such a small difference in hospitalizations (even if a real program effect) seems unlikely to generate the 16-percent reduction in Part A and B expenditures needed to cover the program's fees. Although the treatment group costs in each of the three other programs in the middle panel of Table VI.10 are lower than or equal to those of the control group, the differences are not significant and are too small in each case to outweigh the monthly care coordination fees. Furthermore, the programs exhibit little treatment-control differences in hospitalization rates. Thus the observed differences in Medicare Part A and B expenditures appear to be mostly the result of chance fluctuations rather than true program effects, and even if true savings exist, they are likely to be too small to offset the monthly fee.

Five programs—Mercy, Hospice of the Valley, Georgetown, Quality Oncology, and QMed—are more likely to be cost neutral (bottom panel of Table VI.10). Mercy is the clearest case, with a large and highly significant difference in hospital admissions (27 percent, $p = 0.003$). While the estimated treatment-control difference of $154 per member per month ($p = 0.105$) is not enough to offset fully the average monthly fee that Mercy receives ($257 per month for enrolled patients), the statistical variation around this estimate is large enough that neither the hypothesis that the savings in Part A and B expenditures are zero nor the hypothesis that the savings are greater than the average care coordination fees received per month the patients are followed up can be rejected. QMed offers an ambiguous case, because the difference in Part A and B expenditures (-$80, or 12 percent of the control group mean) virtually offsets the low monthly fee of $96. While the treatment group had only 4-percent fewer hospital admissions than the control group ($p = 0.740$), a small change in hospital use could be enough to offset QMed's relatively low program fee. The two other programs (Hospice of the Valley and Georgetown) are somewhat similar to Mercy in that both have treatment groups with

hospitalization rates 10 percent or more below those of their respective control groups (though neither difference is statistically significant), and neither hypothesis can be rejected. However, the evidence to support cost neutrality is considerably weaker for both of them than for Mercy. Georgetown has a small sample (about 100 patients each in the treatment and control groups), and Hospice of the Valley has virtually no likelihood that it has generated savings in Part A and B expenditures (p = 0.990). Nonetheless, given the test results, these are included along with Quality Oncology, QMed, and Mercy in the category of potentially cost neutral programs at the midpoint of the 4-year demonstration.

The differences between the expenditure results for the 1-year postenrollment period and the first 25 months of operation warrant some further discussion, because these data are useful for assessing the robustness of the results (Table VI.11). Altogether, seven of the programs are classified as possibly or definitely not cost neutral in either analysis, two (Quality Oncology and QMed) are classified as probably cost neutral in both analyses, and six are classified differently in the two analyses. The three programs that appear possibly to be cost neutral in the 25-month analysis but were probably or definitely not cost neutral in the 1-year follow-up estimates (Hospice of the Valley, Georgetown, and Mercy) appear to have had larger effects as their experience has grown and as patients were exposed to them for a longer period. This pattern, which was expected, will be subjected to closer scrutiny as more data become available and allow for separate assessment of how program effects vary with beneficiaries' length of time enrolled and with program experience.

The only odd difference in results between the two analyses is for CorSolutions. The smaller treatment-control difference in costs and hospitalizations over the full time period since startup may be due to a number of factors, but it appears to be due primarily to the program's unusual patient intake pattern. During its first year, CorSolutions was only modestly successful

TABLE VI.11

A COMPARISON OF COST NEUTRALITY OVER THE 25 MONTHS SINCE PROGRAM STARTUP AND OVER THE YEAR AFTER ENROLLMENT

Cost Neutrality, Year After Enrollment	Cost Neutrality, 25 Months Since Startup: Possibly	Probably Not	Definitely Not
Possibly	QMed Quality Oncology	Medical Care Development CorSolutions	Avera
Probably Not	Georgetown	University of Maryland Jewish Home and Hospital	
Definitely Not	Mercy Hospice of the Valley		Charlestown Washington University Carle CenVaNet Health Quality Partners

in enrolling patients (619 after the first 12 months). However, the program implemented an aggressive and successful approach to enrollment at about the time of its first anniversary. Enrollment then picked up markedly over the next several months. As a result, the average length of enrollment for CorSolutions's patients in the 25-month analysis was only 9.7 months. If few effects are likely to be observed on patients until they have been enrolled for a full year, CorSolutions's high proportion of patients in the 25-month analysis who are recent enrollees would suggest that the estimated treatment-control differences are attenuated by these short-tenure patients.[11] Future analyses of data on patient contacts and for a longer period will help to sort out these competing explanations.

