MATHEMATICA Policy Research

REPORT

FINAL REPORT

# Evaluation of the Medicare Coordinated Care Demonstration: Final Report for the Health Quality Partners' Program

October 28, 2015

Greg Peterson
Jelena Zurovac
Anne Mutti
Cara Stepanczuk
Randall Brown

Submitted to:
U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
Center for Medicare & Medicaid Innovation
7500 Security Boulevard
Baltimore, MD 21244-1850

Project Officer: Patricia Markovich
Contract Number: HHSM-500-2014-00429G

Submitted by:
Mathematica Policy Research
1100 1st Street, NE
12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director and Co-Principal Investigator: Greg Peterson
Deputy Project Director and Co-Principal Investigator: Jelena Zurovac
Reference Number: 50005.00.400.471.000

This page has been left blank for double-sided copying.

## ACKNOWLEDGMENTS

The authors would like to thank the many people who contributed to this report. Mathematica staff who worked on the project include Carol Razafindrakoto, Mark Flick, Xiaofan Sun, Lisa Shang, and Huihua Lu, all of whom provided diligent programming support; Bridget Gutierrez, Sharon Peters, and Amanda Bernhardt, who edited the report; and Sharon Clark, who carefully produced this report. We also thank Mathematica staff who led the earlier phases of the evaluation, including Debbie Peikes, Jennifer Schore, and Arnold Chen. We thank Ken Coburn, Sherry Marcantonio, and their staff at Health Quality Partners (HQP), who took the time from their busy schedules to meet with us multiple times over the life of the project and supplied the contacts data sets on which several of the analyses in this report were based. We thank the Medicare beneficiaries who participated in the HQP program. We also wish to thank Patricia Markovich and William Clark at the Centers for Medicare & Medicaid Services for serving as project officers and reviewing this report.

This page has been left blank for double-sided copying.

## CONTENTS

## TABLES

This page has been left blank for double-sided copying.

# EXECUTIVE SUMMARY

This report presents final evaluation results for the Medicare Coordinated Care Demonstration (MCCD). The MCCD, authorized by the Balanced Budget Act of 1997 (BBA), was a large-scale randomized trial that the Centers for Medicare & Medicaid Services (CMS) conducted between 2002 and 2014 to rigorously assess the impacts of care coordination programs for Medicare fee-for-service (FFS) beneficiaries with chronic illnesses. The BBA authorized CMS to extend or expand programs that an independent evaluation found were able to (1) reduce Medicare Part A and Part B expenditures or (2) improve care quality and patient and provider satisfaction without raising total expenditures. Based on favorable impact results, CMS extended one program—Health Quality Partners (HQP)—several times. The program ran for a total of 12 years. This report focuses on findings for HQP; earlier evaluation reports included all MCCD programs.

**Background.** From 2002 to 2010, HQP did not reduce hospitalizations or expenditures for its full population. However, for the 15 percent of all enrollees who met "high-risk" criteria that we defined, the treatment group beneficiaries (n=164) had significantly lower hospitalizations (34 percent) than the control group beneficiaries (n=158) and significantly lower Medicare expenditures (22 percent), even after inclusion of program fees (Burwell 2014, based on analyses by Mathematica). These beneficiaries had at least one of three chronic conditions—(1) congestive heart failure (CHF), (2) chronic obstructive pulmonary disease (COPD), (3) or coronary artery disease (CAD)—and one or more hospitalizations in the year prior to enrollment. Furthermore, for the additional 28 percent of enrollees with CAD but no hospitalization in the year before enrollment (the "CAD-only" group), HQP treatment group enrollees had significantly lower two-year mortality rates than the control group. However, they did not have significantly fewer hospitalizations nor lower Medicare Part A and B expenditures. In October 2010, CMS extended the HQP program, but only for beneficiaries who met the high-risk or CAD-only criteria. HQP enrolled new beneficiaries (over 99 percent of whom met the high-risk criteria and for whom HQP received $281 per beneficiary per month) and continued to serve the CAD-only and high-risk beneficiaries who had enrolled earlier. Despite the strongly favorable findings before the extension, interim evaluation results since the extension found that the program did not measurably reduce hospitalizations, mortality rates, or emergency department (ED) visits during the first 44 months of the extension. This report extends those interim findings to include beneficiaries who enrolled through the end of the program and outcomes through the end of program operations.

**Research questions**. This evaluation addresses three research questions:

1. What intervention did HQP provide before and after the 2010 extension?
2. What impact did HQP's program have during the extension on mortality, service use, and Medicare expenditures? How do those impacts compare to the impacts for high-risk beneficiaries before the extension?
3. What are likely explanations for any observed differences in impacts before and after the extension?

Addressing these questions can help CMS determine whether to disseminate HQP's model of care management in current care management initiatives that have similar aims of improving quality of care while reducing expenditures, as well as what modifications or cautions to suggest concerning the model.

**Methods**. To describe the intervention before and after the extension (question 1), we relied on prior project reports, recent interviews with HQP program administrators and nurse care managers, and detailed data from HQP on the contacts that nurses had with program enrollees before and after the extension. We estimated program impacts during the extension (question 2) as the regression-adjusted difference in outcomes for beneficiaries randomly assigned to the treatment group, which received program services, and the control group, which did not receive program services. The regressions adjusted for chance differences between patients at baseline and improved the precision of the estimates. All outcomes were developed from Medicare FFS claims and enrollment data. To identify likely explanations for changes in impacts before and after the extension for high-risk enrollees (question 3), we developed five hypotheses (described below), which were informed by discussions with HQP about why impacts might have declined. We tested each hypothesis, to the extent possible, using the HQP contact data or new impact analyses. When comparing the intervention and impacts before and after the extension, we focused on the high-risk population (excluding the CAD-only enrollees) because (1) HQP enrolled only high-risk beneficiaries after the extension, and (2) this is the only group for which HQP reduced Medicare expenditures prior to the extension.

**Question 1: What intervention did HQP provide before and after the 2010 extension?** HQP's intervention contained core elements that remained constant before and after the extension—including frequent in-person contacts between nurses and enrollees (for assessments and patient education), coordination of care with physicians and social services, and use of data to ensure consistent delivery of services. However, after the extension, HQP began identifying prospective enrollees through hospital discharge records rather than through billing data from participating primary care practices, as it had in the pre-extension period. The change was made to ensure that all new enrollees met the high-risk criteria. This change in identification method led to substantial differences between the pre-extension and post-extension high-risk enrollees on characteristics and service use measured during the year prior to their enrollment. On average, the 483 high-risk beneficiaries (treatment and control) who enrolled after the extension were older, had more complex medical conditions, and were more likely to have used home health or skilled nursing facilities than high-risk beneficiaries who enrolled prior to the extension. To meet the needs of a more uniformly high-risk patient population after the extension, HQP reduced caseloads (from 108 to 75 cases per nurse), increased the timeliness of assessments and visits following hospital stays, and spent more time addressing psychosocial needs. HQP also offered fewer group classes because the population was more homebound and less able to participate in the classes. Program contact data showed that after the extension HQP provided more intensive services to the high-risk population than it did before the extension, providing 37 percent more total contacts and 135 percent more in-person contacts in the first year after patient enrollment.

**Question 2: What impact did HQP's program have during the extension on mortality, service use, and Medicare expenditures? How do those impacts compare to the impacts for high-risk beneficiaries before the extension?** During the extension, we found no measurable

differences between the treatment and control groups for hospitalizations, outpatient ED visits, two-year mortality rates, and Medicare expenditures (Part A and B) or any expenditure component. Because there was no reduction in service use to offset program fees that averaged $260 per member per month, the program increased total Medicare expenditures by an estimated 16 percent ($p = 0.08$). These impacts contrast strongly with the positive findings found for high-risk beneficiaries before the extension (described above). The differences between the two samples in impact estimates for hospitalizations, ED visits, and expenditures were not only large in magnitude, they were all statistically significant ($p = 0.003$ for hospitalizations, $p = 0.03$ for expenditures including program fees), indicating that the observed differences were likely real and not due to chance.

**Question 3: What are likely explanations for any observed differences in impacts before and after the extension?** We tested five hypotheses identified in consultation with HQP as potential explanations for the decline in impacts. Below, we summarize each hypothesis and discuss whether the evidence supported it.

1. **Patient tenure in the program was shorter after the extension.** The evidence does not support this hypothesis because program impacts before the extension appeared within the first three years of patient enrollment, but did not do so after the extension over the same length of follow-up.
2. **Usual care for the control group improved, decreasing the marginal benefit of HQP's services for reducing hospitalizations.** There is strong evidence for this hypothesis. Program effects on hospitalizations disappeared in the extension period because the control group's regression-adjusted hospitalization rates were lower (improved) for the extension cohort than the pre-extension cohort, whereas the treatment group's outcomes were essentially the same for the two cohorts.
3. **The population was too ill or complex to benefit from the intervention.** The evidence does not support this hypothesis because after weighting the post-extension population to resemble the pre-extension population on observable baseline characteristics the impact estimates did not improve. However, we were unable to control for unobservable characteristics, such as functional limitations or caregiver status, which may have differed between the pre- and post-extension high-risk groups and could have influenced the ability of the program to improve outcomes.
4. **The intervention became less intensive for high-risk beneficiaries after the extension.** Overall, there is little support for this hypothesis. On most measures, the program became more intensive for the high-risk enrollees after the extension (more total contacts per person, more in-person contacts, more provider contacts—even after accounting for the greater acuity of the high-risk beneficiaries). However, participation in group classes did decline—even after adjusting for changes in the patient population—and this may have contributed to the decline in impacts.
5. **Service disruptions from near shutdowns in 2010 and 2013 made the program less effective.** There is no support for this hypothesis. Although the contact rate did decline for patients during these disruptive periods (by about 10 percent), the impact estimates did not improve after removing these periods when calculating sample members' outcomes for both the treatment and control groups.

Based on these results, the most likely explanation for the decline in impacts is that improvements in usual care[1] decreased the marginal benefit of HQP's services for reducing hospitalizations—although, it is unclear what aspects of usual care improved to reduce hospitalizations for the control group. The improvements may have resulted from the expansion of services that overlap with those provided by HQP, such as hospital-based transitional care or care management offered through medical homes. Or, it could be that general improvements in patient self-care, medications, or other treatment regimens for patients with CHF, COPD, or CAD made HQP's interventions less useful than they were before the extension. The decrease in group classes may have also contributed to the decline in impacts. An important limitation of the analysis is that we can only identify factors that likely contributed to the decline in impacts. It is not possible to know with certainty or even known probability exactly what caused the declines.

**Implications.** The results of this evaluation substantially reduce the confidence that HQP's model of fee-for-service care management can improve quality and reduce Medicare expenditures in today's health care environment, at least in the geographic region in which HQP operated. The most likely explanation for the decline in effects appears to be that, because usual care has improved over time, the gaps in care that HQP met during the pre-extension period (which led to substantial declines in hospitalizations) may not exist to the same degree in the post-extension period. If this were true, it could be very difficult for the HQP model to achieve its earlier impacts again. However, the model could continue to be effective in other regions if the quality gaps that HQP's program fills are as large, or larger, there as they were in HQP's service area before the extension.

[1] By "usual care" we mean the care (medical or supportive services) that Medicare beneficiaries in the control group received. While beneficiaries in the control group were not eligible for HQP's care management services, they were eligible for all medical care regularly covered by Medicare.

# I. INTRODUCTION

The Medicare Coordinated Care Demonstration (MCCD) was a large-scale randomized trial that the Centers for Medicare and Medicaid Services (CMS) conducted between 2002 and 2014 to rigorously assess the impacts of care coordination programs for Medicare fee-for-service (FFS) beneficiaries with chronic illnesses. In June 2013, CMS extended the sole remaining program in the demonstration (out of the original 15 programs) through December 2014. That program, Health Quality Partners (HQP), ended in 2014, after having run for 12 years (2002–2014).

This report presents the final evaluation for the MCCD, focusing on (1) HQP's impacts on service use, survival, and Medicare expenditures during the most recent period of operations (2010-2014), (2) comparing these impacts to those attained earlier in the demonstration, and (3) identifying likely explanations for the changes in impacts.

This page has been left blank for double-sided copying.

## II. BACKGROUND

### A. The Medicare Coordinated Care Demonstration

As mandated by the Balanced Budget Act of 1997 (BBA), the MCCD was designed to test the use of various interventions to improve (1) patient adherence to treatment regimens; (2) physician adherence to treatment guidelines; and (3) coordination among providers and patients for targeted, chronically ill FFS Medicare beneficiaries. The BBA also mandated a concurrent, independent evaluation of the MCCD to continue for the duration of the demonstration and include reports to Congress describing the programs and their impacts. The BBA authorized CMS to extend or expand programs that either reduced Medicare Part A and Part B expenditures or improved care quality and patient and provider satisfaction without raising total expenditures.

From 2002 to 2014, the number of programs participating in the MCCD declined from an initial 15 programs to a single program. Of the original 15 programs, 11 were granted extensions from 2006 to 2008. Two of those programs, which were determined to be either likely or potentially cost neutral, received extensions for another two years from 2008 to 2010. One program (Mercy Medical Center North Iowa) chose to stop in 2010 but the other—Health Quality Partners (HQP)—sought a further extension. In October 2010, HQP was granted a further extension through June 2013, but only for beneficiaries at higher risk of future service use—the group for which the program was effective (as demonstrated by reducing hospitalizations and Medicare expenditures, even after accounting for the monthly program fees that averaged $114 per beneficiary per month [PBPM]). These beneficiaries either (1) had coronary artery disease (CAD), congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), or diabetes and at least one hospital stay in the year before program enrollment (the "high-risk" group, for which CMS paid HQP $281 PBPM during the extension) or (2) had CAD but no hospital stay in the previous year (the "CAD-only" group, for which CMS paid HQP $83 PBPM).[2] HQP could enroll new beneficiaries who met these criteria and continue to serve beneficiaries who enrolled prior to the 2010 extension and met the criteria at enrollment or at the start of the extension. CMS also permitted HQP to expand into new geographic areas in order to test if the program might succeed for a more socioeconomically diverse population than the program's original service area (Doylestown, Pennsylvania). Finally, in June 2013, CMS granted HQP an additional extension through December 2014 to provide for a longer period to evaluate the program's impacts and to allow the enrolled beneficiaries to transition to other care management programs at the demonstration's conclusion. The program ended as scheduled in December 2014.

### B. HQP's impacts

The MCCD evaluation has shown that HQP had highly favorable effects for high-risk patients before the 2010 extension (2002 to 2010) but not thereafter (2010 to 2014). From 2002

[2] CMS paid different rates for these two groups because, based on interim impact results described in Section II.B, the program was expected to generate larger reductions in Part A and B expenditures for the high-risk group than the CAD-only group.

to 2010, for the 15 percent of all enrollees who met high-risk criteria similar to those defined[3] above, HQP reduced hospitalizations by 34 percent and reduced Medicare expenditures (including program fees) by 22 percent (Burwell 2014, based on analysis by Mathematica). For the additional 28 percent of enrollees in the CAD-only group, HQP substantially reduced two-year mortality rates but did not measurably reduce hospitalizations or Medicare Part A and B expenditures (Schore et al. 2011). However, from 2010 to 2014, the program had no measurable impacts on expenditures, hospital use, or mortality for either patient group (Zurovac et al. 2014). This discrepancy raises the question of what factors may be driving the differences in measured performance for these outcomes before and after the 2010 extension.

[3] The only difference in the criteria is that, in Burwell (2014), beneficiaries who had diabetes (but not CAD, CHF, or COPD) and one or more hospitalizations in the year before enrollment were not counted as high risk. These beneficiaries were added to the high-risk definition during the 2010 extension because it was clinically plausible that the program would benefit these beneficiaries. In practice, this change did not substantially increase the size of the high-risk group because most beneficiaries who had diabetes and a prior hospitalization also had one of the other qualifying conditions (CAD, CHF, or COPD) and so already met the high-risk criteria.

## III. EVALUATION GOALS AND RESEARCH QUESTIONS

### A. Goals

CMS had two goals for the evaluation:

1. To use the additional program enrollment and the additional follow-up period (relative to the last report [Zurovac et al. 2014]) to gain a complete picture of HQP's program on service use, Medicare expenditures, and mortality during the extension (2010 to 2014)
2. To compare the impacts during the extension to impacts before the extension for high-risk enrollees and, if they continued to be very different, to determine likely explanations for the differences

Meeting these two goals would help CMS determine how confidently to disseminate HQP's model in current care management initiatives with similar aims to improve quality and lower Medicare expenditures. Furthermore, for those who wished to replicate HQP's early success, the information could help determine what areas they should focus on.

### B. Research questions

To meet these evaluation goals, the evaluation addressed three research questions:

1. What intervention did HQP provide after the extension (2010 to 2014) and how did it compare to the intervention before the extension (2002 to 2010)?
2. What impact did HQP's program have during the extension (2010 to 2014) on mortality, service use, and Medicare expenditures and how do those impacts compare to the impacts for high-risk beneficiaries before the extension?
3. What are likely explanations for any observed differences in impacts before and after the extension? Specifically, to what extent can any or some combination of the following explain any observed differences in impacts:
    - Changes in the patient population or tenure in the program
    - Changes in the intervention
    - Changes in the usual care provided to beneficiaries in the control group
    - Disruptions in the intervention in 2010 and 2013 when the program nearly shut down because it received an extension in the same month it was scheduled to end

This page has been left blank for double-sided copying.

# IV. QUESTION 1: HQP'S INTERVENTION BEFORE AND AFTER THE 2010 EXTENSION

## A. Methods

We used three methods and data sources to address the first question of how HQP's intervention compared before and after the extension. First, we used prior evaluation documents and recent interviews with HQP administrators and nurse care managers to qualitatively describe the intervention and how it changed over time. Second, we analyzed Medicare claims data to describe the characteristics (for example, demographics, chronic conditions, and recent service use) of beneficiaries who enrolled after the extension, all of whom met the high-risk definition. We also compared these beneficiaries to the high-risk subset of all beneficiaries who enrolled before the extension. Third, we analyzed HQP's detailed data set of nurse contacts to determine quantitatively how the number and types of program contacts changed before and after the extension for high-risk beneficiaries. These data sets recorded every contact that the nurse care managers had with or on behalf of program enrollees throughout the life of the program (2002 to 2014), including individual contacts and contacts during group classes. By linking the contact data to the claims data, we could identify the contact type and frequency specifically for high-risk beneficiaries before and after the extension.

## B. Findings

### 1. Program enrollment

#### a. Changes in the method for identifying enrollees

HQP substantially changed its method for identifying prospective enrollees after the extension to ensure that all new enrollees met the high-risk definition established in the 2010 extension. (After the extension, HQP received $281 PBPM to serve each of these high-risk beneficiaries).[4]

Before the extension, HQP identified prospective enrollees by assisting physician practices that had agreed to partner with HQP to run reports from their billing or scheduling systems that would identify patients who were potentially eligible for the program. HQP enrolled beneficiaries with a wide range of chronic conditions (including common and less severe conditions such as hypertension and rarer and more severe conditions such as heart failure). HQP stratified the enrollees into three risk levels, and provided varying levels of services depending on the level. CMS paid HQP different monthly fees depending an enrollee's risk level (ranging from $50 to $130 PBPM).

After the extension, HQP changed to partnering with hospitals and health systems as a way to accurately identify beneficiaries who met the high-risk criteria. HQP enrolled beneficiaries in three steps. The first step was to partner with hospitals to identify patients who might be eligible for the demonstration. HQP negotiated formal collaboration agreements with two new health

[4] The terms of the 2010 extension allowed HQP to enroll CAD-only beneficiaries for whom they would have been paid $83 PBPM; however, HQP decided not to enroll these beneficiaries.

systems—(1) Crozer-Keystone Health System and (2) St. Mary Medical Center—as well as with Doylestown Hospital, with which HQP had a long-standing relationship.[5] From their claims data, hospitals produced quarterly lists of patients who met the criteria of a hospital stay in the prior year along with one of the qualifying diagnoses (listed on claims). HQP also used hospital records to apply exclusion criteria, one of which was that an enrollee had to have identified a primary care provider.[6] A small portion of patients was identified by floor nurses and by practices. The second step was for HQP nurse care managers to contact the primary care providers to ask them to review lists of potential enrollees and remove those who did not meet eligibility criteria. This contact also allowed the nurse care managers to begin developing a collaborative relationship with the providers for managing patients' care. The third step was for HQP nurses to reach out to the patients designated by the physicians as meeting eligibility criteria to describe how HQP's intervention worked and to elicit their voluntary participation.

After the extension, HQP also broadened the geographic area that it served. In the post-extension period, about 60 percent of enrollees were from the same geographic area as served in the pre-extension period (that is, upper Bucks County and Lehigh, Montgomery, and Northampton counties). The remaining 40 percent came from new counties around Philadelphia, Pennsylvania (that is, lower Bucks County and Chester and Delaware counties).

### b. Characteristics of enrollees during the extension

Between October 2010 and June 2014[7], HQP enrolled 483 beneficiaries (treatment and control), 59 percent in the original service area in Doylestown and the remaining 41 percent from the new service areas outside of Philadelphia (Table IV.1). These enrollees were roughly two to three times more likely than the national Medicare FFS average to have CAD, CHF, COPD, or stroke. Their average hospitalization rate in the year before enrollment was 1.7 hospitalizations—almost six times the national average of 0.3. However, only 29 percent of enrollees had been hospitalized for one of the conditions (CAD, CHF, COPD, or diabetes) that qualified them for the high-risk group in the year before enrollment (as indicated by the condition being listed as the primary diagnosis for the hospital stay). About half (52 percent) of enrollees used Medicare home health care in the year before enrollment and 21 percent used a skilled nursing facility (SNF). The extension population was almost exclusively white and non-Hispanic, with few enrollees eligible for Medicaid. The median household income of the zip codes where beneficiaries lived ($79,432) was well above the national average ($51,371) and the unemployment rate was lower (7.6 percent versus 8.1 percent).

[5] Under the pre-extension period, Doylestown Hospital agreed to share some of the costs associated with hiring HQP's nurse care managers.

[6] HQP excluded patients they identified as having amyotrophic lateral sclerosis, Alzheimer's disease, dementia, a diagnosis or history of cancer (other than skin) within the past five years, end-stage renal disease, HIV/AIDS, Huntington's disease, psychoses or schizophrenia, as well as organ transplant candidates and residents of—and those who intended to become residents of—long-term care facilities.

[7] Although HQP continued to serve beneficiaries through December 2014, they were required to stop new enrollment in June 2014 so that all enrollees could potentially be exposed to the intervention for at least six months.

Table IV.1 Pre-enrollment characteristics of high-risk beneficiaries who enrolled before and after the extension (percentages unless otherwise noted)

| | | Medicare FFS average (2012) (n = 32 million) | Before the extension (n = 368) | After the extension (n = 483) | Difference[a] (percentage change) |
|---|---|---|---|---|---|
| Age | < 65 | 16.7 | 0.0 | 0.0 | 0.0 (0.0) |
| | 65–74 | 45.5 | 37.2 | 36.4 | -0.8 (-2.1) |
| | 75–84 | 25.4 | 48.4 | 38.1 | -10.3 (-21.2) |
| | > or = 85 | 12.4 | 14.4 | 25.5 | 11.1 (76.8) |
| Male | | 44.7 | 51.4 | 44.1 | -7.3 (14.1) |
| Race and ethnicity[b] | Black, non-Hispanic | 10.4 | 1.4 | 2.5 | 1.1 (82.9) |
| | Hispanic | 2.6 | 0.3 | 0.2 | -0.1 (-23.8) |
| Medicaid buy-in[c] | | 21.0 | 2.7 | 2.5 | -0.2 (-8.6) |
| Resident of original service area | | n.a. | 100.0 | 58.6 | -41.4 (-41.4) |
| Diagnosis[d] | CAD | 29.8 | 82.9 | 76.6 | -6.3 (-7.6) |
| | CHF | 15.3 | 38.0 | 48.4 | 10.4 (27.3) |
| | Diabetes | 28.0 | 42.4 | 46.2 | 3.8 (8.9) |
| | COPD | 11.8 | 26.1 | 42.0 | 15.9 (61.1) |
| | Cancer[e] | NA | 13.0 | 15.5 | 2.5 (19.0) |
| | Stroke | 4.0 | 12.0 | 12.8 | 0.9 (7.4) |
| | Depression | 15.9 | 14.4 | 25.1 | 10.6 (73.9) |
| | Alzheimer's and dementia | 11.1 | 4.9 | 8.9 | 4.0 (82.0) |
| | Osteoporosis | 6.7 | 17.9 | 20.3 | 2.4 (13.1) |
| | Rheumatoid arthritis | 30.3 | 33.2 | 41.2 | 8.0 (24.3) |
| | CKD | 16.2 | 14.4 | 33.3 | 18.9 (131.4) |
| | Atrial fibrillation | 8.2 | 33.4 | 31.7 | -1.7 (-5.2) |
| Number of chronic conditions (out of 12 above) | | 1.5 | 3.3 | 4.0 | 0.7 (20.9) |
| In the year before enrollment | Annualized hospitalizations (number) | 0.295 | 1.450 | 1.665 | 0.215 (14.8) |
| | Any hospital stay for CAD, CHF, COPD, or diabetes[f] | NA | 32.3 | 29.0 | -3.4 (-10.4) |
| | Any use of home health | NA | 34.8 | 52.0 | 17.2 (49.4) |
| | Any use of a SNF | NA | 7.9 | 21.1 | 13.2 (168.0) |
| Character-istics of beneficiary's zip code of residence | Median household income ($) | 51,371 | 81,742 | 79,432 | -2,310 (-2.8) |
| | College degree or more | 28.5 | 40.8 | 38.3 | -2.4 (-5.9) |
| | Unemployment rate | 8.1 | 7.1 | 7.6 | 0.5 (6.5) |

Sources: Medicare National Claims History File, Standard Analytic File, and Enrollment Databases and the American Community Survey. Medicare FFS totals come from the CCW, Medicare Beneficiary Prevalence for Chronic Conditions for 2003 Through 2012, Table B.2 (https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_b2.pdf). Monthly expenditures and annualized hospitalizations are exceptions and come from the 2013 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, Table V.1 (http://downloads.cms.gov/files/TR2013.pdf) and the Health Indicators Warehouse developed by the National Center for Health Statistics (http://www.healthindicators.gov/Indicators/Hospital-inpatient-Medicare-admissions-per-1000-beneficiaries_2001/Profile/ClassicData), respectively. Household income comes from the 2012 American Community Survey, Table S1901 (Income in the Past 12 Months [in 2012 Inflation-Adjusted Dollars]). Education status comes from the 2012 American Community Survey, 5-

Year Estimates, Table S15003 (Educational Attainment). Unemployment rate comes from the Bureau of Labor Statistics Current Population Survey, Table 1 (http://www.bls.gov/cps/aa2012/cpsaat01.pdf).

Notes: The "before the extension" sample includes beneficiaries who enrolled between April 1, 2002, and March 31, 2010, who also met the "high-risk" criteria at enrollment—that is, they had CAD, CHF, COPD, or diabetes and one or more hospitalizations in the year before enrollment. The "after the extension" includes beneficiaries who enrolled from October 1, 2010, to June 30, 2014, all of whom met the high-risk criteria. The sample sizes include beneficiaries assigned to treatment or control groups.

[a]The difference is the mean for the beneficiaries enrolled after the extension minus the mean for the beneficiaries enrolled before the extension.

[b]Includes all (not only FFS) Medicare beneficiaries who were enrolled on or after January 1, 2012. The total number of beneficiaries is 53.6 million.

[c]Medicaid buy-in indicates that the beneficiary is eligible for both Medicare and Medicaid. The Medicare FFS average was approximated using the percentage of Medicare beneficiaries who were dual eligibles in 2010. See http://kff.org/medicaid/state-indicator/duals-as-a-of-medicare-beneficiaries.

[d]Diagnoses were based on the CCW definitions, version 1.6. The definitions use a look-back period of one year before enrollment for COPD, stroke, and depression and two years for CAD, CHF, and diabetes. The evaluation used a two-year look-back period for dementia, rather than the three years used by CCW, because of the limits of the Medicare claims data extracted for the analysis.

[e]This category excludes skin cancer.

[f]Inpatient hospital stays with CAD, CHF, COPD, or diabetes listed as the primary diagnosis on the inpatient claim, indicating that one of these conditions was the primary reason for the stay. The list of specific diagnoses that counted as CAD, CHF, COPD, or diabetes came from the CCW definitions, version 1.6.

CAD = coronary artery disease; CCW = CMS Chronic Conditions Data Warehouse; CHF = congestive heart failure; CKD = chronic kidney disease; COPD = chronic obstructive pulmonary disease; FFS = fee-for-service; SNF = skilled nursing facility.

NA = not available.

n.a. = not applicable.

### c. Comparison of high-risk enrollees before and after the extension

Before the extension (from April 2002 through March 2010), HQP enrolled 2,219 beneficiaries (treatment and control). Of these, 368 beneficiaries (17 percent) met the high-risk criteria at enrollment, based on our analysis of their Medicare claims before enrollment.

Even though the criteria that we applied retrospectively to identify the subset of pre-extension enrollees who were high-risk were the same criteria HQP used prospectively to identify high-risk beneficiaries after the extension, the pre- and post-extension high-risk groups differed substantially on baseline characteristics. The post-extension high-risk enrollees had, on average, more chronic conditions and were older (Table IV.1). The differences among chronic disease prevalence were largest for chronic kidney disease (CKD) (33.3 percent versus 14.4 percent), depression (25.1 percent versus 14.4 percent), and COPD (42 percent versus 26.1 percent). Moreover, 25.5 percent of post-extension beneficiaries were 85 or older, compared to 14.4 percent of pre-extension beneficiaries. In addition, post-extension enrollees used more SNFs and home health services and had more hospitalizations in the year before enrollment. For example, 52 percent of post-extension enrollees had used Medicare-covered home health services versus 34.8 percent of pre-extension enrollees. Similarly, 21.1 percent of post-extension beneficiaries used Medicare-covered SNF services, compared to 7.9 percent of pre-extension beneficiaries. These prior service use patterns support HQP's impression that post-extension enrollees were more limited functionally than their pre-extension counterparts. Finally, post-extension enrollees had an average of 1.67 hospitalizations compared to an average of 1.45 hospitalizations among pre-extension enrollees.

The older age and complexity of high-risk beneficiaries after the extension is likely driven by the change in the process for enrolling beneficiaries, although the exact mechanism is not clear. HQP suggested that one possible explanation is that, after the extension, program enrollees may have been more likely to use medical specialists as their primary providers (rather than primary care providers). Such beneficiaries are likely to be older and have more complex conditions. Before the extension, HQP identified prospective enrollees who the partnering primary care practices had seen in the last 18 months (based on billing or scheduling records). After the extension, HQP filtered the hospital records to those beneficiaries who listed a primary care provider, but their last visit to that provider could have been years ago. This explanation could be tested by examining claims to see whether visits to primary providers were less frequent and visits to specialists were more frequent after the extension. Such an analysis was beyond the scope of this evaluation.

Although HQP did expand into new areas after the extension, the available measures did not show large changes in the socioeconomic status of enrollees. Medicaid enrollment rates were about the same in the two periods – 2.5 percent and 2.7 before and after the extension, respectively. Similarly, the median household income of the zip codes in which enrollees lived was only slightly lower in the post-extension period than in the pre-extension period ($79,432 versus $81,742).

### 2. Program features

#### a. Core elements that remained the same pre- and post-extension

HQP's intervention retained core features before and after the extension. The intervention consistently centered on nurses engaging with patients one-on-one. The relationship began when the patient agreed to participate. Nurses then conducted an initial assessment with the patient to identify problems. Assessments were conducted in person and, whenever possible, in the patient's home. They served to guide the focus of care. Accordingly, if there were immediate problems that needed to be addressed, the nurse first focused on those. Otherwise, ongoing monitoring and reassessment occurred, with the nurse asking about any changes in patients' conditions, reconciling medications, recommending preventive care, and reviewing action plans that reminded patients what to do if they experienced early warning signs of exacerbations of their conditions.

Patient education was also a focus of these one-on-one sessions—which included the topics of disease etiology, disease signs and symptoms, proper use of medications, nutrition, physical activity, weight loss, self-management skills, home safety, strategies for coping with chronic illness, and availability of community resources. HQP developed disease-specific core curricula as well as many visual aids to help patients understand key messages. Nurse care managers also assisted patients in health behavior change (for example, changing diet to lose weight or reduce sodium intake, increasing physical activity, or quitting smoking) by tailoring their recommendations based on the patient's readiness to change. They also used graphs of a patient's own clinical measures over time as education tools because they visually linked patient behavior to specific outcomes.

HQP nurses also worked with physicians and other health care and social service professionals to coordinate care—for example, to reconcile duplicative or conflicting medication

prescriptions. Most communication was by phone or fax, but sometimes nurses attended physician appointments with patients.

Lastly, HQP's approach was characterized by the use of data to ensure consistent delivery of services. The program used a dashboard of key measures, including process measures and clinical outcomes, to monitor its service provision. If program managers identified gaps in care provided by individual nurses (for example, the average length of time between visits with patients exceeded thresholds), they met with the nurses (in a nonpunitive approach) to address barriers to delivering the intervention as planned.

### b. Programmatic changes after the extension

HQP reported making four management and care protocol changes as it began to serve a more uniformly high-risk population in the post-extension period. First, it increased staffing and decreased caseloads. To accommodate the expected increase in frequency of visits to patients' homes, HQP reduced the target caseload per nurse care manager from 108 to 75 cases per full-time equivalent (FTE).[8] Based on monthly caseload data provided by HQP for 2008 through 2013, HQP generally hit these targets, although there were times caseloads were higher than the targets (especially near or around times when HQP was unsure of whether CMS would soon end the demonstration). In addition, two of the most experienced nurse care managers assumed supervisory roles. HQP also hired 14 additional nurse care managers post-extension. The additional care managers were needed to address staff attrition and to meet the needs of enrollment in new areas.

Second, HQP required care managers to conduct more timely initial assessments with the highest-risk cases (that is, within 7 days of intake) and more timely intervention following hospitalizations (that is, visit within 3 days of discharge). Care managers were also required to increase their use of action plans to communicate goals and steps for patients to take if self-monitoring showed that key indicators had changed for the worse. For example, CHF patients who exceeded their target weight should call the nurse care manager. Recognizing that care management needs continued in nursing home facilities, HQP also began to actively encourage nurse care managers to visit patients in nursing and short-term rehabilitation facilities.

Third, HQP reduced the number of group classes offered. In the pre-extension period, these programs were run by nurse care managers and included classes in general cardiovascular education (the Heart Healthy Workshop), weight loss (the Lifestyle, Exercise, Attitudes, Relationships, Nutrition or LEARN program) and weight maintenance, increasing physical activity (Active Living Every Day, chair exercise classes, walking groups), stress management, and balance and mobility (FallProof). In addition to providing important health education, the classes also served to reduce social isolation among enrollees and associated adverse health effects, such as depression and anxiety. In the post-extension period, however, HQP found that the increased complexity of patient medical problems compared to those of patients in the pre-extension period, and the greater travel distances as HQP expanded into new areas, made the

[8] Although these targets were specific to MCCD enrollees (based on unique MCCD participants per FTE devoted to them), in practice nurse caseloads included both MCCD beneficiaries and Medicare beneficiaries that HQP served through its care management contract with Aetna, which began in 2009.

logistics of offering group classes more challenging. More of the beneficiaries were homebound and unable to attend classes. Obtaining a critical mass of patients within a certain geographic area for certain classes was no longer as feasible. Therefore, HQP only offered Chair Exercise, LEARN weight management, and weight maintenance programs in the post-extension period.

Finally, during the post-extension period, HQP nurse care managers spent more of their time addressing psychosocial needs because more patients had issues with substance abuse, intimate partner and family violence, caregiver stress, anxiety, depression, bipolar disorders, and other psychiatric issues. To address these needs, care managers reported spending more time (1) identifying patients' problems, (2) providing patient and family counseling or education about the conditions and self-care, (3) making referrals for psychiatric and substance abuse services, (4) addressing problems with informal and formal caregivers, and (5) coordinating care with physicians.

### c. Program contacts for high-risk beneficiaries before and after the extension

HQP nurses had higher rates of individual contacts with patients, caregivers, and providers in the post-extension period than in the pre-extension period. When looking at high-risk enrollees in their first year of enrollment, the overall monthly contact rate with individuals (including patients, caregivers, and providers) was 3.22 contacts during the post-extension period—more than 50 percent higher than the pre-extension rate of 2.13 contacts. The rates for the post-extension sample were also higher than rates for the pre-extension sample in the second year of enrollment (by 60 percent) and the third year (by 29 percent). The average number of monthly individual visits that occurred in person was 135 percent higher in the post-extension period than in the pre-extension period (1.6 visits versus 0.68 visits) in the first year after enrollment. That difference was even larger among beneficiaries in the second year of enrollment (Table IV.2).

Not surprisingly, HQP nurses had fewer group visits in the post-extension period—reflecting not only HQP's decision to offer fewer group education classes but also beneficiaries not attending those classes as frequently. Group visits declined by 81.8 percent in the first year after enrollment from a rate of 0.44 to 0.08 group contacts per month (Table IV.2). The lower rates persisted regardless of year of enrollment. Far fewer beneficiaries participated in group classes in the first year after enrollment during the post-extension period (12 percent versus 43 percent for the pre-extension period).

In looking just at nurse contacts with individuals (Table IV.3), the proportions of contacts with patients, caregivers, and providers also changed from the pre-extension period to the post-extension period.[9] A greater proportion of contacts was made with providers and caregivers and a smaller proportion was made with patients by themselves. For example, the portion of contacts devoted to caregiver contact nearly doubled. This change in distribution of visits likely reflects the realities of working with a sicker patient population with greater health care needs and greater likelihood of dementia and depression.

[9] Contacts with individuals are generally one-on-one interactions, but also include contacts where nurses interacted with both patients and their families as well as with both patients and their physicians.

## Table IV.2. Program contacts for high-risk beneficiaries in the treatment group before and after the extension, by year of enrollment

| | | Year after enrollment | Before the extension (April 2002–September 2010) | After the extension (October 2010–December 2014) | Difference (percentage change) |
|---|---|---|---|---|---|
| Contacts per person-month, average number (% of sample with any contact of that type during the enrollment year) | Individual[a] | 1 | 2.13 (100%) | 3.22 (100%) | 1.09 (51.2%) |
| | | 2 | 1.73 (97%) | 2.76 (94%) | 1.03 (59.5%) |
| | | 3 | 1.64 (93%) | 2.11 (92%) | 0.47 (28.7%) |
| | Group[b] | 1 | 0.44 (43%) | 0.08 (12%) | -0.36 (-81.8%) |
| | | 2 | 0.34 (28%) | 0.07 (9%) | -0.27 (-79.4%) |
| | | 3 | 0.38 (26%) | 0.08 (4%) | -0.3 (-78.9%) |
| | Individual or group | 1 | 2.57 (100%) | 3.30 (100%) | 0.73 (28.4%) |
| | | 2 | 2.06 (97%) | 2.83 (94%) | 0.77 (37.4%) |
| | | 3 | 2.02 (93%) | 2.18 (92%) | 0.16 (7.9%) |
| | Individual, in-person[c] | 1 | 0.68 (98%) | 1.60 (100%) | 0.92 (135%) |
| | | 2 | 0.48 (89%) | 1.32 (92%) | 0.84 (175%) |
| | | 3 | 0.49 (88%) | 1.19 (88%) | 0.70 (143%) |
| Sample size | | 1 | 215 | 223 | n.a. |
| | | 2 | 159 | 179 | n.a. |
| | | 3 | 151 | 50 | n.a. |

Source: Authors' analysis of program data collected and provided by Health Quality Partners (HQP).

Notes: In the pre-extension period, the sample for each year of enrollment includes high-risk beneficiaries in the treatment group who (1) were alive at the start of the enrollment year and (2) enrolled early enough for the enrollment year to end before the pre-extension period ended on September 30, 2010. In the post-extension period, the sample for each year of enrollment includes high-risk beneficiaries in the treatment group who (1) were alive at the start of the enrollment year and (2) enrolled after the extension started (on October 1, 2010) and early enough for the enrollment year to end before the program ended on December 31, 2014.

The mean contact rate in each enrollment year was calculated in two steps. First, for each person in the sample, we calculated his or her average number of contacts during the period (which was calculated by summing the number of contacts during the year, dividing by the number of days alive in the year, and multiplying by 30.5). Second, we took the average across all members in the sample for the year. The means include beneficiaries who, although they were alive and enrolled in the program, did not receive any contacts in the year.

Contacts are limited to those that were successful—that is, we removed contacts that the care manager listed as "attempted" or "no show."

[a]Contacts that a nurse care manager had with or on behalf of an individual beneficiary.

[b]A beneficiary's participation in a group class offered by HQP.

[c]Contacts with patients, caregivers, or providers in any of these settings: home, assisted living facility, HQP office visit, inpatient hospital, long-term nursing home, other health care facility, physician office visit, or short-term skilled nursing or rehabilitation facility.

n.a. = not applicable.