[11] Alternatively, if the program diverted resources from intensive care coordination during the early period to more recruiting and enrollment, or if it did not have enough nurses to deliver interventions of the same intensity (or the same quality, if the new nurses were less experienced), a decrease in impacts on patient hospitalizations and costs would be expected. However, these explanations seem unlikely because CorSolutions did not use the same staff for patient recruiting and intervention delivery, and because the program drew its nurses from a large telephone call center.

VII. SUMMARY AND CONCLUSION

Patients and physicians were generally very satisfied with the program, but few programs had statistically detectable effects on patients' behavior or use of Medicare services. Treating only statistically significant treatment-control differences as evidence of program effects, the results show:

- Few effects on beneficiaries' overall satisfaction with care
- A sizable increase in the percentage of beneficiaries reporting they received health education on various topics, including diet, exercise, warning signs, and their condition or treatment
- No clear effects on patients' adherence or self-care
- Favorable effects for only two programs each on the quality of preventive care, the number of preventable hospitalizations, and patients' well-being
- Reduced hospitalizations for only 1 of the 15 programs over the first 25 months of program operations (the average number of hospitalizations was 10 percent or more lower for the treatment than control groups in another 4 programs, but these differences were not statistically significant)
- No reduction in expenditures for Medicare Part A and B services for any program

Despite the absence of statistically significant treatment-control differences on Medicare expenditures for traditional services, it is possible that some of the programs are cost neutral to date. This could be true because the large variation in Medicare expenditures and the small number of beneficiaries enrolled in some programs make it difficult to draw definitive conclusions—for nine programs, treatment-control differences in Part A and B expenditures are not statistically different from zero, but they also are not significantly different from the average fee paid to the program. Based on the patterns of differences in hospitalizations, Part A and B expenditures, and Medicare expenditures including the care coordination fees, six of the

programs are not cost neutral, four probably are not, and five may be over the first 25 months since program startup.

A. NO SINGLE FACTOR STANDS OUT AS KEY TO A SUCCESSFUL INTERVENTION

Given the limited number of programs that show any promise of reducing beneficiaries' need for hospitalizations and saving money, or of improving the quality of care they receive, there is relatively little assessment to be done of "what works." The one program for which there were statistically significant estimates of reductions in hospital use, Mercy Medical Center in Iowa, differed from the other programs in that it had by far the highest proportion of contacts conducted in person (two-thirds), and it excelled at problem identification and care planning, patient education, and improving communications and coordination between patients and physicians. Its staffing was also rated in the top quintile. The program also had large impacts on patient education, as judged from the patient survey, and was rated highly by the patients' physicians.

In the evaluator's discussions with the programs on the reasons for their effects or lack of effects to date, Mercy's staff attributed the reductions in hospitalizations they achieved primarily to getting patients to see their physician quickly when symptoms worsened or problems arose. By identifying looming problems before they became severe, and convincing patients of the urgency of seeing a physician (or contacting physicians directly on behalf of patients when necessary), Mercy staff felt they were able to prevent the patients' health from deteriorating to the point where a hospital admission would be necessary. They felt this preventive effect typically arose through quickly getting patients on needed medications or different dosages of their current medications.

The four other programs for which the treatment group had 10- to 20-percent fewer hospitalizations than the control group (though these differences were not statistically significant) also scored highly on one or more domains. For example, Georgetown and Health Quality Partners both scored in the top quintile on initial assessment. Quality Oncology scored in the top quintile on four domains—Staffing, Information Technology, Ongoing Monitoring, and Quality Management.

Programs that seemed to improve preventive care (Carle and Health Quality Partners) also scored very well on patient survey indicators and tended to receive high ratings on the scoring algorithm for the quality of their intervention. Carle scored higher than all other programs on 5 of the 10 indicators. Health Quality Partners scored at the top on patient education. However, neither of these programs generated reductions in Medicare expenditures in the observed follow-up period. This lack of impact on expenditures may be due to the fact that these two programs' enrollees had far lower preenrollment expenditures than any other program's enrollees (except for QMed, which was comparable).