Table IV.3. Content of nurse care manager contacts with or on behalf of high-risk beneficiaries in the first year after enrollment

| | | Before the extension (April 2002–September 2010) | After the extension (October 2010–December 2014) | Difference (percentage change) |
|---|---|---|---|---|
| **Any contact** | | | | |
| Mean monthly contacts rate (% of all contacts), by whom the contact was with (mutually exclusive[a]) | Patient | 1.58 (74%) | 2.06 (64%) | 0.48 (30%) |
| | Provider[b] | 0.20 (10%) | 0.49 (15%) | 0.29 (145%) |
| | Caregiver[c] | 0.18 (8%) | 0.50 (16%) | 0.32 (178%) |
| | Other | 0.16 (8%) | 0.17 (5%) | 0.01 (6%) |
| Nurse care manager continuity, %[d] | | 90 | 89 | -1 (-1%) |
| **Contacts with patients** | | | | |
| Mean monthly contact rate (% of all patient contacts), by contact mode (mutually exclusive[e]) | Telephone | 0.81 (51%) | 0.54 (26%) | -0.27 (-33%) |
| | Home visit | 0.43 (28%) | 1.06 (51%) | 0.62 (143%) |
| | Letter or email | 0.11 (7%) | 0.10 (5%) | -0.01 (-9%) |
| | Inpatient visit | < 0.005 (0%) | 0.03 (1%) | 0.02 (504%)[f] |
| | HQP office visit | 0.16 (10%) | 0.08 (4%) | -0.08 (-49%) |
| | Physician visit | 0.04 (3%) | < 0.005 (0%) | -0.04 (-93%) |
| | Short-term SNF, rehab, or nursing home | < 0.005 (0%) | 0.08 (4%) | 0.08 (2,761%)[f] |
| | Administrative[g] | < 0.005 (0%) | 0.17 (8%) | 0.16 (3,971%)[f] |
| | Other | 0.01 (1%) | 0.0 (0%) | -0.01 (-100%) |
| Mean monthly contact rate (% of all patient contacts), by contact content (not mutually exclusive[i]) | Assessment[h] | 0.86 (54%) | 1.46 (71%) | 0.60 (70%) |
| | Monitoring | 1.22 (77%) | 1.55 (75%) | 0.34 (28%) |
| | Education | 1.19 (75%) | 1.52 (74%) | 0.33 (28%) |
| | Identification of services[j] | 0.03 (2%) | 0.05 (3%) | 0.03 (98%) |
| | Action plan | --[k] | 1.29 (63%) | n.a.[l] |
| | Programs applied | 1.17 (74%) | 1.33 (65%) | 0.16 (14%) |
| Mean contact duration, minutes | | 33.0 | 39.6 | 6.7 (21%) |
| **Contacts with providers[b]** | | | | |
| Mean contact rate (% of all provider contacts), by contact mode (mutually exclusive[m]) | Telephone | 0.15 (77%) | 0.38 (77%) | 0.22 (146%) |
| | Letter or email | < 0.005 (2%) | 0.01 (1%) | < 0.005 (30%) |
| | Inpatient visit | < 0.005 (1%) | 0.01 (3%) | 0.01 (1,085%)[f] |
| | HQP office visit | < 0.005 (1%) | < 0.005 (0%) | < 0.005 (-52%) |
| | Physician visit | 0.01 (7%) | 0.02 (4%) | 0.01 (47%) |
| | Short-term SNF, rehab, or nursing home facility | < 0.005 (0%) | 0.02 (4%) | 0.02 (2,328%)[f] |
| | Administrative[g] | 0.02 (12%) | 0.05 (11%) | 0.03 (122%) |
| | Other | < 0.005 (2%) | 0.00 (0%) | < 0.005 (-100%) |
| Mean contact rate (% of all provider contacts), by contact content (not mutually exclusive[n]) | Assessment[h] | 0.01 (7%) | 0.04 (8%) | 0.03 (191%) |
| | Monitoring | 0.14 (69%) | 0.44 (89%) | 0.30 (216%) |
| | Education | 0.03 (15%) | 0.07 (15%) | 0.04 (136%) |
| | Identification of services[j] | 0.01 (6%) | 0.02 (5%) | 0.01 (113%) |
| | Action plan | --[k] | 0.08 (16%) | n.a.[l] |
| | Programs applied | 0.11 (55%) | 0.13 (27%) | 0.02 (22%) |
| Mean contact duration, minutes | | 10.0 | 10.2 | 0.1 (1%) |

Source: Authors' analysis of program data collected and provided by Health Quality Partners.

Notes: The pre-extension sample includes high-risk beneficiaries who enrolled between the program start on April 1, 2002, and September 30, 2009, meaning they could be followed up for a full year before the pre-

extension period ended in September 2010. The post-extension sample includes high-risk beneficiaries who enrolled between October 1, 2010, and December 31, 2013, meaning they could be followed up for a full year before the program ended in December 2014.

The mean contact rate is calculated in two steps: (1) for each person in the sample, calculate his or her average number of contacts during the year (calculated by summing the number of contacts during the year, dividing by the number of days alive in the year, and multiplying by 30.5) and (2) take the average across all members in the sample for the year. The means include beneficiaries who, although they were alive and enrolled in the program, did not receive any contacts in the year.

Contacts are limited to those that were successful—that is, we removed contacts that the care manager listed as "attempted" or "no show."

[a]Because these contacts are mutually exclusive, the sum of the contact rates across the categories equals the total contact rate (and the percentages sum to 100).

[b]Contacts with providers included those that care managers listed as being with health care providers, hospice providers, inpatient case managers, physicians, visiting nurses, or physicians and patients together.

[c]Contacts with caregivers included those that care managers listed as being with family or patient and family.

[d]The mean percentage of beneficiary contacts in the year that are with their primary nurse care manager (defined as the care manager who provided the most contacts in the year). We first calculated, for each person, the percentage of total contacts in the year that were with his or her primary nurse care manager. Then, we took the average across all members in the sample.

[e]Because these contacts are mutually exclusive, the sum of the contact rates across the categories equals the total rate of contacts with patients (and the percentages sum to 100).

[f]These percentages are large because the denominator is very small (< 0.005).

[g]Administrative contacts included those that care managers listed as the following types: administrative, coordination or documentation, scheduling call, fax received, fax sent.

[h]Contacts for assessment included any contact that the nurse care manager indicated was for assessment (initial or reassessment), for example, to administer the Sutter Questionnaire (an assessment tool), a disease-specific health risk assessment, or a care transition assessment.

[i]Because these contacts are not mutually exclusive, the sum of the contact rates across the categories exceeds the total rate of contacts with patients (and the percentages sum to more than 100).

[j]Contacts for identification services included any contact that the nurse care manager indicated was for identifying Medicare-covered services (such as hospice, rehabilitation, or diabetes education) or non-Medicare covered services (such as transport or meal services).

[k]Care managers created action plans during contacts with patients or providers but did not record them in HQP's contact database before the extension.

[l]Percentage change is not applicable because the denominator (mean before the extension) is zero.

[m]Because these contacts are mutually exclusive, the sum of the contact rates across the categories equals the total rate of contacts with providers (and the percentages sum to 100).

[n]Because these contacts are not mutually exclusive, the sum of the contact rates across the categories exceeds the total rate of contacts with providers (and the percentages sum to more than 100).

HQP = Health Quality Partners; SNF = skilled nursing facility.

n.a. = not applicable.

**Nurse contacts with patients.** Looking just at the data on nurse contacts with or on behalf of patients during the first year of enrollment (Table IV.3), we see that the number of monthly individual nurse contacts with patients was higher after the extension than before. This increase was present regardless of year of enrollment (data for years 2 and 3 not shown because patterns were similar across years). In the first year after enrollment post-extension, nurses had an average of 2.06 contacts per month with patients, a 30 percent increase over the pre-extension rate of 1.58 contacts per month.

The number of home visits was much higher in the post-extension period than in the pre-extension period, with 51 percent of all contacts with patients occurring in the patient's home. In

contrast, just 28 percent of all patient contacts were in the patient's home in the pre-extension period. In addition, we see that nurse care managers were far more likely to visit patients in nursing or short-term rehabilitation facilities in the post-extension period. In the first year after enrollment during the post-extension period, 16 percent of patients were visited in these settings, compared to less than 2 percent of their pre-extension counterparts. By the third year after enrollment in the post-extension period, nearly one-quarter of enrollees were visited in a nursing or rehabilitation facility (data not shown).

The time spent during patient contacts was also higher post-extension than pre-extension (40 minutes versus 33 minutes)—perhaps reflecting the greater needs of these sicker patients. The mix of services as defined by the broad categories of assessments, monitoring, and education during individual visits was largely the same pre- and post-extension, except that more time was taken with patient assessments.

Continuity in the relationship with the nurse care manager remained quite high across the pre-extension and post-extension periods (90 percent and 89 percent, respectively). Continuity is measured as the percentage of visits that were with the same nurse in a given year. HQP is concerned that that nurse assignment changes over multiple years may not have been well tolerated by patients. Our measure does not capture changes in assignment over multiple years.

**Nurse contacts with providers.** Nurses' monthly rate of contact with providers, which includes physicians and their staff, inpatient case managers, visiting nurses, and hospice providers, increased from 0.20 contacts in the pre-extension period to 0.49 contacts in the post-extension period (first panel of Table IV.3). The majority of these contacts were made by phone. Indeed, the proportion of contacts with providers by phone remained largely unchanged across the two periods at about 77 percent. Similarly, the mean contact duration remained constant at 10 minutes.

This page has been left blank for double-sided copying.

# V. QUESTION 2: PROGRAM IMPACTS BEFORE AND AFTER THE EXTENSION

## A. Methods

### 1. Impacts after the extension

To assess the impacts of the HQP program during the extension period, we repeated the analyses in the last evaluation report (Zurovac et al. 2014) but used a longer follow-up period, more beneficiaries, and additional outcomes. The follow-up period covered the full length of the extension, from October 1, 2010, to December 31, 2014, adding seven months from the last report. The sample included enrollees through June 30, 2014, adding seven months to the enrollment period and 74 new enrollees (treatment and control). Outcomes included the following:

- All-cause hospitalizations (number per person per year)
- Medicare expenditures (Part A and B without care coordination fees, Part A and B with fees, Part A only) ($ PBPM)
- Medicare expenditures broken out by component—for example, inpatient expenditures, home health care ($ PBPM)
- Outpatient emergency department (ED) visits (number per person per year), which are ED visits that did not end in an inpatient admission (this measure includes observational stays that began as ED visits)
- Two-year mortality rates, in which the sample was limited to those enrolled early enough to be followed up for at least two years

To generate unbiased estimates of program impacts, we used MCCD's randomized design—the gold standard for program evaluation. The impact estimates are the differences in outcomes between the treatment and control groups, adjusted for beneficiary demographics, chronic conditions, and recent service use measured at baseline (the time of program enrollment). We used an intent-to-treat design, following patients for all months that they were alive and enrolled in FFS Medicare after they enroll in HQP's program, regardless of whether they remained active in the program.[10] Before estimating impacts, we verified that the treatment and control groups were similar at baseline on measured demographics, service use, and chronic conditions, as we would expect following random assignment. All outcomes and patient characteristics used as covariates were constructed from Medicare claims and enrollment data. We collected claims updated through March 31, 2015, ensuring at least three months claims run-out from the end of the outcome period on December 31, 2014.

To increase the precision of the estimates and to account for chance differences between the treatment and control groups, we adjusted treatment-control differences for patient

[10] The intent-to-treat design limits the bias in estimates that could result from comparing outcomes for those who actually received treatment to those who did not. In previous analyses, we also explored whether program impacts on mortality or entry into managed care (both of which remove a beneficiary from the sample) could bias the impact estimates for key outcomes. We concluded that the possible bias, if any, is very small and does not drive overall findings (Zurovac et al. 2014).

characteristics measured at baseline. Consistent with prior reports, we used a $p < 0.10$ threshold (two-tailed test) to determine statistical significance. We assessed the extent to which beneficiaries in the treatment group actually received HQP services by calculating the average percentage of a beneficiary's follow-up months for which HQP submitted a bill for services rendered.

As in the last report, we assessed impacts during the extension with two populations: (1) those who had enrolled since the 2010 extension, virtually all of whom met the high-risk definition (n = 483, treatment and control)[11], and (2) those who had enrolled at any time (2002 to 2014) but who met the current eligibility criteria at enrollment and were alive and observable[12] for at least part of the extension period (n = 736). The second approach increases the statistical power to detect effects (because the sample size is larger), but does not exclusively test HQP's current model because the outcomes for earlier enrollees may also be influenced by their time in the program before the extension. We also estimated impacts during the extension for all 1,186 beneficiaries who enrolled at any time from 2002 to 2014 and met the 2010 eligibility criteria at enrollment, which included 736 high-risk beneficiaries as well as 450 beneficiaries who enrolled before the extension but met the CAD-only eligibility criteria at enrollment (that is, they had CAD but no hospitalization in the prior year). Because our focus in this report is in on the high-risk population, we present results for the sample of 1,186 beneficiaries in Appendix A.

To identify which beneficiaries met the high-risk criteria at baseline, we determined eligibility using claims data for pre-extension enrollees and HQP's own method of identification for post-extension enrollees. Before adopting this approach, we verified—using claims—that virtually all (98 percent) of the beneficiaries HQP identified as high risk did in fact meet these criteria.

## 2. Comparison with impacts from before the extension

We compared the impact estimates for high-risk beneficiaries who enrolled after the extension to those for high-risk beneficiaries who enrolled before the extension, as reported in the evaluation's fifth report to Congress (Burwell 2014, based on analysis by Mathematica). The impact estimates before the extension are based on high-risk beneficiaries (n = 322, treatment and control) who enrolled between April 2002 and June 2009, with outcomes measured through June 2010. In addition to comparing point estimates for key outcomes, we formally assessed whether the impact estimates before and after the extension were different from one another. This is important to ensure that measured differences reflect true differences in impacts, not chance differences in populations or outcomes. This statistical assessment requires an estimate of the variance of the difference in impact estimates for the two populations. Because the populations and outcome periods for the pre- and post-extension analysis do not overlap, this variance is equal to the sum of the variances of the pre- and post-extension impact estimates.

[11] HQP also enrolled two beneficiaries who met the CAD-only criteria; these beneficiaries are excluded from the analysis.

[12] For enrollees to be observable, they must be in FFS, have both Part A and B coverage and Medicare as the primary payer, and have been alive for at least part of any one month.

## B. Findings

### 1. Impacts after the extension

#### a. Approach A: Impacts for those who enrolled after the extension

**Balance at baseline.** Between October 1, 2010, and June 30, 2013, HQP enrolled 483 beneficiaries (treatment and control) who met the high-risk definition. The treatment and control groups were similar at baseline on measured demographics, service use, and chronic conditions (Table V.1). (See Chapter IV for a description of enrollee characteristics.)

**Impacts.** We found no measurable differences between the treatment and control groups for hospitalizations, outpatient ED visits, two-year mortality rates, and Medicare expenditures (Part A and B) or any expenditure component (Table V.2). However, due to small sample sizes, these estimates are statistically imprecise, as indicated by large confidence intervals for the differences between the treatment and control groups. For example, the 90 percent confidence interval for hospitalizations ranges from the treatment group outcomes being 12 percent lower (better) than the control group's outcomes to being 29 percent higher (worse). Including program fees that averaged $260 PBPM, the treatment group's total Medicare expenditures were 15.8 percent higher than the control group's expenditures. However, this difference was not statistically significant ($p = 0.17$). On average, HQP provided treatment services to beneficiaries for 93 percent of their follow-up months, verifying that the vast majority of the treatment group received HQP services.

#### b. Approach B: Impacts for those who enrolled at any time

**Baseline characteristics.** Between April 1, 2002, and June 30, 2014, HQP enrolled 736 beneficiaries who met the high-risk criteria at enrollment and who were still enrolled in FFS Medicare for at least one day during the extension (Table V.3). The treatment and control groups were similar at baseline on measured demographics, service use, and chronic conditions, as we would expect following random assignment (Table V.3).

**Impacts.** In terms of hospitalizations, outpatient ED visits, and expenditures without fees (Table V.4), we found no measurable differences between the treatment and control groups during the extension period. These estimates are statistically more precise than those restricted to beneficiaries enrolling after the extension, as reflected in narrower confidence intervals. After factoring in program fees, the program *increased* total Medicare expenditures by an estimated 15.6 percent ($p = 0.08$). HQP provided treatment services to beneficiaries for 85 percent of their follow-up months, verifying that the vast majority of the treatment group received HQP services.

### 2. Comparison with impacts from before the extension

For all outcomes assessed (hospitalizations and Medicare Part A and B expenditures, with and without program fees), the difference in impacts before and after the extension were large and statistically significant ($p$-values ranged from $< 0.01$ for hospitalizations to 0.04 for Part A and B expenditures) (Table V.5). The confidence intervals for these estimates are wide, however, indicating that the true differences in impacts could be considerably smaller or larger than the differences in point estimates.

The mean number of follow-up months was 49.4 before the extension but only 25.5 after the extension, raising the possibility that the difference in impact estimates may be due to differences in patient tenure in the program (a hypothesis we discuss in the next chapter).

The fact that the differences in impact estimates are statistically significant emphasizes that the large observed differences in impacts before versus after the extension are likely real. The challenge becomes explaining what factors likely drove this change, which we turn to next.

Table V.1. Pre-enrollment characteristics for high-risk beneficiaries who enrolled during the extension (2010–2014), by treatment status (percentages unless otherwise noted)

| | | | Health Quality Partners' enrollees | | | |
|---|---|---|---|---|---|---|
| | | Medicare FFS average (2012) (n = 32 million) | Treatment group mean (n = 241) | Control group mean (n = 242) | Difference | *p*-value |
| Age | < 65 | 16.7 | 0.0 | 0.0 | 0.0 | n.a. |
| | 65–74 | 45.5 | 40.2 | 32.6 | 7.6 | 0.22[a] |
| | 75–84 | 25.4 | 35.7 | 40.5 | -4.8 | |
| | > or = 85 | 12.4 | 24.1 | 26.9 | -2.8 | |
| Male | | 44.7 | 46.1 | 42.1 | 3.9 | 0.39 |
| Race and ethnicity[b] | Black, non-Hispanic | 10.4 | 1.7 | 3.3 | -1.6 | 0.25 |
| | Hispanic | 2.6 | 0.0 | 0.4 | -0.4 | 0.32 |
| Medicaid buy-in[c] | | 21.0 | 2.5 | 2.5 | 0.0 | 0.99 |
| Resident of original service area | | n.a. | 58.5 | 58.7 | -0.2 | 0.97 |
| Diagnosis[d] | CAD | 29.8 | 79.3 | 74.0 | 5.3 | 0.17 |
| | CHF | 15.3 | 51.0 | 45.9 | 5.2 | 0.26 |
| | Diabetes | 28.0 | 43.6 | 48.8 | -5.2 | 0.25 |
| | COPD | 11.8 | 42.3 | 41.7 | 0.6 | 0.90 |
| | Cancer[e] | NA | 16.6 | 14.5 | 2.1 | 0.52 |
| | Stroke | 4.0 | 13.7 | 12.0 | 1.7 | 0.58 |
| | Depression | 15.9 | 24.9 | 25.2 | -0.3 | 0.94 |
| | Dementia and Alzheimer's | 11.1 | 9.5 | 8.3 | 1.3 | 0.62 |
| | Osteoporosis | 6.7 | 19.9 | 20.7 | -0.7 | 0.84 |
| | Rheumatoid arthritis | 30.3 | 41.5 | 41.0 | 0.6 | 0.90 |
| | CKD | 16.2 | 34.0 | 32.6 | 1.4 | 0.75 |
| | Atrial fibrillation | 8.2 | 32.4 | 31.0 | 1.4 | 0.75 |
| Number of chronic conditions (of 12 above) | | 1.5 | 4.1 | 4.0 | 0.1 | 0.38 |
| In the year before enrollment | Annualized hospitalizations (number) | 0.295 | 1.604 | 1.726 | -0.123 | 0.23 |
| | Medicare Part A and B expenditures ($ PBPM) | 860 | 2,390 | 2,521 | -131 | 0.51 |
| | Any hospital stay for CAD, CHF, COPD, or diabetes | NA | 29.0 | 25.6 | 3.4 | 0.40 |
| | Any use of home health | NA | 51.9 | 52.1 | -0.2 | 0.97 |
| | Any use of a SNF | NA | 18.7 | 23.6 | -4.9 | 0.19 |

| | | | Health Quality Partners' enrollees | | | |
|---|---|---|---|---|---|---|
| | | Medicare FFS average (2012) (n = 32 million) | Treatment group mean (n = 241) | Control group mean (n = 242) | Difference | *p*-value |
| Characteristics of enrollees' residence zip code, mean | Median household income ($) | 51,371 | 79,715 | 79,151 | 564 | 0.75 |
| | College degree or more | 28.5 | 38.8 | 37.9 | 1.0 | 0.42 |
| | Unemployment rate | 8.1 | 7.5 | 7.6 | -0.1 | 0.76 |

Sources: Medicare National Claims History File, Standard Analytic File, and Enrollment Databases and the American Community Survey. Medicare FFS totals come from the CCW, Medicare Beneficiary Prevalence for Chronic Conditions for 2003 Through 2012, Table B.2 (https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_b2.pdf). Monthly expenditures and annualized hospitalizations are exceptions and come from the 2013 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, Table V.1 (http://downloads.cms.gov/files/TR2013.pdf) and the Health Indicators Warehouse developed by the National Center for Health Statistics (http://www.healthindicators.gov/Indicators/Hospital-inpatient-Medicare-admissions-per-1000-beneficiaries_2001/Profile/ClassicData), respectively. Household income comes from the 2012 American Community Survey, Table S1901 (Income in the Past 12 Months [in 2012 Inflation-Adjusted Dollars]). Education status comes from the 2012 American Community Survey, 5-Year Estimates, Table S15003 (Educational Attainment). Unemployment rate comes from the Bureau of Labor Statistics Current Population Survey, Table 1 (http://www.bls.gov/cps/aa2012/cpsaat01.pdf).

Notes: The sample includes beneficiaries enrolled from October 1, 2010, to June 30, 2014. All beneficiaries met the "high-risk" criteria—that is, they had CAD, CHF, COPD, or diabetes and one or more hospitalizations in the year before enrollment.

[a]Only one *p*-value is reported for the treatment-control differences in age because a chi-squared test was used to determine whether the overall age distribution for the treatment group was different from the distribution for the control group.

[b]Includes all (not only FFS) Medicare beneficiaries who were enrolled on or after January 1, 2012. Total number of beneficiaries is 53.6 million.

[c]Medicaid buy-in indicates that the beneficiary is eligible for both Medicare and Medicaid. The Medicare FFS average was approximated using the percentage of Medicare beneficiaries who were dual eligibles in 2010. See http://kff.org/medicaid/state-indicator/duals-as-a-of-medicare-beneficiaries.

[d]Diagnoses were based on the 2010 version of CCW definitions. The definitions use a look-back period of one year before enrollment for atrial fibrillation, cancer, COPD, depression, osteoporosis, and stroke and two years for CAD, CHF, CKD, diabetes, and rheumatoid arthritis. The evaluation used a two-year look-back period for dementia and Alzheimer's, rather than the three years used by CCW, because of the limits of the Medicare claims data extracted for the analysis.

[e]This category excludes skin cancer.

CAD = coronary artery disease; CCW = CMS Chronic Conditions Data Warehouse; CHF = congestive heart failure; CKD = chronic kidney disease; COPD = chronic obstructive pulmonary disease; FFS = fee-for-service; SNF = skilled nursing facility; PBPM = per beneficiary per month.

NA = not available.
n.a. = not applicable.

Table V.2. Program impacts on service use, Medicare expenditures, and mortality during the extension among high-risk beneficiaries who enrolled during the extension

| | | Control group mean | Treatment group mean | Difference (90% confidence interval) | Percent difference | p- value |
|---|---|---|---|---|---|---|
| **Program impacts on outcomes** | | | | | | |
| Annualized hospitalizations (#/person/year) | | 0.743 | 0.808 | 0.065 (-0.086, 0.216) | 8.8 | 0.48 |
| Annualized outpatient ED visits (#/person/year) | | 0.747 | 0.757 | 0.010 (-0.135, 0.156) | 1.4 | 0.91 |
| Two-year mortality rates[a] (%) | | 15.5 | 14.3 | -1.2 | -7.7 | 0.72 |
| Medicare Part A and B expenditures ($ PBPM) | Without program fees | 1,748 | 1,764 | 16 (-315, 347) | 0.9 | 0.94 |
| | With program fees (mean fee paid = $260) | 1,748 | 2,024 | 276 (-55, 607) | 15.8 | 0.17 |
| Medicare Part A expenditures ($ PBPM) | Total | 943 | 973 | 30 (-234, 295) | 3.2 | 0.85 |
| | Inpatient hospital | 671 | 669 | -2 (-230, 226) | -0.3 | 0.99 |
| | Skilled nursing facility | 179 | 191 | 13 (-53, 79) | 7.1 | 0.75 |
| | Hospice | 33 | 36 | 3 (-26, 32) | 7.8 | 0.88 |
| | Home health (Part A portion) | 56 | 71 | 16 (-2, 34) | 28.3 | 0.15 |
| | Other Part A services | 4 | 5 | 1 (-4, 7) | 35.5 | 0.66 |
| Medicare Part B expenditures ($ PBPM) | Total | 805 | 790 | -14 (-128, 99) | -1.8 | 0.84 |
| | Outpatient facilities (ED) | 35 | 32 | -3 (-15, 9) | -7.8 | 0.71 |
| | Outpatient facilities (not ED) | 161 | 169 | 8 (-36, 53) | 5.1 | 0.76 |
| | Physician services (delivered in the ED or hospital) | 100 | 93 | -7 (-32, 19) | -6.7 | 0.67 |
| | Physician services (delivered outside the ED or hospital) | 125 | 134 | 9 (-4, 21) | 7.0 | 0.25 |
| | Home health (Part B portion) | 66 | 59 | -7 (-36, 21) | -10.8 | 0.68 |
| | Lab work and radiology | 63 | 64 | 1 (-7, 7) | 1.6 | 0.99 |
| | Durable medical equipment | 72 | 48 | -24 (-85, 37) | -33.5 | 0.51 |
| | Other Part B services | 183 | 192 | 9 (-37, 55) | 5.0 | 0.74 |
| **Sample sizes, average length of follow-up, and receipt of treatment services** | | | | | | |
| Number of beneficiaries | | 242 | 241 | n.a. | n.a. | n.a. |
| Mean number of follow-up months | | 25.6 | 25.5 | n.a. | n.a. | n.a. |

| | Control group mean | Treatment group mean | Difference (90% confidence interval) | Percent difference | p- value |
|---|---|---|---|---|---|
| Mean percentage of follow-up months during which beneficiary received treatment[b] | 0.0 | 93.0 | n.a. | n.a. | n.a. |

Sources: Medicare Enrollment Database, National Claims History File, Standard Analytic File, and the Mathematica randomization file.

Notes: The sample includes beneficiaries who enrolled between October 1, 2010, and June 30, 2014, with outcomes measured from the randomization month until December 31, 2014. All beneficiaries in the sample met the high-risk definition—that is, they had CAD, CHF, COPD, or diabetes and one or more hospitalizations in the year before enrollment.

The control group mean is weighted but not regression-adjusted. The treatment group mean equals the control group mean plus the regression-adjusted difference between treatment and control. The percentage difference is calculated by dividing the regression-adjusted difference between the treatment and control groups by the control group mean.

Treatment-control differences were adjusted for baseline characteristics to increase the precision of the estimates and to account for chance differences between the two groups. Negative estimates of treatment-control differences indicate that hospitalizations or expenditures were lower for the treatment group.

For all outcomes except mortality, the regressions use observations that are weighted according to the proportion of the follow-up period during which each sample member met CMS's demonstration-wide requirements: being in fee-for-service, having both Part A and Part B coverage, having Medicare as the primary payer, and being alive for a part of the month. Weights were calculated separately for the treatment and control groups. For the two-year mortality analyses, every beneficiary received a weight of 1 because all sample members could be followed up for at least two years.

[a]The sample for the two-year mortality analysis was limited to beneficiaries who enrolled by December 31, 2012, and could therefore be followed up for a full two years (n = 185 for treatment and n = 184 for control).

[b]Calculated as follows: (1) for each beneficiary, calculate the percentage of follow-up months during which he or she received treatment services (a beneficiary was considered to have received services in a given month if HQP submitted a bill for services for that beneficiary in that month) and (2) find the average across treatment group beneficiaries.

CAD = coronary artery disease; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; ED = emergency department; FFS = fee-for-service; HQP = Health Quality Partners; PBPM = per beneficiary per month.

NA = not available.
n.a. = not applicable.

Table V.3. Pre-enrollment characteristics for high-risk beneficiaries who enrolled at any time during the program (2002–2014), by treatment status (percentages unless otherwise noted)

| | | | Health Quality Partners' enrollees | | | |
|---|---|---|---|---|---|---|
| | | Medicare FFS average (2012) (n = 32 million) | Treatment group mean (n = 374) | Control group mean (n = 362) | Difference | *p*-value |
| Age | < 65 | 16.7 | 0.0 | 0.0 | 0.0 | n.a. |
| | 65–74 | 45.5 | 42.2 | 35.4 | 6.9 | |
| | 75–84 | 25.4 | 38.8 | 42.8 | -4.0 | 0.16[a] |
| | > or = 85 | 12.4 | 19.0 | 21.8 | -2.8 | |
| Male | | 44.7 | 47.6 | 44.5 | 3.1 | 0.40 |
| Race and ethnicity[b] | Black, non-Hispanic | 10.4 | 1.1 | 2.8 | -1.7 | 0.09* |
| | Hispanic | 2.6 | 0.3 | 0.3 | -0.0 | 0.98 |
| Medicaid buy-in[c] | | 21.0 | 2.4 | 2.2 | 0.2 | 0.86 |
| Resident of original service area | | n.a. | 73.3 | 72.4 | 0.9 | 0.79 |
| Diagnosis[d] | CAD | 29.8 | 80.5 | 76.2 | 4.2 | 0.16 |
| | CHF | 15.3 | 45.2 | 40.9 | 4.3 | 0.24 |
| | Diabetes | 28.0 | 42.5 | 45.3 | -2.8 | 0.45 |
| | COPD | 11.8 | 35.0 | 37.0 | -2.0 | 0.57 |
| | Cancer[e] | NA | 16.3 | 12.4 | 3.9 | 0.13 |
| | Stroke | 4.0 | 13.1 | 11.0 | 2.1 | 0.39 |
| | Depression | 15.9 | 21.4 | 21.0 | 0.4 | 0.90 |
| | Dementia and Alzheimer's | 11.1 | 8.0 | 6.4 | 1.7 | 0.38 |
| | Osteoporosis | 6.7 | 20.1 | 19.1 | 1.0 | 0.73 |
| | Rheumatoid arthritis | 30.3 | 39.8 | 37.0 | 2.8 | 0.43 |
| | CKD | 16.2 | 26.5 | 25.1 | 1.3 | 0.68 |
| | Atrial fibrillation | 8.2 | 31.0 | 31.0 | 0.1 | 0.98 |
| Number of chronic conditions (of 12 above) | | 1.5 | 3.8 | 3.6 | 0.2 | 0.16 |
| In the year before enrollment | Annualized hospitalizations (number) | 0.295 | 1.576 | 1.578 | -0.002 | 0.98 |
| | Medicare Part A and B expenditures ($ PBPM) | 860 | 2,188 | 2,246 | -57 | 0.70 |
| | Any hospital stay for CAD, CHF, COPD, or diabetes | NA | 30.2 | 27.1 | 3.1 | 0.35 |
| | Any use of home health | NA | 44.9 | 45.3 | -0.4 | 0.92 |
| | Any use of a SNF | NA | 15.2 | 17.1 | -1.9 | 0.49 |

| | | | Health Quality Partners' enrollees | | | |
|---|---|---|---|---|---|---|
| | | Medicare FFS average (2012) (n = 32 million) | Treatment group mean (n = 374) | Control group mean (n = 362) | Difference | *p*-value |
| Characteristics of enrollees' residence zip code, mean | Median household income ($) | 51,371 | 80,254 | 79,835 | 418 | 0.77 |
| | College degree or more | 28.5 | 39.1 | 38.6 | 0.5 | 0.63 |
| | Unemployment rate | 8.1 | 7.4 | 7.4 | 0 | 0.77 |

Sources: Medicare National Claims History File, Standard Analytic File, and Enrollment Databases and the American Community Survey. Medicare FFS totals come from the CCW, Medicare Beneficiary Prevalence for Chronic Conditions for 2003 Through 2012, Table B.2 (https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_b2.pdf). Monthly expenditures and annualized hospitalizations are exceptions and come from the 2013 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, Table V.1 (http://downloads.cms.gov/files/TR2013.pdf) and the Health Indicators Warehouse developed by the National Center for Health Statistics (http://www.healthindicators.gov/Indicators/Hospital-inpatient-Medicare-admissions-per-1000-beneficiaries_2001/Profile/ClassicData), respectively. Household income comes from the 2012 American Community Survey, Table S1901 (Income in the Past 12 Months [in 2012 Inflation-Adjusted Dollars]). Education status comes from the 2012 American Community Survey, 5-Year Estimates, Table S15003 (Educational Attainment). Unemployment rate comes from the Bureau of Labor Statistics Current Population Survey, Table 1 (http://www.bls.gov/cps/aa2012/cpsaat01.pdf).

Notes: The sample includes beneficiaries enrolled from April 1, 2002, to June 30, 2014. All beneficiaries met the "high-risk" criteria—that is, they had CAD, CHF, COPD, or diabetes and one or more hospitalizations in the year before enrollment.

[a]Only one *p*-value is reported for the treatment-control differences in age because a chi-squared test was used to determine whether the overall age distribution for the treatment group was different from the distribution for the control group.

[b]Includes all (not only FFS) Medicare beneficiaries who were enrolled on or after January 1, 2012. Total number of beneficiaries is 53.6 million.

[c]Medicaid buy-in indicates that the beneficiary is eligible for both Medicare and Medicaid. The Medicare FFS average was approximated using the percentage of Medicare beneficiaries who were dual eligibles in 2010. See http://kff.org/medicaid/state-indicator/duals-as-a-of-medicare-beneficiaries.

[d]Diagnoses were based on the 2010 version of CCW definitions. The definitions use a look-back period of one year before enrollment for atrial fibrillation, cancer, COPD, depression, osteoporosis, and stroke and two years for CAD, CHF, CKD, diabetes, and rheumatoid arthritis. The evaluation used a two-year look-back period for dementia and Alzheimer's, rather than the three years used by CCW, because of the limits of the Medicare claims data extracted for the analysis.

[e]This category excludes skin cancer.

CAD = coronary artery disease; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; ED = emergency department; FFS = fee-for-service; HQP = Health Quality Partners; PBPM = per beneficiary per month.

NA = not available.
n.a. = not applicable.

*Significantly different from zero at the .1 level, two-tailed test.

Table V.4. Program impacts on service use and expenditures during the extension (2010-2014) among high-risk beneficiaries who enrolled at any time during the program (2002–2014)

| | | Control group mean | Treatment group mean | Difference (90% confidence interval) | Percentage difference | *p*-value |
|---|---|---|---|---|---|---|
| **Program impacts on outcomes** | | | | | | |
| Annualized hospitalizations (#/person/year) | | 0.723 | 0.761 | 0.039 (-0.080, 0.157) | 5.4 | 0.59 |
| Annualized outpatient ED visits (#/person/year) | | 0.729 | 0.670 | -0.059 (-0.174, 0.056) | -8.1 | 0.40 |
| Medicare Part A and B expenditures ($ PBPM) | Without program fees | 1,694 | 1,724 | 29 (-218, 277) | 1.7 | 0.85 |
| | With program fees (mean fee paid = $260) | 1,694 | 1,959 | 264 (17, 512) | 15.6 | 0.08* |
| Medicare Part A expenditures ($ PBPM) | | 935 | 988 | 53 (-143, 250) | 5.7 | 0.66 |
| Medicare Part B expenditures ($ PBPM) | | 760 | 736 | -24 (-107, 59) | -3.1 | 0.64 |
| **Sample sizes, average length of follow-up, and receipt of treatment services** | | | | | | |
| Number of beneficiaries | | 362 | 374 | n.a. | n.a. | n.a. |
| Mean number of follow-up months | | 30.7 | 30.8 | n.a. | n.a. | n.a. |
| Mean percentage of follow-up months during which beneficiary received treatment[a] | | 0.0 | 85.3 | n.a. | n.a. | n.a. |

Sources: Medicare Enrollment Database, National Claims History File, Standard Analytic File, and the Mathematica randomization file.

Notes: The sample includes beneficiaries enrolled from April 1, 2002, to June 30, 2014. All beneficiaries met the "high-risk" criteria—that is, they had CAD, CHF, COPD, or diabetes and one or more hospitalizations in the year before enrollment.

The control group mean is weighted but not regression-adjusted. The treatment group mean equals the control group mean plus the regression-adjusted difference between treatment and control. The percentage difference is calculated by dividing the regression-adjusted difference between the treatment and control groups by the control group mean.

Treatment-control differences were adjusted for baseline characteristics to increase the precision of the estimates and to account for chance differences between the two groups. Negative estimates of treatment-control differences indicate that hospitalizations or expenditures were lower for the treatment group.

The regressions use observations that are weighted according to the proportion of the follow-up period during which each sample member met CMS's demonstration-wide requirements: being in fee-for-service, having both Part A and Part B coverage, having Medicare as the primary payer, and being alive for part of the month. Weights were calculated separately for the treatment and control groups.

[a]Calculated as follows: (1) for each beneficiary, calculate the percentage of follow-up months during which he or she received treatment services (a beneficiary was considered to have received services in a given month if HQP submitted a bill for services for that beneficiary in that month) and (2) find the average across treatment group beneficiaries.

CAD = coronary artery disease; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; ED = emergency department; HQP = Health Quality Partners; PBPM = per beneficiary per month.

n.a. = not applicable.

*Significantly different from zero at the .1 level, two-tailed test.

## Table V.5. Comparison of impact estimates before and after the extension for high-risk beneficiaries

| | | | Before the extension (2002–2010) | After the extension (2010–2014) | Difference[a] |
|---|---|---|---|---|---|
| Impact estimates (90 percent confidence interval) | Annualized number of hospitalizations (#/person/year) | | -0.293*** (-0.458, -0.129) | 0.065 (-0.086, 0.216) | 0.359*** (0.136, 0.581) |
| | Medicare Part A and B expenditures ($ PBPM) | Without program fees | -425** (-698, -152) | 16 (-315, 347) | 441* (13, 689) |
| | | With program fees[b] | -313* (-587, -40) | 276 (-55, 607) | 589** (161, 1,017) |
| Sample size (treatment and control) | | | 322 | 483 | 161 |
| Mean number of follow-up months | | | 49.4 | 25.5 | -23.9 |

Sources: Impact estimates for the period before the extension (2002–2010) are from the Fifth Report to Congress for the evaluation of the MCCD (Burwell 2014). Impact estimates for the period after the extension (2010–2014) are from Table V.2 of this report.

Notes: For the impact estimates before the extension, outcomes were measured from April 1 2002, to June 30, 2010, for beneficiaries who enrolled by June 30, 2009 and met the high-risk criteria at enrollment. Outcome and sample definitions for the post-extension sample are the same as those reported in Table V.2 of this report.

The impact estimates in each period (before and after the extension) represent the difference in the regression-adjusted outcomes for beneficiaries in the treatment and control groups. The *p*-values correspond to tests that the differences in outcomes for the two treatment and control groups are zero. The difference in impact estimates is the difference in the estimates in each period; the *p*-value is for the test that the difference in impact estimates for the two periods is zero.

[a]Difference between the sample size, follow-up months, or impact estimates after the extension versus before the extension.

[b]The mean fee that HQP received for each high-risk beneficiary in the treatment group, per month, was $112 before the extension and $260 after the extension.

MCCD = Medicare Coordinated Care Demonstration; PBPM = per beneficiary per month.

*/**/***Significantly different from zero at the .1/.05/.01 levels, respectively, two-tailed test.

# VI. QUESTION 3: LIKELY EXPLANATIONS FOR THE DECLINE IN IMPACTS

## A. Overview of analytic approach

We identified likely explanations for the observed decline in impacts on service use and Medicare expenditures in three steps. First, we developed hypotheses that—based on discussions with HQP administrators and nurse care managers and our own judgment—plausibly explain the decline in program impacts. We describe those hypotheses in the next section. Next, we tested those hypotheses to the extent possible by using descriptive data and new impact estimates, as applicable. Third, we interpreted the combined evidence, drawing conclusions about which, if any, hypothesized explanations seemed the most likely. Although we identified *likely* explanations and ruled out others, it was not possible to tell with certainty which changes were the ones that caused the decline in effects.