Programs that exhibited no effects on hospitalizations, costs, or quality-of-care indicators gave a range of reasons why they were unable to reduce the need for hospitalizations. Reasons included the still-short time frame over which the analysis was conducted; the belief that some of their patients were either too debilitated or not sick enough to benefit from their interventions; and the belief that physicians in their service areas had an intractable tendency to send patients to the emergency room, rather than to find time for office visits when patients exhibited worsening symptoms.

Looking across the characteristics of the five programs most likely to be cost neutral over the first 2 years of operation and of the two that appear to have improved the quality of care seems to confirm the finding in Chen et al. (2000) that no single program feature or characteristic

seems to be associated with a greater likelihood of program "success." Nor does the absence of a particular feature seem to doom a program to relative failure. However, how *well* programs perform their functions (based on information obtained from program staff and assessed by the evaluator) does seem to be associated with program success.

While no firm conclusions can be drawn as yet about which Medicare Coordinated Care Demonstration programs really are effective, because samples are still relatively small and the follow-up period relatively short, those programs that are most promising to date share few common structural features. Two of the programs with the most success in improving quality (Health Quality Partners and Carle) operate in rural areas, as does Mercy, the sole program with statistically significant effects on the number of hospitalizations. Yet Avera and Medical Care Development also operate in rural areas and show no such promising results to date. Two of the programs with the most favorable expenditure results (Quality Oncology and Georgetown) have fewer than 100 treatment group members—Medical Care Development is the only other program serving fewer than 300 patients. However, the results for these two programs may be due more to the imprecision of the estimates than to the excellence of the interventions. The five other relatively successful programs have substantially more patients. All four programs whose care coordinators have average caseloads of 50 or fewer patients are among the most effective programs, but the other three relatively effective programs have average caseloads in the highest range (over 75 patients). Three of the five programs operated by commercial disease management programs were among the top seven performers, but the four other strong performers had hospitals, clinics, or academic medical centers as hosts. Other program characteristics examined seem equally unrelated to whether a program was one of the more effective seven.

The relationship between how well programs reportedly performed certain functions appears to have a somewhat stronger association with performance than the structural characteristics. Strong performance in any particular domain does not appear to be necessary or sufficient for a program to be relatively successful. However, there are some clear patterns of association between how programs scored on the 10 domains examined and their ability to improve quality or generate reasonably favorable expenditure comparisons. The domains most strongly associated with the promising programs are Staffing (the five programs with the highest ratings on staffing were all among the seven most effective programs), Improving Communications and Coordination (five of the six top programs on this domain were promising programs), Patient Education (four of the top five programs were promising), and Quality Management and Outcome Measurement (four of the top five programs were promising). Characteristics decidedly *not* associated with stronger quality or cost performance included Improving Provider Practice, Service and Resource Arranging, Information Technology, and (perhaps surprisingly) Ongoing Monitoring. For each of these characteristics, only one or two of the five top-rated programs were among the seven programs classified as most promising to date.

Finally, the characteristics of the patients enrolled appear to be unrelated to the relative success of the programs to date. Three of the seven promising programs targeted patients with a single disease; the other four targeted multiple diseases. All three of the programs that enrolled patients with average preenrollment Medicare expenditures of under $600 per month were among the top seven performers, but three others of the top performers were among the six programs whose patients had average expenditures in excess of $2,000 per month. None of the other patient characteristics examined (age, education, income, or race) appeared to be related to programs' likelihood of success.

The current findings suggest that hiring excellent staff and performing certain key functions well are the most important determinants of the likelihood that a program might successfully improve patient outcomes or save enough in Medicare expenditures to cover the cost of its intervention. The results to date are thus consistent with findings from Chen et al. (2000) that a few factors were common to most successful programs, including hiring well-trained, experienced nurses with at least a baccalaureate degree, but that many other factors, such as having sophisticated electronic health records, were not required.

B. THE FINDINGS FOR THE FIRST 2 YEARS ARE NOT HIGHLY FAVORABLE, BUT THEY COULD IMPROVE

Although none of the impact estimates available at this time suggest that the demonstration programs are having large effects on patients' behavior or outcomes, effects on Medicare service use and expenditures might be observed when the full 4 years of data on all patients become available. Physicians have been responding favorably to the programs—an important factor, given the widespread recognition that few care coordination programs are likely to succeed without significant cooperation and reinforcement from patients' physicians (Chen et al. 2000; Schore et al. 1999). Even more important, patients appear to have formed a bond with their care coordinators and trust their judgment.