## B. Hypotheses for why program impacts declined

We have five hypotheses for why the estimates of program impacts on service use and expenditures declined after the 2010 extension:

- **Hypothesis 1: Tenure.** Program impacts may get larger the longer a beneficiary is enrolled in the program. Because the extension period lasted a little over four years (October 1, 2010, to December 31, 2014) and the pre-extension period lasted a little over eight years (April 1, 2002, to September 31, 2010), the smaller impacts during the extension may be due to enrollees having a shorter time enrolled in the program, on average.
- **Hypothesis 2: Improvements in usual care.** Since 2010, other health care organizations have begun providing care management services that may overlap with the services that HQP provides. For example, some hospitals in the region have enhanced their transitional care, in part to reduce financial penalties for high hospital readmission rates. Some primary care practices have begun adopting medical home principles, which include care management. If usual care for both the treatment and control group now includes care management, HQP's services may not add the same value in terms of being able to decrease service use and expenditures. Improvements in medications or other treatment regimens for patients with CHF, COPD, or CAD may also have improved outcomes for the control group.
- **Hypothesis 3: Change in population served.** After the extension, HQP began identifying prospective enrollees through hospital discharge records (instead of by reviewing patient charts at participating primary care practices) and modestly expanded into new service areas. According to both HQP and our claims analyses, the high-risk group after the extension has a higher prevalence of chronic conditions and is older than the pre-extension high-risk group. HQP staff also think these later enrollees have, on average, lower income and health literacy levels. These changes in the population could reduce impacts on service use and expenditures if any, or some combination, of the following is true:
    - The post-extension high-risk population required more frequent contacts than the earlier group did (for example, to detect and prevent acute exacerbations of illnesses that occur more frequently), but HQP's intervention remained the same.
    - New enrollees were less able or willing to make the substantial changes in self-care needed to reduce hospitalizations, regardless of HQP's intervention.

  - Lower socioeconomic status meant that nurses needed to first develop a foundation in health literacy and establish connections with community resources (for example, transportation to appointments) before the program could have impacts; therefore, delaying impacts.

- **Hypothesis 4: Changes in the intervention.** Several core elements of the intervention may have changed after the extension. First, facing a consistently high-risk caseload, HQP's nurses may not have been able to contact high-risk enrollees as frequently (overall or in person) after the extension as they had been able to do before the extension. (Before the extension, the care managers served patients with a wide range of risk levels and could triage their contacts for those at highest risk at any point in time). HQP lowered its target caseload from 108 to 75 to accommodate the high-risk caseloads, but this decrease may not have been enough. Second, because they served a more homebound population, HQP stopped offering several group classes, which may have been important for earlier impacts for at least some of the enrollees (for example, by reducing social isolation). Third, with the HQP staff growing as its Aetna business expanded and with nurse care managers needing to cover more for one another, it is possible that the continuity of care declined. That is, a patient may have seen different care managers more often, compromising the relationship between care managers and patients. Finally, because HQP began enrolling beneficiaries through relationships with hospitals rather than physician offices, it is possible that relationships between nurses and physicians deteriorated, resulting in either fewer contacts with physicians or physicians being less willing to modify their treatment regimens based on HQP's input.

- **Hypothesis 5: Disruption from possible shutdowns.** HQP administrators have told us that there were two periods during the extension when HQP was not able to deliver the intervention as fully as intended:

  - The first 14 months of the extension (October 2010 through November 2011)—when HQP needed to quickly hire and train nurse care managers to maintain the intervention (the extension was granted less than a month before the project was scheduled to end) and needed to re-establish relationships with patients, most of whom HQP had told the program was ending

  - The four months spanning the June 30, 2013 extension (May 2013 through August 2013), which also occurred within a month of the planned end date and created similar disturbances as described above

  During these two periods (totaling 18 months), HQP administrators think that they were not able to deliver the same intensity and quality of intervention, particularly in-person contacts with patients. They also noticed that some patients became frustrated by the start-and-stop nature of the program and became disengaged (for example, by disenrolling). It is possible that the declines in effects over the full extension period were driven by the disruptions caused during these periods.

We explored another hypothesis, but after interviews with HQP staff, determined it was not plausible. We initially thought that, because HQP identified patients through hospital discharge records and not through physician offices after the extension, physicians may have been less involved in selecting patients they thought would benefit most from the intervention. However,

HQP explained that before the extension physicians did not filter the lists of potential enrollees for those they thought would benefit the most.

In developing Hypothesis 4 (changes to the intervention), we considered whether any of the six features that prior MCCD studies identified as key to reducing hospitalizations changed after the extension. Brown et al. (2012) identified the following six features—all of which HQP had before the extension—as distinguishing the programs that successfully reduced hospitalizations for high-risk beneficiaries from those that did not: (1) frequent in-person meetings with patients to supplement phone calls, (2) occasional in-person contact between nurses and the patients' physicians, (3) nurses acting as a communications hub for providers, (4) delivering evidence-based patient education to patients, (5) providing strong medication management, and (6) providing timely and comprehensive transitional care after hospitalizations. Based on discussions with HQP, we determined that only one of these factors—frequent in-person contacts—may have declined after the extension, and this is built into the hypothesis. HQP's program retained all of the other features, and in some cases strengthened them, after the intervention. For example, HQP strengthened its transitional care program since, after the extension, all new hospital partners agreed to provide HQP with timely data on when their enrollees were hospitalized, allowing for timely intervention. Prior to the extension, only one of the participating hospitals was able to provide such data.

## C. Testing hypothesis 1: Tenure in the program

### 1. Methods

To test the hypothesis that impacts declined because the average length of enrollment (tenure) shortened after the extension, we re-estimated impacts for high-risk beneficiaries before and after the extension while controlling for tenure. Specifically, we estimated impacts in enrollees' first year of follow-up, second year of follow-up, and one to three years of follow-up, if those periods fell fully before or after the extension. In each period, the sample was limited to beneficiaries who enrolled early enough to have at least six months of potential exposure to the intervention and who were observable for at least one day during the follow-up period. The impact estimate was the regression-adjusted difference in outcomes (hospitalizations, outpatient ED visits, and Medicare expenditures) for the treatment and control beneficiaries, using the same regression and weighting methods described earlier.

If the hypothesis is correct, the patterns in impacts on hospitalizations, outpatient ED visits, and Medicare expenditures before and after the extension should be similar after controlling for year of enrollment.

### 2. Results

**Sample sizes and average follow-up.** The analysis of impacts in the first year of follow-up and years 1 through 3 of follow-up included 366 pre-extension enrollees and 483 post-extension enrollees (treatment and control) (Table VI.1). The sample sizes dropped modestly for the second-year analysis because fewer people enrolled early enough for their second year to start before the end of the relevant period. The mean number of follow-up months was similar in

Table VI.1. Comparison of program impacts among high-risk beneficiaries during the first three years of patient follow-up, if those years occurred before versus after the extension

| | | | Before the extension (April 2002–September 2010) | | | During the extension (October 2010–December 2014) | | |
|---|---|---|---|---|---|---|---|---|
| | | | First year | Second year | First through third[a] | First year | Second year | First through third[a] |
| **Program impacts on outcomes** | | | | | | | | |
| Treatment-control difference, adjusted (*p*-value) | Annualized hospitalizations (#/person/year) | | -0.119 (0.37) | -0.227 (0.11) | -0.206** (0.04) | 0.074 (0.54) | 0.077 (0.59) | 0.069 (0.46) |
| | Annualized outpatient ED visits (#/person/year) | | -0.210** (0.0495) | -0.312* (0.053) | -0.254*** (0.007) | 0.059 (0.60) | -0.027 (0.85) | 0.014 (0.88) |
| | Medicare Part A and B expenditures ($ PBPM) | Without program fees | -379 (0.14) | -245 (0.27) | -360** (0.048) | -196 (0.49) | 293 (0.36) | 19 (0.93) |
| | | With program fees | -263 (0.30) | -131 (0.56) | -244 (0.18) | 76 (0.79) | 549* (0.09) | 280 (0.17) |
| | Medicare Part A expenditures without program fees ($ PBPM) | | -282 (0.21) | -191 (0.29) | -272* (0.08) | -101 (0.68) | 208 (0.45) | 32 (0.84) |
| **Sample sizes, average length of follow-up, and receipt of treatment services** | | | | | | | | |
| Sample size (treatment and control) | | | 366 | 348 | 366 | 483 | 386 | 483 |
| Mean number of eligible follow-up months (treatment and control) | | | 11.7 | 10.2 | 29.2 | 11.4 | 10.3 | 24.5 |
| Mean percentage of follow-up months during which beneficiaries received treatment[b] | | | 94.7 | 91.7 | 93.3 | 96.7 | 91.3 | 93.3 |

Sources: Medicare Enrollment Database, National Claims History File, Standard Analytic File, and the Mathematica randomization file.

Notes: The research sample depends on the outcome period and follow-up year. Before the extension, the sample for the first year of follow-up and follow-up years 1 through 3 included beneficiaries enrolled from April 1, 2002, to March 31, 2010. This ensured that each beneficiary could potentially be observed for at least six months before measuring outcomes. The second year of follow-up included beneficiaries enrolled between April 1, 2002, and September 30, 2009. During the extension, the first year of follow-up and first through third years of follow-up included beneficiaries enrolled between October 1, 2010, and June 30, 2014. The second year of follow-up included beneficiaries enrolled between October 1, 2010, and December 31, 2013.

The research sample includes beneficiaries who met (1) the high-risk definition at randomization and (2) CMS's demonstration-wide requirements for at least one month during the follow-up period. To be "high risk," a beneficiary needed to have CAD, CHF, COPD, or diabetes and at least one hospitalization in the year before randomization. To meet CMS's eligibility criteria in a month, a beneficiary needed to (1) be alive and enrolled in Medicare Part A and B, (2) have Medicare as the primary payer of medical bills, and (3) not be enrolled in a comprehensive HMO.

Outcomes were measured during the patient follow-up year or years and were weighted according to the proportion of months in a year that a sample member met CMS's demonstration-wide requirements. Weights were calculated separately for the treatment and control groups. Treatment-control differences were adjusted for baseline characteristics to increase the precision of the estimates and to account for chance differences between the two groups. Negative estimates of treatment-control differences indicate that hospitalizations or expenditures were lower for the treatment group.

TABLE VI.1 (*continued*)

[a]This pooled three-year period helps us take advantage of greater power to detect impacts as compared to one-year analyses.

[b]Calculated as follows: (1) for each beneficiary, calculate the percentage of follow-up months during which he or she received treatment services (a beneficiary was considered to have received services in a given month if HQP submitted a bill for services for that beneficiary in that month) and (2) find the average across treatment group beneficiaries.

CAD = coronary artery disease; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; ED = emergency department; HQP = Health Quality Partners; PBPM = per beneficiary per month.

*/**/***Significantly different from zero at the .1/.05/.01 levels, respectively, two-tailed test.

the first and second years of follow-up for the pre- and post-extension samples, illustrating that the impact estimates effectively controlled for tenure. The mean number of follow-up months for the analysis for years 1 to 3 is a bit higher for the pre-extension sample than the post-extension sample (29.2 months versus 24.5 months).

**Impacts.** Before the extension, the program reduced hospitalizations by 0.21 per person per year ($p = 0.04$) and outpatient ED visits by 0.25 per person per year ($p = 0.01$), in the first three follow-up years (Table VI.1). It reduced Medicare Part A and B expenditures without program fees by $360 PBPM ($p = 0.048$). However, the program did not measurably change expenditures with program fees. This was likely due to low power to detect an effect—the point estimate indicated decrease and was large in magnitude ($244 PBPM, $p = 0.18$). After the extension, the program did not have a statistically significant effect on hospitalizations, outpatient ED visits, or Medicare expenditures, except for the second year of follow-up when the program increased expenditures with program fees by an estimated $549 PBPM ($p = 0.09$).

**Support for the hypothesis.** These results indicate that the differences in impacts before versus after the extension are not due to differential tenure in the program.

## D. Testing hypothesis 2: Improvements in usual care

### 1. Methods

If improvements in usual care drove the decline in program impacts, then—assuming the population that HQP served before and after the extension remained the same—we would expect the following pattern. First, before the extension, treatment group hospitalizations would be lower than the control group's outcomes. Second, after the extension, the treatment group's hospitalizations would remain the same but the control group's outcomes would improve (decrease), erasing program effects.

Because HQP's high-risk population did not remain the same before and after the extension, we used a regression to control for the changes in patient population. This regression included 849 high-risk beneficiaries, 366 who enrolled before the extension and 483 who enrolled after the extension. Outcomes (hospitalizations and outpatient ED visits) were measured through three years of enrollment to control for patient tenure. Specifically, for beneficiaries who enrolled before the extension, outcomes were measured from the month of enrollment through three years or the start of the extension, whichever came earlier. For beneficiaries who enrolled after the extension (October 1, 2010), outcomes were measured from the month of enrollment through the earlier of three years or the end of the program (December 31, 2014). Explanatory variables included beneficiary characteristics at enrollment (for example, chronic conditions and prior service use), treatment status, a binary variable for enrollment cohort (pre- or post-extension), and an interaction between treatment status and enrollment cohort.

If the hypothesis is true, we would expect that after controlling for patient characteristics, the treatment group's outcomes would remain about the same before and after the extension, but the outcomes for the control group would improve substantially, erasing program impacts.

## 2. Results

**Impact estimates.** As expected, the impact estimates for the pre-extension and post-extension periods in this analysis (see Table VI.2) are similar to those reported in Table V.2. The small differences resulted from differences in how the regressions controlled for beneficiary characteristics at baseline.[13]

**Changes in outcomes for the treatment and control groups**. The patterns over time depended on the outcome (Table VI.2):

- **Hospitalizations.** The annualized hospitalization rate for the post-extension treatment group was essentially the same as the hospitalization rate for the pre-extension treatment group, after controlling for differences in patient characteristics (Table VI.2). In contrast, the post-extension control group's hospitalization rate was 34 percent lower than the pre-extension control group's rate (0.544 hospitalizations versus 0.825 hospitalizations), a difference that was highly significant ($p = 0.009$). This pattern means that the large impact estimate before the extension disappeared during the extension because the control group's risk-adjusted outcomes improved. This result is robust to sensitivity tests that reduced the influence of outlier values by trimming the observations with rates above the 99th (or 98th) percentile to the 99th (or 98th) percentile.
- **Outpatient ED visits.** The treatment group's outpatient ED visit rate increased by 84 percent from the pre- to the post-extension period ($p < 0.001$), even after accounting for the increase in severity of illness. In contrast, the control group's ED visit rate essentially remained the same. Therefore, the large impact estimate before the extension disappeared because the treatment group's outcomes worsened, not because the control group's outcomes improved.

**Support for hypothesis.** The hospitalization results provide strong support for the hypothesis. Because improvements in hospitalization rates for the control group erased the program impacts after the extension, it is likely that improvements in usual care decreased the marginal value of HQP's services. These results cannot identify which aspects of usual care improved. As noted in the hypothesis, possibilities include increased provision of transitional care and the growth of medical homes, which include care management for high-risk beneficiaries. For example, representatives from both health systems that HQP began working with during the extension—St. Mary's and Crozer-Keystone—said that their health systems offer their own transitional care and care management programs. In fact, these services substitute for HQP's services. For both health systems, if a beneficiary from the system is randomly assigned to receive HQP services, the health system does not provide its own care management or transitional care services, relying instead on the services from HQP.

[13] Specifically, the model in this section assumes that the influence of the covariates on outcomes is the same for the pre- and post-extension samples. However, in the model used for Table V.2, the effects of covariates on outcomes is allowed to vary for the two samples because the regressions are run separately for the two samples.

## Table VI.2. Hospitalizations and outpatient ED visits for high-risk beneficiaries before and after the extension, by treatment status, controlling for changes in the patient population

| | | Before the extension | After the extension | Difference, adjusted | *p*-value for difference |
|---|---|---|---|---|---|
| Annualized hospitalizations (#/person/year) | Treatment | 0.632 | 0.614 | -0.018 | 0.86 |
| | Control | 0.825 | 0.544 | -0.281 | 0.009*** |
| | Difference, adjusted | -0.193 | 0.069 | 0.262 | 0.05* |
| | *p*-value for difference | 0.06* | 0.44 | 0.05* | n.a. |
| Annualized outpatient ED visit (#/person/year) | Treatment | 0.415 | 0.766 | 0.351 | < 0.001*** |
| | Control | 0.663 | 0.739 | 0.075 | 0.45 |
| | Difference, adjusted | -0.248 | 0.027 | 0.276 | 0.03** |
| | *p*-value for difference | 0.009*** | 0.74 | 0.03** | n.a. |
| Sample sizes | Treatment | 188 | 241 | n.a. | n.a. |
| | Control | 178 | 242 | n.a. | n.a. |
| | Combined | 366 | 483 | n.a. | n.a. |

Source: Medicare Enrollment Database, National Claims History File, Standard Analytic File, and the Mathematica randomization file.

Notes: The results in this table are based on two regressions, one for each outcome (hospitalizations and outpatient ED visits). Each regression includes 849 observations. The observations fall into two groups—outcomes for (1) high-risk beneficiaries who enrolled before the extension (from April 1, 2002, to March 31, 2010 [n = 366, treatment and control]) and (2) high-risk beneficiaries who enrolled after the extension (October 1, 2010, to June 30, 2014 [n = 483]). Outcomes are measured from the month of enrollment to the earlier of (1) three years after enrollment; (2) the end of the pre-extension period (September 30, 2010) if in the pre-extension sample or the post-extension period (December 31, 2014) if in the post-extension sample; or (3) death. The regression includes (1) a binary indicator for treatment status (1 = treatment, 0 = control); (2) whether a beneficiary is in the post-extension (post = 1) or pre-extension sample (post = 0); (3) an interaction between the treatment and post binary variable; and (4) a series of characteristics measured in claims at the time of program enrollment (for example, demographics, chronic conditions, and recent service use). The observations are weighted by the number of months that a person was eligible for the demonstration during his or her follow-up month. A person is eligible for the demonstration if he or she is enrolled in Medicare FFS, is not enrolled in a comprehensive HMO, and has Medicare as the primary payer for medical bills.

The difference between the treatment and control groups before the extension (and its *p*-value) equals the coefficient estimate (and its *p*-value) from the regression for the treatment binary variable. The difference between the control group before and after the extension equals the coefficient estimate for the post binary variable. The difference between the treatment and control group after the extension equals the coefficient estimate for the interaction of the treatment and post binary variables plus the coefficient for the treatment binary. Finally, the difference between the treatment group before and after the extension equals the coefficient estimate for the post variable plus the coefficient for the interaction between the treatment and the control groups.

The mean hospitalizations or ED visit rate for the control group before the extension is set to the weighted mean among the sample members in that group. The rates for three other cells in the table—(1) treatment group before the extension, (2) control group after the extension, and (3) treatment group after the extension—are calculated by adding the appropriate coefficient estimates for the treatment or post variables from the regression model to the mean rate for the control group before the extension.

ED = emergency department.

*/**/***Significantly different from zero at the .1/.05/.01 levels, respectively, two-tailed test.

n.a. = not applicable.

It is clear that HQP is unique in providing intensive, long-term nurse care management. Representatives from both St. Mary's hospital and Crozer-Keystone indicated that HQP is unique in its ability to follow patients across multiple settings, including into patients' homes. However, it may be that other programs offering more limited services are now providing the select interventions that had been critical to HQP's earlier success, which at the time only HQP was providing.

Despite the convincing results for hospitalizations, the patterns for outpatient ED visits do not fit the hypothesis. For ED visits, it appears that HQP's treatment services are not as effective as they had been earlier, not that usual care improved and reduced the added value of HQP's services to reduce ED visits.

## E. Testing hypothesis 3: Changes in the population

### 1. Methods

To test the hypothesis that change in the high-risk population contributed to the decline in impacts after the extension, we examined whether the post-extension impact estimates improve if we weight the beneficiaries in the post-extension population so that they resemble the pre-extension population.

We completed this analysis in three steps:

First, we developed weights for each observation in the post-extension sample. To develop the weights, we estimated a logistic regression to predict the likelihood of belonging to the pre-extension sample. This regression included 849 high-risk beneficiaries: 366 who enrolled between April 1, 2002, and March 31, 2010 (the pre-extension sample) and 483 who enrolled between October 1, 2010, and June 30, 2014 (the post-extension sample). Predictor variables included all available patient characteristics measured at baseline that might influence study outcomes and the likelihood of being in the pre-extension sample.[14] From the regression results, we generated a propensity score, $\hat{p}$, for each of the 849 high-risk beneficiaries, where $\hat{p}$ equals the predicted likelihood of being in the pre-extension sample. The distribution of propensity scores (see Appendix B) between the pre- and post-extension groups showed that (1) the predictor variables were able to distinguish between enrollees' likelihood of being in the pre-extension sample (as indicated by pre-extension enrollees tending to have relatively high propensity scores), but (2) there was substantial overlap in the propensity scores for the two groups. The second finding meant that there were beneficiaries in the post-extension sample who resembled pre-extension enrollees on observable characteristics. We then computed a weight for each post-extension enrollee, equal to $\frac{\hat{p}}{1-\hat{p}}$, which gives greater weight to post-extension enrollees who were similar to pre-extension enrollees (as indicated by $\hat{p}$ close to 1).

[14] This included all of the covariates we used in the main regression models (described earlier) and additional baseline characteristics: number of hospitalizations in the two years before enrollment and 12 separate variables for whether enrollees were hospitalized for each of 12 chronic conditions in the year before enrollment.

We then normalized the weights so they summed to the post-extension sample size. Additional details are in Appendix B.

Second, we confirmed that (1) the re-weighted post extension sample resembled the pre-extension sample (see Section VI.E.2.a), and (2) that the treatment and control groups remained balanced at baseline after the re-weighting (see Appendix B).

Third, we re-estimated post-extension impact regressions reported in Section VI.C (with outcomes measured for up to three years of patient follow-up) using the new weights. In these regressions, we also expanded the list of covariates in the regression so they matched the set used for generating the propensity scores (described above). Because the set of covariates changed, we also re-estimated impacts during the post-extension period without the new weights, but with the full list of covariates, to ensure a head-to-head comparison with the weighted pre-extension estimation, and also to ensure that the impact estimates were similar to those reported in Section VI.C.

If the change in the patient population was indeed one of the factors contributing to the decline in post-extension impacts, the impact results would show improvements in service use and Medicare expenditures in the post-extension period after weighting the post-extension sample.

## 2. Results

### a. Comparison of pre- and post-extension samples before and after weighting

Propensity score weighting was very successful in making the post-extension enrollees resemble the pre-extension enrollees on patient characteristics defined at baseline (Table VI.3). Before weighting, the post-extension sample was older, more likely to have chronic conditions, and more likely to use home health or SNF than the pre-extension sample. Many of these differences exceed 0.25 standardized differences,[15] a common threshold for identifying large differences (Institute of Education Sciences 2014). After weighting, standardized mean differences were 0.08 or smaller for all variables, indicating that the pre- and post-extension groups are similar after weighting. For example, before weighting, the largest of the differences was in the proportion of enrollees who have CKD (33.3 percent versus 14.5 percent); after weighting, there is essentially no difference between the groups—14.2 percent versus 14.5 percent for the post- and the pre-extension samples, respectively. The omnibus test of perfect similarity between pre- and post-extension enrollees on all variables cannot be rejected ($p > 0.99$), confirming that the two groups are similar.

We also confirmed that after the weighting, the post-extension treatment and control groups remained balanced at baseline for all measured characteristics (Appendix Table B.1), with balance becoming slightly better or worse for specific variables.

[15] The standardized difference between two groups for an outcome is the difference between the means of the outcome for the two groups divided by the standard deviation of the outcome across all observations in the two groups.

Table VI.3. Baseline characteristics for high-risk beneficiaries who enrolled before and after the extension, with and without weighting (percentages unless otherwise noted)

| | | | Without weighting[a] | | | With weighting[b] | | |
|---|---|---|---|---|---|---|---|---|
| | | Pre-extension enrollees (n = 366) | Post-extension enrollees (n = 483) | Difference | Standardized difference | Post-extension enrollees (n = 483) | Difference | Standardized difference |
| Age | < 65 | 0.0 | 0.0 | 0.0 | n.a. | 0.0 | 0.0 | n.a. |
| | 65–74 | 36.9 | 36.4 | 0.5 | 0.01 | 36.4 | 0.5 | 0.01 |
| | 75–84 | 48.6 | 38.1 | 10.5*** | 0.21 | 51.6 | -3.0 | -0.04 |
| | > or = 85 | 14.5 | 25.5 | -11.0*** | -0.27 | 12.0 | 2.5 | 0.08 |
| Male | | 51.1 | 44.1 | 7.0** | 0.14 | 49.6 | 1.5 | 0.02 |
| Race and ethnicity | Black, non-Hispanic | 1.4 | 2.5 | -1.1** | -0.08 | 0.8 | 0.6 | 0.07 |
| | Hispanic | 0.3 | 0.2 | 0.1 | 0.01 | 0.2 | 0.1 | 0.02 |
| Medicaid buy-in[c] | | 2.7 | 2.5 | 0.3 | 0.16 | 3.9 | -1.1 | -0.04 |
| Original reason for Medicare entitlement | Disability | 5.7 | 9.5 | -3.8** | -0.14 | 4.6 | 1.2 | 0.06 |
| | ESRD | 0.0 | 0.2 | -0.2 | -0.06 | 0.0 | 0.0 | -0.06 |
| Diagnosis[d] | CAD | 82.8 | 76.6 | 6.2** | 0.15 | 82.8 | 0.0 | 0.00 |
| | CHF | 38.0 | 48.5 | -10.5*** | -0.21 | 37.3 | 0.7 | 0.01 |
| | Diabetes | 42.6 | 46.2 | -3.6 | -0.07 | 42.0 | 0.6 | 0.01 |
| | COPD | 26.0 | 42.0 | -16.1*** | -0.34 | 24.7 | 1.3 | 0.03 |
| | Cancer[e] | 13.1 | 15.5 | -2.4 | -0.07 | 14.8 | -1.7 | -0.04 |
| | Stroke | 12.0 | 12.8 | -0.8 | -0.02 | 13.5 | -1.5 | -0.03 |
| | Depression | 14.5 | 25.1 | -10.6*** | -0.26 | 14.0 | 0.5 | 0.01 |
| | Dementia and Alzheimer's | 4.9 | 8.9 | -4.0** | -0.15 | 4.5 | 0.5 | 0.02 |
| | Osteoporosis | 18.0 | 20.3 | -2.3 | -0.06 | 21.4 | -3.4 | -0.06 |
| | Rheumatoid arthritis | 33.3 | 41.2 | -7.9** | -0.16 | 31.7 | 1.6 | 0.03 |
| | CKD | 14.5 | 33.3 | -18.9*** | -0.44 | 14.2 | 0.3 | 0.01 |
| | Atrial fibrillation | 33.3 | 31.7 | 1.7 | 0.04 | 33.6 | -0.2 | 0.00 |
| Hospitalized for the following conditions in the year before enrollment | CAD | 18.9 | 11.2 | 7.7*** | 0.22 | 20.0 | -1.1 | -0.02 |
| | CHF | 7.9 | 8.5 | -0.6 | -0.02 | 6.7 | 1.2 | 0.04 |
| | Diabetes | 0.6 | 2.1 | -1.5* | -0.13 | 0.5 | 0.1 | 0.02 |
| | COPD | 5.2 | 7.2 | -2.1 | -0.08 | 4.7 | 0.5 | 0.02 |
| | Cancer[e] | 0.0 | 0.2 | -0.2 | -0.06 | 0.0 | 0.0 | -0.06 |
| | Depression | 0.3 | 0.6 | -0.4 | -0.05 | 0.2 | 0.1 | 0.03 |
| | Dementia and Alzheimer's | 0.0 | 0.0 | 0.0 | n.a. | 0.0 | 0.0 | n.a. |
| | Rheumatoid arthritis | 6.0 | 6.2 | -0.2 | -0.01 | 4.8 | 1.2 | 0.05 |
| | CKD | 1.4 | 3.1 | -1.7* | -0.11 | 1.3 | 0.1 | 0.01 |
| | Atrial fibrillation | 7.7 | 6.2 | 1.4 | 0.06 | 7.9 | -0.3 | -0.01 |

TABLE VI.3 (*continued*)

| | | | Without weighting[a] | | | With weighting[b] | | |
|---|---|---|---|---|---|---|---|---|
| | | Pre-extension enrollees (n = 366) | Post-extension enrollees (n = 483) | Difference | Standardized difference | Post-extension enrollees (n = 483) | Difference | Standardized difference |
| | Hip fracture | 1.9 | 2.1 | 0.2 | -0.01 | 1.3 | 0.6 | 0.05 |
| In the year before enrollment | Annualized hospitalizations (number) | 1.45 | 1.67 | -0.22*** | -0.21 | 1.44 | 0.01 | 0.00 |
| | Any use of home health | 35.0 | 52.0 | -17.0*** | -0.35 | 32.2 | 2.8 | 0.05 |
| | Any use of a SNF | 7.9 | 21.1 | -13.2*** | -0.37 | 7.6 | 0.4 | 0.02 |
| Two years before enrollment | Annualized hospitalizations (number) | 1.01 | 1.21 | -0.20*** | -0.24 | 1.00 | 0.01 | 0.01 |
| | Number of conditions treated per person[f] | 3.33 | 4.02 | -0.69*** | -0.44 | 3.34 | -0.01 | -0.01 |
| Characteristics of enrollees' zip code of residence, mean | Median household income ($) | 81,776 | 79,432 | 2,343* | 0.12 | 83,197 | -1,421 | -0.02 |
| | College degree or more | 40.8 | 38.3 | 2.4*** | 0.19 | 41.6 | -0.8 | -0.02 |
| | Unemployment rate | 7.1 | 7.6 | 0.5*** | 0.22 | 7.0 | 0.1 | 0.03 |
| **Omnibus test for balance on baseline characteristics[g]** | | | | | | | | |
| *p*-value | | | n.a. | n.a. | 0.00 | n.a. | n.a. | > 0.99 |

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

Notes: The post-extension sample includes beneficiaries enrolled from October 1, 2010, to June 30, 2014. The pre-extension sample includes beneficiaries enrolled from April 1, 2002, to March 31, 2010. All beneficiaries met the "high-risk" criteria—that is, they had CAD, CHF, COPD, or diabetes and one or more hospitalizations in the year before enrollment. This analysis uses additional covariates not used in other impact regressions—for example, hospitalization for each of several conditions in the year before enrollment and sociodemographic makeup of enrollees' residence zip codes.

[a]Observations for pre-extension enrollees were not weighted.

[b]Observations were weighted using a new weight that equals the product of (1) the propensity score weight that ensures similarity of the post-extension sample to the pre-extension sample, and (2) the weight used for the original impact regression, which is proportional to the number of follow-up months for the each enrollee.

[c]Medicaid buy-in indicates that the beneficiary is eligible for both Medicare and Medicaid. The Medicare FFS average was approximated using the percentage of Medicare beneficiaries who were dual eligibles in 2010. See http://kff.org/medicaid/state-indicator/duals-as-a-of-medicare-beneficiaries.

[d]Diagnoses were based on the 2010 version of CCW definitions. The definitions use a look-back period of one year before enrollment for atrial fibrillation, cancer, COPD, depression, osteoporosis, and stroke and two years for CAD, CHF, CKD, diabetes, and rheumatoid arthritis. The evaluation used a two-year look-back period for dementia and Alzheimer's, rather than the three years used by CCW, because of the limits of the Medicare claims data extracted for the analysis.

[e]This category excludes skin cancer.

[f]Number of conditions for which one is treated is counted over a one- or two-year period, depending on the look-back period.

[g]Results from an overall chi-squared test indicate the likelihood of observing differences in the explanatory variables as large as the differences we observed if, in fact, the pre-extension and post-extension populations were perfectly balanced. The value of $p = 0.00$ for the omnibus test suggests that the two groups are not well-balanced, because we reject the null hypothesis that their characteristics are identical.

TABLE VI.3 (*continued*)

CAD = coronary artery disease; CCW = CMS Chronic Conditions Data Warehouse; CHF = congestive heart failure; CKD = chronic kidney disease; COPD = chronic obstructive pulmonary disease; ESRD = end-stage renal disease; FFS = fee-for-service; SNF = skilled nursing facility.

n.a. = not applicable.

*/**/***Significantly different from zero at the .1/.05/.01 levels, respectively, two-tailed test.

### b. Impact estimates before and after weighting

The impact estimates using the original weights (Table VI.4) but with the more exhaustive list of covariates were very similar to the estimate with the original weights that also used the original list of covariates (Table VI.1). This finding indicates that post-extension impact estimates are robust to inclusion of additional covariates.

With the new weights, the program still did not measurably affect hospitalizations or outpatient ED visits in the first three years of patient follow-up of the post-extension period. The point estimates remained essentially the same for outpatient ED visits and became slightly larger (less favorable) for hospitalizations and Medicare Part A and B expenditures, although neither was statistically significant. Without weighting, total Medicare expenditures with program fees were $315 higher for the treatment than the control group. This difference was not statistically significant at conventional levels ($p = 0.13$), but it became slightly larger ($380) and statistically significant ($p = 0.02$) after weighting the post-extension enrollees so they resemble the pre-extension enrollees.

The mean hospitalizations rates and Medicare expenditures (for treatment and control) were lower after weighting than before weighting, as was to be expected, given that the weighting process gives greater weight to those who are younger and healthier and who, therefore, more closely resemble the pre-extension high-risk group.[16]

### c. Support for hypothesis

We hypothesized that the post-extension finding of no impacts would change toward measurable improvements in outcomes if changes in population were one of the factors influencing the change in impacts. Given the findings that again show no measurable improvement in outcomes and even a statistically significant increase in Medicare expenditures with program fees, there is no evidence that changes in population reduced post-extension impacts. Note, however, that the changes in population may have *indirectly* influenced the decline in impacts by prompting HQP to modify the intervention (for example, to not offer group classes).

A limitation of propensity score weighting is that we cannot control for patient characteristics that are not available in the data. For example, we learned from HQP that the post-extension high-risk group differs from the pre-extension high-risk sample with regard to

[16] We also assessed whether the pre-extension impacts would differ if we estimated them after re-weighting the pre-extension sample to resemble the post-extension enrollees who were older and had more complex health conditions than pre-extension enrollees. Previous analyses have found that the pre-extension program improved hospitalization rates and Medicare expenditures. After weighting, the program was associated with smaller improvements in hospitalizations and Medicare expenditures without fees (and these differences were no longer statistically significant), but with larger improvements for ED visits. These results suggest the pre-extension program might have been less effective if it had been delivered to an older, frailer, and more medically complex population. However, because the intervention and usual care changed after the extension, these pre-extension impact results do not demonstrate that the converse—serving a healthier population during the extension—would have improved impacts during the extension.

Table VI.4. Program impacts in the first three years of patient follow-up during the extension period, before and after weighting the post-extension sample to resemble pre-extension enrollees at baseline

| | | After the extension without propensity weights[a] | | | After the extension with propensity weights[b] | | |
|---|---|---|---|---|---|---|---|
| | | Treatment group mean | Control group mean | Difference (adjusted) (*p*-value) | Treatment group mean | Control group mean | Difference (adjusted) (*p*-value) |
| Annualized hospitalizations (#/person/year) | | 0.813 | 0.749 | 0.064 (0.50) | 0.630 | 0.524 | 0.106 (0.18) |
| Annualized outpatient ED visits (#/person/year) | | 0.735 | 0.751 | -0.016 (0.86) | 0.638 | 0.655 | -0.017 (0.83) |
| Medicare Part A and B expenditures ($ PBPM) | Without program fees | 1,812 | 1,758 | 54 (0.80) | 1,401 | 1,284 | 118 (0.47) |
| | With program fees | 2,073 | 1,758 | 315 (0.13) | 1,664 | 1,284 | 380** (0.02) |
| **Sample sizes and average length of follow-up** | | | | | | | |
| Sample size | | 241 | 242 | 483 | 241 | 242 | 483 |
| Mean number of eligible follow-up months (treatment and control) | | 24.5 | 24.6 | 24.5 | 24.5 | 24.6 | 24.5 |

Sources: Medicare Enrollment Database, National Claims History File, Standard Analytic File, Mathematica randomization file, and the American Community Survey.

Notes: The sample includes beneficiaries enrolled from October 1, 2010, to June 30, 2014. All beneficiaries met the "high-risk" criteria—that is, they had CAD, CHF, COPD, or diabetes and one or more hospitalizations in the year before enrollment.

The control group mean is weighted but not regression-adjusted. The treatment group mean equals the control group mean plus the regression-adjusted difference between treatment and control. The percentage difference is calculated by dividing the regression-adjusted difference between the treatment and control groups by the control group mean.

Treatment-control differences were adjusted for baseline characteristics to increase the precision of the estimates and to account for chance differences between the two groups. This analysis uses additional covariates not used in other impact regressions—for example, hospitalization for each of several conditions in the year before enrollment and sociodemographic makeup of enrollees' residence zip codes. Negative estimates of treatment-control differences indicate that hospitalizations or expenditures were lower for the treatment group.

[a]Observations were weighted according to the proportion of the follow-up period during which each sample member met CMS's demonstration-wide requirements: being in fee-for-service, having both Part A and Part B coverage, having Medicare as the primary payer, and being alive for part of the month. Weights were calculated separately for the treatment and control groups.

[b]Observations were weighted using a new weight that equals the product of (1) the propensity score weight, set to equal $\frac{\hat{p}}{1-\hat{p}}$, where $\hat{p}$ is the propensity score weight that ensures similarity of the post-extension sample to the pre-extension sample, and (2) the weight used for the original impact regression, which is proportional to the number of months during which each sample member met CMS's demonstration-wide requirements. Use of this weight ensures that the post-extension enrollees resemble pre-extension enrollees on characteristics used in the propensity score weighting model. See Appendix B, Table B.1 for details on how the two populations compare after weighting.

CAD = coronary artery disease; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; ED = emergency department; PBPM = per beneficiary per month.

**Significantly different from zero at the .05 level, two-tailed test.

functional limitations. Even though we cannot capture functional limitations directly in claims, we are measuring a rich set of health conditions and prior service use (such as SNF and home health) that are likely to be correlated with many functional limitations. Still, there may be other unmeasured characteristics that we did not capture, which remains a limitation of this analysis. These characteristics could include the following, all of which HQP's nurse care managers indicated changed after the extension—health literacy, availability of a caregiver at home, or alcohol or substance abuse.

## F. Testing hypothesis 4: Changes in the intervention

### 1. Methods

To test the hypothesis that impacts declined because the intervention changed, we used HQP's detailed nurse contact data and regressions to assess whether the hypothesized changes in the intervention occurred, after controlling for changes in the high-risk patient population after the extension. These changes are a necessary but not sufficient condition for the hypothesis to be correct; the intervention might have changed in the hypothesized direction, but those changes might not have caused the decline in impacts.

We used a linear regression model that measured changes in the frequency and type of contacts, controlling for the increase in health risk found among the post-extension treatment group population (as compared to the pre-extension high-risk population). We used HQP's contact data, which contains observations for each contact a nurse care manager had with or on behalf of a patient throughout the duration of the program (from 2002 to 2014). Similar to the process we used for the implementation analysis in Section IV.B, we limited the sample in this analysis to all high-risk treatment group beneficiaries who enrolled before or after the extension and early enough to be followed up for a full year within the pre- or post-extension period (215 and 223 beneficiaries, respectively). We selected four types of contacts as outcome variables (calculated as the mean number of successful contacts per beneficiary per month), which represent core elements of HQP's program: all individual contacts, individual contacts in person, individual contacts with a physician, all group contacts, and all individual and group contacts. We also used as an outcome variable the percentage of total contacts provided by a beneficiary's primary nurse care manager to approximate the continuity of care. Each outcome variable was measured during the first, second, and third years of a beneficiary's enrollment. The covariates, all measured at baseline, included (1) demographics (age, gender, and race), (2) original reason for Medicare entitlement (old age, disability, or end-stage renal disease), (3) dual eligibility status (approximated by Medicaid buy-in), (4) indicators of health risk (instances of 12 chronic diseases), and (5) recent service use (hospital admissions, Medicare expenditures, use of home health, and SNF). We also included indicators for program tenure (specifically, whether the contact occurred during a participant's first, second, or third year in the program) and whether the observation was in the pre- or post-extension period.

### 2. Results

#### a. Program contacts

HQP provided a *more* intensive intervention to high-risk beneficiaries who enrolled after the extension than to high-risk beneficiaries who enrolled before the extension, after controlling for patient characteristics (Table VI.5). On average, HQP staff provided:

Table VI.5. Program contacts with patients in the first year of enrollment before and after the extension, with and without controlling for changes in patient population

| Contact type[a] | Mean contact rate (#/beneficiary/month) | | | | Mean contact rate (#/beneficiary/month) | | | |
|---|---|---|---|---|---|---|---|---|
| | Before the extension | After the extension, unadjusted | Unadjusted difference (%) | *p*-value for unadjusted difference | Before the extension | After the extension, adjusted | Adjusted difference (%) | *p*-value for adjusted difference |
| Individual, any | 2.13 | 3.22 | 1.10 (51.5) | < 0.001*** | 2.13 | 2.77 | 0.64 (30.2) | 0.004*** |
| Individual, in-person | 0.68 | 1.60 | 0.92 (135.1) | < 0.001*** | 0.68 | 1.41 | 0.73 (107.7) | < 0.001*** |
| Individual, with provider | 0.20 | 0.49 | 0.29 (143.3) | < 0.001*** | 0.20 | 0.35 | 0.15 (73.1) | 0.048** |
| Group[b] | 0.44 | 0.08 | -0.36 (-81.3) | < 0.001*** | 0.44 | 0.10 | -0.35 (-78.5) | < 0.001*** |
| Individual (any) or group | 2.57 | 3.30 | 0.73 (28.5) | 0.001*** | 2.57 | 2.87 | 0.29 (11.4) | 0.190 |
| Individual, with primary nurse care manager (%) | 90.49 | 89.34 | -0.01 (-1.26) | 0.282 | 90.49 | 90.04 | -0.44 (-0.5) | 0.704 |

Sources: HQP data on program contacts and Mathematica analysis of Medicare enrollment databases and claims.