The absence of large effects on the patient adherence measures may be somewhat discouraging for programs, but it does not necessarily imply that the programs are not having any effect on patients' behavior. Relative to the control group, patients of several programs reported better access to information and appointments and better communication among their providers. Furthermore, the finding that program patients were not significantly more likely to report eating a healthy diet or exercising regularly may have a positive explanation—it is possible that, as a result of program education, the treatment group had higher standards of what constitutes

"healthy" or "regular." If that is true, their actual adherence may be better than the control group's, but the survey measures reported here may not reflect it. In addition, in many cases, behavioral change takes time; some changes do not occur until patients have experienced an adverse event that makes them recognize the value of adhering to advice from their physicians or care coordinators. Programs report that they expect it to take a few years to observe changes in their patients' behavior and the effects of those behaviors on their health and service use. The observed improvements in preventive care in some programs also may not result in lower hospitalizations or costs for a few years. Thus, there is reason to believe that some programs may have effects over the longer run.

C. THE FINAL REPORT WILL COVER 4 PROGRAM YEARS

The results presented here are not the final word on the programs' impacts—changing ingrained behaviors of physicians and patients and improving communications among non-integrated fee-for-service providers are all difficult tasks to achieve. Furthermore, even if achieved, such improvements in the processes of care may not yield statistically discernable improvements in patient well-being or reductions in Medicare costs over the first 2 years of program operations. Thus, the estimates presented here may differ from those that will be observed over the full 4 years of operations. Nonetheless, this report provides (to our knowledge) the largest single random-assignment study to date of disease management/case management programs, and only the second evaluation ever conducted of such programs in a Medicare fee-for-service setting. (The first was by Schore et al. 1999.)

The next evaluation report will assess the effectiveness of the demonstration programs by estimating program impacts on Medicare service use, expenditures, and quality of care over the first 4 years of program operations. The report will also describe the features of the programs or target populations associated with effectiveness (if any). CMS has extended the end dates by 2

years to 2008 for all 11 of the 15 demonstration programs that requested extensions. The remaining programs (the three smallest programs and Charlestown, which is starting a new demonstration program) ended in 2006 as originally planned. CMS granted the extensions because the Balanced Budget Act of 1997 authorizes CMS to continue any programs that are found to be cost-effective after the demonstration ends. The Act defines cost-effectiveness as either (1) reducing Medicare expenditures, or (2) not increasing Medicare expenditures while increasing the quality of services furnished and beneficiary and provider satisfaction. The new end dates allow 11 of the demonstration programs to continue operating until the final evaluation findings are available. The extension allows any of these programs that the final report finds to be cost effective to remain operating rather than shutting down in 2006 and having to restart later.

REFERENCES

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

Aubry, Barbara. "Bolstering Disease Management Programs." *Healthplan,* July–August 2000, pp. 11–12.

Bodenheimer, Thomas. "Disease Management—Promises and Pitfalls." *New England Journal of Medicine*, vol. 340, no. 15, April 15, 1999, pp. 1202–1205.

Bodenheimer, Thomas, Edward Wagner, and Kevin Grumbach. "Improving Primary Care for Patients with Chronic Illness." *Journal of the American Medical Association*, vol. 288, no. 15, October 16, 2002, pp. 1909–1914.

Brown, Randall, Deborah Peikes, Arnold Chen, Judy Ng, Jennifer Schore, and Clara Soh. "Second Report to Congress on the Evaluation of the Medicare Coordinated Care Demonstration." Princeton, NJ: Mathematica Policy Research, Inc., December 2006.

Brown, Randall, Jennifer Schore, Nancy Archibald, Arnold Chen, Deborah Peikes, Karen Sautter, Bridget Lavin, Sherry Aliotta, and Todd Ensor. "Coordinating Care for Medicare Beneficaries: Early Experiences of 15 Demonstration Programs, Their Patients, and Providers." Report to Congress. Princeton, NJ: Mathematica Policy Research, Inc., May 2004.

Brown, Randall, Deborah Peikes, Eric Schone, Nazmul Khan, Arnie Aldridge, and Lucy Lu. "Waiver Cost Estimates for the Medicare Coordinated Care Demonstration." Princeton, NJ: Mathematica Policy Research, Inc., August 31, 2001.