[a]A contact that the nurse care manager made with or on behalf of a beneficiary.

[b]A beneficiary's participation in a group class.

**/***Significantly different from zero at the .05/.01 levels, respectively, two-tailed test.

- 30 percent more individual contacts, either in person or by phone ($p < 0.01$)
- 108 percent more in-person contacts ($p < 0.01$)
- 73 percent more contacts with patients' providers ($p < 0.05$)

The number of group contacts decreased by 79 percent ($p < 0.01$). In terms of total contacts (group or individual), the decrease in group classes largely offset the increase in individual contacts. Therefore, the total contact rate was only 11 percent higher after the extension, and this difference was not statistically significant. However, the shift away from group classes toward individual contacts, particularly in-person contacts, signals an increase in the intensity of the intervention.

Nurse continuity (as measured by the percentage of all contacts in the first year of enrollment that are with the primary nurse) was very high, at about 90 percent, in both the pre- and post-extension periods.

### b. Support for hypothesis

The results support only one component of the hypothesis regarding changes in implementation: the large, statistically significant decrease in group contacts might help explain why impacts have declined among high-risk beneficiaries in the post-extension period. If group contacts conferred benefits to program participants beyond that of a one-on-one visit (a type of contact which increased in the post-extension period), the large decline in group contacts could have affected impacts.

The results are not consistent with the other parts of the hypothesis that could link program implementation to the lack of impacts for high-risk beneficiaries in the post-extension period:

- The large, statistically significant increase in individual contacts with (or on behalf of) high-risk enrollees indicates that, on average, high caseloads did not impede the provision of services in the post-extension period. Nurse care managers were able to provide considerably more services in person with (and on behalf of) high-risk beneficiaries after the extension.
- The large, statistically significant increase in contacts with providers indicates that, on average, nurse care managers had more frequent interaction with physicians in the post-extension period. Thus, there is no evidence that relationships between nurse care managers and physicians were weaker or affected impact estimates for high-risk beneficiaries in the post-extension period. It is possible that—even if the rates of contacts with providers increased—the quality of those contacts deteriorated, for example, providers may have been less willing to provide nurses with medical information or to act on medical information supplied by the nurses. However, two long-serving HQP nurse care managers indicated that, overall, relationships with providers remained as strong, or improved, after the extension (although providers in some newly participating practices were not as receptive to HQP's nurses).
- The high proportion of total contacts provided by a beneficiary's primary nurse care manager that persisted across time periods indicates that, on average, nurse care managers were no less consistent in providing care with (and on behalf of) high-risk beneficiaries in

the post-extension period. Because the continuity of nurse care managers remained stable, there is no evidence that relationships between nurse care managers and patients were weaker or affected impact estimates for high-risk beneficiaries in the post-extension period.

## G. Testing hypothesis 5: Disruptions from near shutdowns

### 1. Methods

To test the hypothesis that impacts declined because the disruptive periods impeded service delivery, we conducted two complementary analyses.

#### a. Analysis 1: Implementation

First, we examined whether the hypothesized declines in nurse contacts occurred during the 18 months HQP identified as particularly disruptive (see Section VI.B for the dates). Specifically, for each of the 223 treatment group beneficiaries who enrolled after the extension but early enough to be followed up for at least a year, we calculated his or her monthly rate of nurse contact (any contact and in-person contacts) during disruptive and non-disruptive periods, in each of up to three years of follow-up. For each outcome, a single beneficiary could have just one observation (for one year of follow-up, all of which fell in a disruptive or non-disruptive periods) and as many as six observations (for three years of follow-up, each with a disruptive and non-disruptive period). Then, using a regression (n = 769) that controlled for year of follow-up and baseline covariates, we estimated whether the mean nurse contact rate differed during disruptive versus non-disruptive months. The regressions adjusted the standard errors to account for clustering of outcomes for beneficiaries contributing more than one observation in the regression.

If the hypothesis is correct, we would expect to see a decline in nurse contacts during disruptive periods. This is a necessary but not sufficient condition for the hypothesis to be true because even if there were a decline in nurse contacts the contact declines might not necessarily have been the reason the impacts declined.

#### b. Analysis 2: Program impacts

This analysis estimated program impacts separately for the disruptive and non-disruptive periods of the extension. As in earlier analyses, the sample included the beneficiaries (treatment and control) who enrolled between October 2010 and June 2014. For each beneficiary, we calculated his or her hospitalization rates and monthly Medicare expenditures during the extension period separately for the disruptive versus non-disruptive months. A beneficiary whose full follow-up period occurred during a disruptive or a non-disruptive period had only one observation per outcome—that is, his annualized hospitalization rate over his full follow-up period. A beneficiary whose follow-up period spanned disruptive and non-disruptive periods had two observations. The regression for estimating impacts included 830 observations (each a unique person-period), baseline covariates, an indicator for treatment status, an indicator for period (disruptive or non-disruptive), and an interaction between the indicators for treatment status and period. To account for the correlation in outcomes within individuals, we used generalized estimating equations (with a linear link function) and an unstructured correlation matrix for repeated observations. This regression yielded separate impact estimates during the

disruptive and non-disruptive periods and a formal test for whether the difference in impact estimates during those two periods is statistically significant.

If the hypothesis is correct, we would expect to see (1) favorable impacts during the non-disruptive periods and (2) no impacts during the non-disruptive periods.

### 2. Results

#### a. Analysis 1: Implementation

The number of nurse contacts (any or in-person contact) with, or on behalf of patients, was an estimated 9 to 10 percent lower in the disruptive versus non-disruptive periods, after controlling for patient characteristics (Table VI.6). This difference was statistically significant ($p < 0.01$) for the in-person contact rate but not for any contacts ($p = 0.16$). The variation in contact rates is smaller for in-person contacts than any contacts, allowing for greater statistical precision for the in-person estimates.

#### b. Analysis 2: Impacts

After removing the non-disruptive periods, the impact estimates continue to show no measurable program impacts on hospitalizations, outpatient ED visits, or Medicare Part A and B expenditures (Table VI.6). Fewer people had any of their follow-up period during a disruptive period (n = 347, treatment and control) than a non-disruptive period (n = 483). And for those whose follow-up period did include disruptive months, the average number of disruptive months was only 5.4. Therefore, it would be difficult for the impact estimates during the non-disruptive period to change after removing the disruptive periods. The reason that average number of disruptive months is only 5.4—while the total number of disrupted calendar months is much higher (18 months)—is that the sample only includes beneficiaries who enrolled after the 2010 extension. New enrollment after the extension did not start until April 2011, which means that most of the months that HQP considered disruptive following the 2010 extension (October 2010 through November 2011) do not overlap with the outcome periods for the sample members.

The point estimates for hospitalizations and expenditures during the disruptive periods are large and not favorable, but these estimates are not statistically different from zero.

#### c. Support for hypothesis

The results do not provide any support for the hypothesis. The contact rate was modestly lower (about 10 percent) during the disruptive versus non-disruptive periods, indicating that these periods did affect service delivery. However, the impact estimates did not improve after removing the disruptive periods.

## H. Summary of most likely explanations for decline in effects

Table VI.7 summarizes the level of evidence supporting each of the five hypotheses for why the impacts declined for the high-risk beneficiaries after the extension.

Table VI.6. Program implementation and impacts during disruptive and non-disruptive periods of the extension

| | | Non-disruptive periods | Disruptive periods | Difference, adjusted (*p*-value) | Percentage change |
|---|---|---|---|---|---|
| **Program implementation: Nurse contacts with patients in the treatment group** | | | | | |
| Sample size (person-years, treatment only) | | 449 | 320 | n.a. | n.a. |
| Individual contact, any (#/person/month) | | 2.98 | 2.71 | -0.27 (0.16) | -9.2 |
| Individual, in-person (#/person/month) | | 1.48 | 1.33 | -0.15 (0.006***) | -10.0 |
| **Program impacts during the extension** | | | | | |
| Sample size (person-period, treatment and control) | | 483 | 347 | n.a. | n.a. |
| Mean number of follow-up months | | 21.7 | 5.4 | n.a. | n.a. |
| Regression-adjusted difference in outcomes between treatment and control groups (*p*-value) | Annualized number of hospitalizations (#/person/year) | 0.020 (0.83) | 0.246 (0.14) | 0.227 (0.17) | n.a.[a] |
| | Annualized number of outpatient ED visits (#/person/year) | 0.004 (0.97) | 0.044 (0.79) | 0.040 (0.81) | n.a.[a] |
| | Medicare Part A and B expenditures ($ PBPM) | -46 (0.81) | 330 (0.19) | 376 (0.17) | n.a.[a] |

Sources: HQP data on program contacts and Mathematica analysis of Medicare enrollment databases and claims.

Notes: For the implementation analysis (top rows), the unit of observation is the person-year-period. For each treatment group beneficiary who enrolled between October 1, 2010 and December 31, 2013, we calculated a nurse contact rate for each follow-up year (up to three years), separately for the portion of the year that fell during the disruptive versus the non-disruptive months. The mean rate for the non-disruptive periods is the average monthly contact rate during the disruptive part of the year for all person-years where at least part of the year fell within the disruptive periods (n = 449). The mean rate for the disruptive periods equals the mean rate for the non-disruptive periods plus the regression-adjusted difference in rates for the two periods. The regressions controlled for the follow-up year and beneficiary characteristics at enrollment.

For the impact analysis (bottom rows), the unit of observation is the person-period. For each person in the treatment and control group, we calculated his or her outcome (hospitalization rate, ED visit rate, or Medicare expenditures) during the disruptive and non-disruptive months. The impact estimate in each period (disruptive or non-disruptive) is the regression-adjusted difference in outcomes for the treatment group's observations during the period and the comparison group's observations during the period. The difference in impact estimates is the impact estimate during the disruptive periods minus the estimate in the non-disruptive periods; the *p*-value is for the test that the difference is zero.

[a]The impact estimate for the non-disruptive periods is so close to zero that the percentage changes are extremely large and not readily interpretable.

ED = emergency department; HQP = Health Quality Partners; PBPM = per beneficiary per month.

n.a. = not applicable.

***Significantly different from zero at the .01 level, two-tailed test.

Table VI.7. Summary of the empirical support for the five hypotheses for why impacts declined after the extension

| Hypothesis for why impacts declined after the extension | | Level of support for hypothesis | Evidence |
|---|---|---|---|
| 1 Shorter tenure in program after extension. | | None | Program impacts appeared within first 3 years of enrollment before extension; no impacts over same period after the extension. |
| 2. Usual care improved, decreasing the added value of HQP services for reducing hospitalizations and expenditures. | | Strong | Program impacts on hospitalizations after extension disappeared because, after controlling for patient characteristics, outcomes for control group improved, not because outcomes worsened for the treatment group. |
| 3. Population is too ill or complex to benefit from intervention. | | None | Re-weighting post-extension group to resemble pre-extension group at baseline did not improve impact estimates. |
| 4. Intervention changed. | a. Became less intense | None | Number of nurse contacts (total and in person) *increased* after extension, controlling for patient characteristics; it did not decrease as hypothesized. |
| | b. Had lower nurse continuity | None | Nurse continuity, as measured by the share of total contacts with the primary nurse, remained high at 90% before and after the extension. |
| | c. Had fewer provider contacts | None | Number of nurse contacts with providers increased after the extension, controlling for patient characteristics; they did not decrease, as hypothesized. |
| | d. Fewer group classes | Some | Participation in group classes, even after controlling for changes in patient population, was substantially lower after the extension. This is necessary to suggest, but does not prove that, decreases in group classes contributed to the decline in impacts. |
| 5. Service disruptions from near shutdowns made intervention less effective. | | None | Program impacts do not improve after limiting outcomes to the non-disruptive periods. |

Source: Tests of individual hypotheses, as described earlier.

Based on these results, the most likely explanation for the decline in program impacts after the extension is that usual care for the control group improved in ways that decreased the ability of HQP's services to further reduce hospitalizations. This can also explain the decline in impacts on expenditures, because HQP's prior impact on expenditures was driven by its impact on hospitalizations (83 percent of the total decline in Part A and B expenditures before the extension was caused by declines in inpatient and post-acute care expenditures). However, improvements in usual care cannot explain the decline in impacts on outpatient ED visits. The decline in group classes may have contributed to the decline in impacts. Even after controlling for patient characteristics, high-risk beneficiaries after the extension were much less likely to participate in group classes than before the extension. However, the measured decline in group classes (a necessary condition for the hypothesis to be true) does not prove that this decline contributed to the decline in impacts.

This analysis has several limitations, although we can identify likely explanations for the decline in program impacts after the extension, we cannot with certainty or known probability identify which factor(s) actually accounted for the decline in impact estimates. One important

limitation is that, when we controlled for patient characteristics (as we did in several regressions), we could do so based only on characteristics measurable in claims. Although claims data can provide a range of characteristics (including chronic conditions and service use), they do not provide detailed information on functional limitations or caregiver status. Including a richer set of covariates could influence the findings.

Finally, although the analysis for the second research question (see Section V.B) suggests that the differences in impacts for the pre- and post-extension samples are not due to chance, it is important to note that such a possibility exists. This could occur in two ways. First, the original findings for the pre-extension period—based on an analysis of a high-risk subgroup of the full patient population—might have been a "false positive," that is, favorable differences between the treatment and control groups arose due to chance differences between the groups not true program effect. Second, the result in the current study may be a "false negative," indicating that the program does not have an effect when, in fact, it does but the impacts are obscured by statistical noise or chance differences between the treatment and control groups (that led to lower service use and expenditures for the control group). Neither of these explanations is likely, given the strength of the results, but either is possible.

This page has been left blank for double-sided copying.

# VII. DISCUSSION

In this section, we discuss how the evaluation results can help CMS determine (1) how confidently to disseminate HQP's model for possible use in current care management initiatives that aim to improve care quality and lower Medicare expenditures and (2) the areas that organizations that wish to replicate HQP's early success should focus on to maximize their chances of doing so.

## A. Degree of confidence in HQP's model to improve quality and reduce Medicare expenditures

The results in this report substantially reduce the confidence that HQP's model of care management can improve quality and reduce Medicare expenditures in today's environment, at least in the geographic area where it was operating. Before the extension, for the high-risk group, HQP's model met the three-part aim—common to many current CMS initiatives—of improving quality of care, improving the health of populations, and reducing expenditures. However, the results for the most recent period (October 2010 to December 2014) indicate that the model did not reduce hospitalizations or outpatient ED visits relative to what they would have been had enrollees not been in HQP, and so did not reduce Medicare expenditures relative to that standard. After accounting for substantial monthly fees (which averaged $260 per beneficiary per month), the program measurably increased expenditures to Medicare by an estimated 16 percent over what they would have been in the absence of HQP.

The high-risk population became older and substantially more complex after the extension, so one initially plausible theory was that revising the patient outreach method to reach individuals similar to the earlier population would restore the favorable impacts. However, this seems unlikely given the evaluation findings. The impact results during the extension period did not improve after re-weighting the post-extension sample to resemble the pre-extension sample on claims-based patient characteristics and select characteristics of the enrollees' zip code of residence. We recognize that this analysis is limited because we lack information on functional limitations, caregiver resources, health literacy, and severity of illnesses. Nonetheless, the fact that the impact estimates did not improve at all—not that they improved only somewhat—after re-weighting using the many baseline characteristics suggests that the changes in population are not the driving factor behind the decrease in impacts.

Rather, the most likely explanation for the decline in impacts is that improvements in usual care decreased the ability of HQP's services to reduce hospitalizations or expenditures relative to that heightened standard. The program impacts on hospitalizations after the extension disappeared because the risk-adjusted outcomes for the control group improved after the extension, not because the treatment group's outcomes worsened. Because decreases in hospitalization expenditures are the main mechanism by which HQP can reduce total Medicare expenditures (as was the case before the extension),[17] improvements in usual care that decrease

[17] Before the extension, reductions in expenditures for inpatient stays and post-acute care accounted for 83 percent of the total decline in Part A and B expenditures.

HQP's marginal value for reducing hospitalizations would also reduce its value for reducing total expenditures.

Several factors may have contributed to improvements in usual care for the control group. First, programs that provide services that overlap to some degree with HQP's services have expanded since the 2010 extension. This includes growth in patient-centered medical homes (one important component of which is care management for high-risk beneficiaries), establishment of new Accountable Care Organizations in the region, and hospital systems initiating care transitions programs in response to CMS policies that financially incentivize low hospital readmission rates. Thus, the control group might have received similar services to those provided by HQP to its enrollees. Furthermore, whereas one might expect HQP to still have added value if treatment group members received these new usual care services as well as those provided by HQP, in fact that did not occur in some cases. Leaders for both of the hospital systems that HQP newly worked with after the extension (and that provided about 40 percent of post-extension enrollees) indicated that the systems have their own transitional care or care management programs for which beneficiaries in the control group could be eligible, but members of the treatment group were not. These hospital systems used the HQP treatment services as a substitute for their own care management interventions.

Finally, one or more of the following factors that Krumholz (2015) cited as possibly driving the decline in Medicare FFS hospitalization rates nationally might also have contributed to the declines for the HQP control group: (1) new medications or treatment regimens for CHF, CAD, or COPD (such as greater use of statins); (2) increases in exercise, decreases in smoking, and better risk-factor management; (3) greater use of post-acute care (rehabilitation, nursing facilities and home health care) that could reduce likelihood of readmission; and (4) decreased use of hospital care at the end of life.[18]

As a result of improvements in usual care, the gaps in care that HQP met during the pre-extension period, leading to substantial declines in hospitalizations, may not exist to the same degree in the post-extension period. If this is true, it could be very difficult for the HQP model to achieve its earlier impacts in its current service area in today's environment.

However, it is possible that HQP's program could continue to have favorable impacts on hospitalizations and expenditures in other geographic regions. The added value of HQP's program depends on the counterfactual against which outcomes for the treatment group are being compared. If the improvements in usual care have been greater in HQP's service area (in eastern Pennsylvania) than they have been in other parts of the country, then the quality gaps might be as large, or even larger, in those regions as they were in the HQP service area before the extension. If so, HQP's program could continue to have its large favorable effects in these other areas. However, we do not have any direct evidence for this; exploring HQP's effectiveness in other regions would require testing the model again in those areas. Nor is it clear whether HQP's

[18] One possible explanation for the decline in hospitalization rates over time is that providers have shifted care from inpatient stays to observational stays. If true, this could mean that usual care has not actually improved, only that the site or categorization of care has changed. However, a recent study by the Medicare Payment Advisory Commission found that Medicare hospitalization rates have declined in the past decade, even after accounting for increases in observational stays (MedPAC 2015).

intervention would be a more cost-effective way of accomplishing those improvements than would implementing the other changes to FFS care that appear to have taken place in HQP's service area.

The results for outpatient ED visits suggest that other, but not identified, forces might be acting to decrease the ability of HQP's program to reduce outpatient ED visits. For this measure, usual care has not appeared to improve, because the outpatient ED visit rate did not decline for the high-risk beneficiaries in the control group after the extension. None of the tests of the hypotheses provide any compelling evidence for why HQP's program would be less effective in reducing outpatient ED visits after the extension.

## B. Elements of program design to maximize likelihood of replicating early success

Unfortunately, given the evaluation results, it is not possible to identify what specific aspects of program design are needed to maximize the chance of replicating HQP's early success. Rather, the findings highlight the need for HQP—and other care management models with the same aims of improving care while reducing expenditures—to revise the model, the service area, or both to make sure the program is filling gaps in care that (1) exist today within the program's service area and (2) contribute to reducing hospitalizations that are preventable.

In particular, models like HQP's will need to coordinate with other care management or care coordination programs operating in the same service area. These programs have become more common recently as initiatives like patient-centered medical homes, accountable care organizations, and hospital-based transitional care programs expand. One might began an investigation of whether and where important gaps in care coordination still remain in a particular geographic region by looking for the existence of the set of features found in the original MCCD analysis to be consistently associated with success in reducing hospitalizations among high-risk Medicare patients (Brown et al. 2012). This set of factors, which was present in the HQP model both before and after the extension, includes frequent in-person meetings with patients to supplement phone calls, acting as a communications hub for provider, delivering evidence-based patient education, providing strong medication management, and providing timely and comprehensive transitional care after hospitalizations.

This page has been left blank for double-sided copying.

## REFERENCES


Busso, Matias, John DiNardo, and Justin McCrary. "New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators." *The Review of Economics and Statistics*, vol. 96, no. 5, 2014, pp. 885–897.

Brown, Randall, Deborah Peikes, Greg Peterson, Jennifer Schore, and Carol Razafindrakoto. "Six Features of Medicare Coordinated Care Demonstration Programs That Cut Hospital Admissions of High-Risk Patients." *Health Affairs (Project Hope)*, vol. 31, no. 6, 2012, pp. 1156–1166.