Casalino, Lawrence. "Disease Management and the Organization of Physician Practice." *Journal of the American Medical Association*, vol. 293, no. 4, January 26, 2005, pp. 485–488.

Centers for Medicare & Medicaid Services. "Medicare Current Beneficiary Survey." Baltimore, MD: CMS, 2002a. Available at [www.cms.hhs.gov/mcbs/PublDT.asp].

Centers for Medicare & Medicaid Services. "Program Information on Medicare, Medicaid, SCHIP, and Other Programs of the Centers for Medicare & Medicaid Services." Baltimore, MD: CMS, June 2002b.

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.

Congressional Budget Office. "High-Cost Medicare Beneficiaries." Washington, DC: CBO, May 2005.

Congressional Budget Office. "An Analysis of the Literature on Disease Management Programs." Washington, DC: CBO, October 13, 2004. Available at [http://www.cbo.gov/ftpdocs/59xx/doc5909/10-13-DiseaseMngmnt.pdf].

Cubanski, Juliette, Molly Voris, Michelle Kitchman, Tricia Neuman, and Lisa Potetz. *Medicare Chartbook.* Third edition. Menlo Park, CA: The Henry J. Kaiser Family Foundation, summer 2005.

Department of Obstetrics and Gynaecology, Chinese University of Hong Kong. "IntraClass Correlation". Available at [http://department.obg.cuhk.edu.hk/researchsupport/IntraClass_correlation.asp]. Accessed September 28, 2005.

Fox, Peter. "Screening: The Key to Early Intervention for High-Risk Seniors." *Healthplan,* November–December 2000, pp. 56–61.

Guterman, Stuart. "Medicare Disease Management Activities." Paper presented at Invitational Working Meeting for Developing a Research and Data Strategy for Improving Care of Chronically Ill Medicare Beneficiaries, Washington, DC, July 1, 2004.

Hagland, Mark. "Integrating Disease Management." *Healthplan,* January–February 2000, pp. 43–46.

Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, DC: National Academy Press, 2001.

Kerr, Eve, Elizabeth McGlynn, John Adams, Joan Keesey, and Steven Asch. "Profiling the Quality of Care in 12 Communities: Results from the CQI Study." *Health Affairs*, vol. 23, no. 3, May/June 2004, pp. 247–256.

Kozak, Lola Jean, Margaret J. Hall, and Maria F. Owings. "Trends in Avoidable Hospitalizations, 1980–1998." *Health Affairs*, vol. 20, no. 2, March–April 2001, pp. 225–232.

Landis, J. Richard and Gary G. Koch. "The Measurement of Observer Agreement for Categorical Data." *Biometrics*, vol. 33, no. 1, March 1977, pp. 159–174.

Lorig, Kate. "Self-Management Education: More Than a Nice Extra." *Medical Care*, vol. 41, no. 6, June 2003, pp. 699–701.

Lorig, Kate, David Sobel, Anita Stewart, Byron Brown, Jr., Albert Bandura, Philip Ritter, Virginia Gonzalez, Diana Laurent, and Halsted Holman. "Evidence Suggesting that a Chronic Disease Self-Management Program Can Improve Health Status While Reducing Hospitalization." *Medical Care,* vol. 37, no. 1, 1999, pp. 5–14.

Lowenstein, Roger. "The Quality Cure?" *New York Times Magazine*, March 13, 2005.

McCall, Nancy T., Erica Brody, Lee Mobley, and Sujha Subramanian. "Investigation of Increasing Rates of Hospitalizations for Ambulatory Care Sensitive Conditions Among

Medicare Fee-for-Service Beneficiaries." Washington, DC: RTI International, June 2004. Available at [www.cms.hhs.gov/researchers/reports/2004/McCall.pdf].

McGraw, Kenneth O. and S. P. Wong. "Forming Inferences About Some Intraclass Correlation Coefficients." *Psychological Methods*, vol. 1, no. 1, 1996, pp. 30–46.

Medicare Payment Advisory Commission. "Report to Congress: New Approaches to Medicare." Washington, DC: MedPAC, June 2004. Available at [www.medpac.gov/publications/congressional_reports/June04_Entire_Report.pdf].

Moreno, Lorenzo. "Electronic Health Records: Synthesizing Recent Evidence and Current Policy." *Issue Brief,* no. 1. Princeton, NJ: Mathematica Policy Research, Inc., August 2005. Available at [www.mathematica-mpr.com/publications/pdfs/electronichealth.pdf].

Rector, Thomas, and Patricia Venus. "Judging the Value of Population-Based Disease Management." *Inquiry,* vol. 36, summer 1999, pp. 122–126.

Rich, Michael W. "Heart Failure in the Elderly: Strategies to Optimize Outpatient Control and Reduce Hospitalizations." *American Journal of Geriatric Cardiology*, vol. 12, no. 1, January 2003, pp. 19–24.

Rich, Michael W., Valerie Beckman, Carol Wittenberg, Charles L. Leven, Kenneth E. Freedland, and Robert M. Carney. "A Multidisciplinary Intervention to Prevent the Readmissions of Elderly Patients with Congestive Heart Failure." *New England Journal of Medicine*, vol. 333, no. 18, November 2, 1995, pp. 1190–1195.

Riegel, Barbara, Beverly Carlson, Zoe Kopp, Barbara LePetri, Dale Glaser, and Alan Unger. "Effect of a Standardized Nurse Case-Management Telephone Intervention on Resource Use in Patients with Chronic Heart Failure." *Archives of Internal Medicine*, vol. 162, no. 6, March 25, 2002, pp. 705–712.

Roter, Debra, Judith Hall, Rolande Merisca, Beth Nordstrom, Deborah Cretin, and Bonnie Svarstad. "Effectiveness of Interventions to Improve Patient Compliance." *Medical Care,* vol. 36, no. 8, 1998, pp. 1138–1161.

Schore, Jennifer, Randy Brown, and Valerie Cheh. "Case Management for High-Cost Medicare Beneficiaries." *Health Care Financing Review,* vol. 20, no. 4, summer 1999, pp. 87–101.

Shrout, Patrick E. and Joseph L. Fleiss. "Intraclass Correlations: Uses in Assessing Reliability." *Psychological Bulletin,* vol. 86, 1979, pp. 420–428.

Sidorov, Joan, Robert Shull, Janet Tomcavage, Sabrina Girolami, Nadine Lawton, and Ronald Harris. "Does Diabetes Disease Management Save Money and Improve Outcomes? A Report of Simultaneous Short-Term Savings and Quality Improvement Associated with a Health Maintenance Organization-Sponsored Disease Management Program Among Patients Fulfilling Health Employer Data and Information Set Criteria." *Diabetes Care*, vol. 25, no. 4, April 2002, pp. 684–689.

U.S. Department of Health and Human Services. "Medicare and Medicaid Statistical Supplement, 2001." *Health Care Financing Review,* April 2003, Table 12.

Vernarec, Emil. "Health Care Power Shifts to the People." *Business and Health—Special Issue: The State of Health Care in America 1999,* vol. 17, no. 6A, June 1999, pp. 8–13.

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.

Wasson, J., C. Gaudette, F. Whaley, A. Sauvigne, P. Baribeau, and G. Welch. "Telephone Care as a Substitute for Routine Clinic Followup." *Journal of the American Medical Association*, vol. 267, no. 13, April 1, 1992, pp. 1788–1793.

Wenger, Neil S., David H. Solomon, Carol P. Roth, Catherine H. MacLean, Debra Saliba, Caren J. Kamberg, Laurence Z. Rubenstein, Roy T. Young, Elizabeth M. Sloss, Rachel Louie, John Adams, John T. Chang, Patricia J. Venus, John F. Schnelle, and Paul G. Shekelle. "The Quality of Medical Care Provided to Vulnerable Community-Dwelling Older Patients." *Annals of Internal Medicine*, vol. 139, no. 9, November 2003, pp. 740–759.

West, J., N. Miller, K. Parker, D. Senneca, G. Gahandour, M. Clark, G. Greenwald, R. Heller, M. Fowler, and R. Debusk. "A Comprehensive Management System for Heart Failure Improves Clinical Outcomes and Reduces Medical Resource Utilization." *American Journal of Cardiology*, vol. 79, January 1, 1997, pp. 58–63.

Williams, Mark. "Chronic Care Clinics: Why Don't They Work?" *Journal of the American Geriatric Society,* vol. 47, no. 7, July 1999, pp. 908–909.