Burwell, Sylvia. "Fifth Report to Congress on the Evaluation of the Medicare Coordinated Care Demonstration: Findings over 10 Years. Department of Health and Human Services." 2014. Available at http://innovation.cms.gov/Files/reports/MedicareCoordinatedCareDemoRTC.pdf. Accessed September 13, 2015. Report is based on analyses by Mathematica Policy Research.

Guo, S., and M.W. Fraser. "Propensity Score Analysis." *In Statistical Methods and Applications*. Los Angeles, CA: SAGE Publications, 2010.

Hirano, K., and G.W. Imbens. "Estimation of Causal Effects Using Propensity Score Weighting: An Application to Data on Right Heart Catheterization." *Health Services and Outcomes Research Methodology*, vol. 2, no. 3, 2001, pp. 259–278.

Institute of Education Sciences. "What Works Clearinghouse: Procedures and Standards Handbook, Version 3.0." Washington, DC: U.S. Department of Education, IES, 2014. Available at http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19. Accessed September 15, 2014.

Krumholz, H.M., S.V. Nuti, N.S. Downing, S.T. Normand, and Y. Wang. "Mortality, Hospitalizations, and Expenditures for the Medicare Population Aged 65 Years or Older, 1999–2013." *Journal of the American Medical Association*, vol. 314, no. 4, 2015, pp. 355–365.

Medicare Payment Advisory Commission. "Report to the Congress: Medicare Payment Policy." Washington, DC: MedPAC, March 2015. Available at http://www.medpac.gov/documents/reports/mar2015_entirereport_revised.pdf?sfvrsn=0 Accessed October 23, 2015.

Schore, Jennifer, Deborah Peikes, Greg Peterson, Angela Gerolamo, and Randall Brown. "Fourth Report to Congress on the Evaluation of the Medicare Coordinated Care Demonstration." Report submitted to the Centers for Medicare & Medicaid Services. Princeton, NJ: Mathematica Policy Research, February 10, 2011.

Zurovac, Jelena, Greg Peterson, Cara Stepanczuk, and Randy Brown. "Evaluation of the Medicare Coordinated Care Demonstration: Interim Impact Estimates for the Health Quality Partners' Program." Washington, DC: Mathematica Policy Research, November 17, 2014.

This page has been left blank for double-sided copying.

# APPENDIX A

# IMPACTS DURING THE EXTENSION FOR ALL BENEFICIARIES WHO MET THE 2010 ELIGIBILITY CRITERIA AT ENROLLMENT

This page has been left blank for double-sided copying.

This appendix presents baseline characteristics and estimated program impacts during the extension for 1,186 beneficiaries who were enrolled at any point since the program began (April 1, 2002 to December 31, 2014), who met the 2010 extension criteria at enrollment, and were alive and enrolled in FFS Medicare for at least one day during the extension period. The analyzed population includes 736 high-risk enrollees (see Section V.B.1.b for impacts on this population) and 450 CAD-only enrollees.

## A. Baseline characteristics

The enrollees had more chronic conditions and used more services in the year before enrollment than the national average for Medicare FFS beneficiaries (Appendix Table A.1). However, compared with beneficiaries enrolled during the extension (Table IV.1), the sample of all enrollees served during the extension period had fewer chronic conditions (except for the percentage with CAD) lower rates of recent service use. These differences occur because all enrollees served during the extension period include both CAD-only and high-risk enrollees, whereas all enrollees who entered during the extension are high-risk patients. Note, however, that the high-risk beneficiaries who enrolled after the extension are sicker and older than the high-risk beneficiaries who enrolled before the extension (as we discuss in Section IV.B.1.b).

## B. Impacts

In terms of hospitalizations, outpatient ED visits, and expenditures without fees (Appendix Table A.2), we found no measurable differences between the treatment and control groups. Given the large sample size, this analysis has better power than any others shown in this report, and it should reliably detect true impacts that are about 20 percent of the control group mean or larger. After factoring in program fees, the program increased total Medicare expenditures by an estimated 14.4 percent ($p = 0.04$).

Appendix Table A.1. Pre-enrollment characteristics for beneficiaries enrolled at any time during the program (2002–2014) who met 2010 eligibility criteria (CAD or high risk), by treatment status (percentages unless otherwise noted)

| | | | Health Quality Partners' enrollees | | | |
|---|---|---|---|---|---|---|
| | | Medicare FFS average (2012) (n = 32 million) | Treatment group mean (n = 606) | Control group mean (n = 580) | Difference | *p*-value |
| Age | < 65 | 16.7 | 0.0 | 0.0 | 0.0 | n.a. |
| | 65–74 | 45.5 | 46.0 | 40.9 | 5.2 | 0.19[a] |
| | 75–84 | 25.4 | 38.8 | 42.9 | -4.2 | |
| | > or = 85 | 12.4 | 15.2 | 16.2 | -1.0 | |
| Male | | 44.7 | 50.0 | 46.7 | 3.3 | 0.26 |
| Race and ethnicity[b] | Black, non-Hispanic | 10.4 | 0.8 | 2.1 | -1.2 | 0.07* |
| | Hispanic | 2.6 | 0.2 | 0.2 | 0.0 | 0.98 |
| Medicaid buy-in[c] | | 21.0 | 2.0 | 2.2 | -0.3 | 0.75 |
| Resident of original service area | | n.a. | 83.5 | 82.8 | 0.7 | 0.73 |
| Diagnosis[d] | CAD | 29.8 | 88.0 | 85.2 | 2.8 | 0.16 |
| | CHF | 15.3 | 33.0 | 31.0 | 2.0 | 0.47 |
| | Diabetes | 28.0 | 38.0 | 38.3 | -0.3 | 0.91 |
| | COPD | 11.8 | 23.8 | 25.9 | -2.1 | 0.40 |
| | Cancer[e] | NA | 14.4 | 11.9 | 2.5 | 0.21 |
| | Stroke | 4.0 | 9.2 | 8.3 | 1.0 | 0.56 |
| | Depression | 15.9 | 14.0 | 16.2 | -2.2 | 0.30 |
| | Dementia and Alzheimer's | 11.1 | 5.3 | 4.5 | 0.8 | 0.52 |
| | Osteoporosis | 6.7 | 17.2 | 15.7 | 1.5 | 0.49 |
| | Rheumatoid arthritis | 30.3 | 32.7 | 32.4 | 0.3 | 0.92 |
| | CKD | 16.2 | 19.3 | 17.9 | 1.4 | 0.54 |
| | Atrial fibrillation | 8.2 | 22.9 | 25.7 | -2.8 | 0.27 |
| Number of chronic conditions (of 12 above) | | 1.5 | 3.2 | 3.1 | 0.0 | 0.62 |
| In the year before enrollment | Annualized hospitalizations (number) | 0.295 | 0.973 | 0.985 | -0.012 | 0.85 |
| | Medicare Part A and B expenditures ($ PBPM) | 860 | 1,450 | 1,514 | -64 | 0.55 |
| | Any hospital stay for CAD, CHF, COPD, or diabetes | NA | 0.2 | 0.2 | 0.0 | 0.43 |
| | Any use of home health | NA | 27.9 | 29.0 | -1.1 | 0.68 |
| | Any use of a SNF | NA | 9.6 | 10.7 | -1.1 | 0.52 |

| | | | Health Quality Partners' enrollees | | | |
|---|---|---|---|---|---|---|
| | | Medicare FFS average (2012) (n = 32 million) | Treatment group mean (n = 606) | Control group mean (n = 580) | Difference | *p*-value |
| Characteristics of enrollees' residence zip code, mean | Median household income ($) | 51,371 | 80,962 | 80,102 | 861 | 0.44 |
| | College degree or more | 28.5 | 39.8 | 39.1 | 0.7 | 0.31 |
| | Unemployment rate | 8.1 | 7.3 | 7.4 | 0.0 | 0.71 |

Sources: Medicare National Claims History File, Standard Analytic File, Enrollment Databases and the American Community Survey. Medicare FFS totals come from the CCW, Medicare Beneficiary Prevalence for Chronic Conditions for 2003 Through 2012, Table B.2 (https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_b2.pdf). Monthly expenditures and annualized hospitalizations are exceptions and come from the 2013 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, Table V.1 (http://downloads.cms.gov/files/TR2013.pdf) and the Health Indicators Warehouse developed by the National Center for Health Statistics (http://www.healthindicators.gov/Indicators/Hospital-inpatient-Medicare-admissions-per-1000-beneficiaries_2001/Profile/ClassicData), respectively. Household income comes from the 2012 American Community Survey, Table S1901 (Income in the Past 12 Months [in 2012 Inflation-Adjusted Dollars]). Education status comes from the 2012 American Community Survey, 5-Year Estimates, Table S15003 (Educational Attainment). Unemployment rate comes from the Bureau of Labor Statistics Current Population Survey, Table 1 (http://www.bls.gov/cps/aa2012/cpsaat01.pdf).

Notes: The sample includes beneficiaries enrolled from April 1, 2002, to June 30, 2014, who met either (1) the high-risk criteria—that is, they had CAD, CHF, COPD, or diabetes and one or more hospitalizations in the year before enrollment—or (2) the CAD-only criteria—beneficiaries with CAD but no hospitalization in the year before enrollment.

[a]Only one *p*-value is reported for the treatment-control differences in age because a chi-squared test was used to determine whether the overall age distribution for the treatment group was different from the distribution for the control group.

[b]Includes all (not only FFS) Medicare beneficiaries who were enrolled on or after January 1, 2012. Total number of beneficiaries is 53.6 million.

[c]Medicaid buy-in indicates that the beneficiary is eligible for both Medicare and Medicaid. The Medicare FFS average was approximated using the percentage of Medicare beneficiaries who were dual eligibles in 2010. See http://kff.org/medicaid/state-indicator/duals-as-a-of-medicare-beneficiaries.

[d]Diagnoses are based on the 2010 version of CCW definitions. The definitions use a look-back period of one year before enrollment for atrial fibrillation, cancer, COPD, depression, osteoporosis, and stroke and two years for CAD, CHF, CKD, diabetes, and rheumatoid arthritis. The evaluation used a two-year look-back period for dementia and Alzheimer's, rather than the three years used by CCW, because of the limits of the Medicare claims data extracted for the analysis.

[e]This category excludes skin cancer.

CAD = coronary artery disease; CCW = CMS Chronic Conditions Data Warehouse; CHF = congestive heart failure; CKD = chronic kidney disease; COPD = chronic obstructive pulmonary disease; FFS = fee-for-service; SNF = skilled nursing facility; PBPM = per beneficiary per month.

NA = not available.

n.a. = not applicable.

*Significantly different from zero at the .1 level, two-tailed test.

## Appendix Table A.2. Program impacts during the extension on service use and Medicare expenditures among beneficiaries enrolled at any time during the program (2002–2014) who met 2010 eligibility criteria (CAD or high risk)

| | Control group mean | Treatment group mean | Difference (90% confidence interval) | Percentage difference | *p*-value |
|---|---|---|---|---|---|
| **Program impacts on outcomes** | | | | | |
| Annualized hospitalizations (#/person/year) | 0.616 | 0.640 | 0.024 (-0.059, 0.106) | 3.9 | 0.64 |
| Annualized outpatient ED visits (#/person/year) | 0.643 | 0.600 | -0.043 (-0.119, 0.034) | -6.7 | 0.36 |
| Medicare Part A and B expenditures ($ PBPM) — Without program fees | 1,452 | 1,507 | 56 (-110, 222) | 3.9 | 0.58 |
| Medicare Part A and B expenditures ($ PBPM) — With program fees (mean fee paid = $260) | 1,452 | 1,660 | 209 (43, 374) | 14.4 | 0.04** |
| Medicare Part A expenditures | 774 | 809 | 35 (-93, 162) | 4.5 | 0.65 |
| Medicare Part B expenditures | 677 | 698 | 21 (-38, 80) | 3.1 | 0.56 |
| **Sample sizes, average length of follow-up, and receipt of treatment services** | | | | | |
| Number of beneficiaries | 580 | 606 | n.a. | n.a. | n.a. |
| Mean number of follow-up months | 35.7 | 35.9 | n.a. | n.a. | n.a. |
| Mean percentage of follow-up months during which beneficiary received treatment[a] | 0.0 | 79.6 | n.a. | n.a. | n.a. |

Sources: Medicare Enrollment Database, National Claims History File, Standard Analytic File, and the Mathematica randomization file.

Notes: The sample includes beneficiaries enrolled from April 1, 2002, to June 30, 2014. Beneficiaries met either (1) the high-risk criteria—that is, they had CAD, CHF, COPD, or diabetes and one or more hospitalizations in the year before enrollment—or (2) they had CAD but no hospital stay in the previous year (the CAD-only group).

The control group mean is weighted but not regression-adjusted. The treatment group mean equals the control group mean plus the regression-adjusted difference between treatment and control. The percentage difference is calculated by dividing the regression-adjusted difference between the treatment and control groups by the control group mean.

Treatment-control differences were adjusted for baseline characteristics to increase the precision of the estimates and to account for chance differences between the two groups. Negative estimates of treatment-control differences indicate that hospitalizations or expenditures were lower for the treatment group.

The regressions use observations that are weighted according to the proportion of the follow-up period during which each sample member met CMS's demonstration-wide requirements: being in fee-for-service, having both Part A and Part B coverage, having Medicare as the primary payer, and being alive for part of the month. Weights were calculated separately for the treatment and control groups.

[a]Calculated as follows: (1) for each beneficiary, calculate the percentage of follow-up months during which he or she received treatment services (a beneficiary was considered to have received services in a given month if HQP submitted a bill for services for that beneficiary in that month) and (2) find the average across treatment group beneficiaries.

CAD = coronary artery disease; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; ED = emergency department; PBPM = per beneficiary per month.

**Significantly different from zero at the .05 level, two-tailed test.

# APPENDIX B

# METHODS FOR USING PROPENSITY SCORES TO RE-WEIGHT THE POST-EXTENSION SAMPLE TO RESEMBLE THE PRE-EXTENSION SAMPLE

This page has been left blank for double-sided copying.

This appendix provides additional detail on the methods we used to test the hypothesis that program impacts for high-risk beneficiaries declined after the extension because HQP served an older and a more medically complex population. To test this hypothesis, we examined whether the post-extension impact estimates improve if we weight the beneficiaries in the post-extension population so that they resemble the pre-extension population.

In this appendix, we first briefly describe the traditional use of propensity score weighting and how we adapted this technique for the MCCD evaluation. We then provide details about how we specified the propensity score logistic regression, obtained propensity scores, generated propensity score weights, and used the weights to estimate program impacts on a post-extension population that resembles the pre-extension sample. Finally, we explain why the assumptions underlying propensity score analyses are reasonable in our context and why we chose propensity weighting over matching on propensity scores.

## A. Adapting traditional propensity score techniques to the MCCD evaluation

The propensity score method is typically used in non-experimental settings to induce similarity between treatment and comparison groups before treatment effects are assessed (Hirano and Imbens 2001; Guo and Fraser 2010; Busso, DiNardo, and McCrary 2014). In this case, the propensity score is the estimated probability, based on observable characteristics, that a person received treatment. Estimates of program impact are generated by comparing those who actually received treatment to a comparison group that is either matched or weighted using the propensity score so that it is similar to the treatment group on observable characteristics at baseline.

We adapted this method to weight the post-extension sample so it resembles the healthier pre-extension sample of enrollees in order to test whether impacts during the post-extension period would change. That is, the propensity score in our application is the predicted probability, based on observable enrollee characteristics, of an enrollee being in the pre-extension group, rather than being in the treatment group.

After re-weighting, we continue to use the randomized controlled design to estimate impacts. Therefore, one traditional concern about propensity score designs—that they may yield biased *impact* estimates because the treatment and comparison groups might differ in ways that are not observable—does not apply here. Our impact estimates remain unbiased because they rely on random assignment to define the treatment and control groups for estimating impacts. As discussed in Section C of this appendix, re-weighting does not affect treatment and control balance at baseline (as expected, because treatment assignment is uncorrelated with patient characteristics). However, as noted in the main text (Section VI.E), the re-weighted post-extension sample could be different from the pre-extension sample on unobservable characteristics. Although this does not bias impact estimates, it does place some limitation on the extent to which the re-weighted analysis can fully test whether impacts during the extension would differ if provided to beneficiaries who were similar to the pre-extension enrollees on both observable and unobservable characteristics.

## B. Generating propensity score weights

To re-weight the post-extension sample so it resembles the pre-extension sample on observable enrollee characteristics, we estimated a logistic regression for the high-risk beneficiaries enrolled before and after the extension. In this regression, the dependent variable denotes whether an enrollee is part of the pre-extension sample and explanatory variables include all available patient characteristics measured before enrollment (at baseline) that might influence analyzed outcomes—service use and Medicare expenditures. A logistic regression gives the conditional probability of belonging to the pre-extension sample:

$$\Pr(\text{Pre}_i | X_i = x_i) = E(\text{Pre}_i) = \frac{e^{x_i\beta_i}}{1+e^{x_i\beta_i}} \quad (1)$$

where $Pre_i$ is a binary variable equal to 1 for an enrollee *i* if he or she is a pre-extension enrollee and equal to 0 for enrollee *i* if he or she is a post-extension enrollee, $X_i$ is the vector of patient characteristics, and $\beta_i$ is the vector of corresponding estimated parameters. The sample for this model included all 849 high-risk beneficiaries who enrolled between April 1, 2002, and June 30, 2014: 366 (or 43 percent) of whom enrolled before the extension and 483 (or 57 percent) who enrolled after.

The estimated coefficients from the logit model are used to calculate a propensity score for each enrollee, which is the estimated probability that an enrollee with those characteristics was a part of the pre-extension sample. The propensity score thus summarizes the likelihood of each post-extension enrollee to have been a part of the pre-extension sample. Using the estimated logit coefficients $\beta_i$, we apply the equation to each enrollee's data to obtain the predicted probability of being in the pre-extension group, that is, the propensity score $\hat{p}(x)$:

$$\hat{p}(x) = \Pr(\text{Pre}_i = 1 | X_i = x_i) \quad (2)$$

Through this process, we generated predicted probabilities of being in the pre-extension period for both those who were part of the pre-extension sample (n = 366) and those who were not (n = 483). The average predicted probability of being in the pre-extension period across all 849 beneficiaries was 43 percent, which is the percentage of pre-extension enrollees in the overall sample. The average predicted probability among those who were actually in the pre-extension sample was 52 percent; the average among those who were in the post-extension sample was 36 percent.

A key step in this analysis is creation of a weight using the propensity score. To do so, we assigned each pre-extension enrollee a weight of one. Post-extension enrollees receive a weight proportional to how similar they are to pre-extension enrollees: the more similar a given post-extension enrollee is to pre-extension enrollees, the larger the weight in the regression. First, we assigned each post-extension enrollee a weight equal to $\frac{\hat{p}(x)}{1-\hat{p}(x)}$, where $\hat{p}(x)$ is the estimated propensity score from equation 2 (Busso, DiNardo, and McCrary 2014; Guo and Fraser 2010). In the traditional application of propensity score weighting, this weighting approach enables one

to compute the average treatment effect on the treated (ATT); in the current application, the ATT estimate corresponds to the average effect for the post-extension sample.

We then multiplied each enrollee's weight by the normalized weight denoting the number of months that enrollee is eligible during his or her follow-up period. To ensure that the effective sample sizes within treatment and comparison groups remain the same, we then normalized the newly obtained weight by dividing each enrollee's weights by the average weight in the treatment or comparison groups, depending on which group the enrollee belongs to. Finally, we used the combined normalized weight in the impact analyses of post-extension impacts.

As described in Section VI.E of the main text, the pre- and post-extension samples were similar on all baseline characteristics after re-weighting the post-extension sample, which is an important condition for a credible comparison of pre- and post-extension impacts (Table VI.3).

## C. Similarity between the treatment and control groups after weighting

After weighting, but before estimating impacts, a key step is to demonstrate that the re-weighted treatment and control groups for the post-extension analysis are still balanced at baseline. Doing so shows that the control group is still an appropriate counterfactual for the treatment group, allowing for the regression-adjusted differences between the groups to be interpreted as program impacts.

Re-weighting should not affect the balance between the treatment and control groups. Because assignment to treatment is random, it is unrelated to patient characteristics at baseline. Therefore, treatment assignment should be unrelated to the propensity scores, which are based on characteristics at baseline. Because treatment assignment is unrelated to propensity scores and because we weight both treatment and control beneficiaries, re-weighting based on propensity scores should not affect the balance between the treatment and control groups.

To confirm empirically that the treatment and control groups remained balanced after re-weighting, we focused on mean standardized differences as a key measure of differences between the pre- and post-extension groups. The standardized difference for a given variable is the difference in means between the pre- and post-extension enrollees divided by the standard deviation for that variable that is pooled across the two groups. We regarded differences as large if they are equal to or greater than 0.25 standard deviations, an industry standard for dissimilarity of treatment and control groups in randomized studies that cannot be corrected with regression adjustment (for example, see Institute of Education Sciences 2014).

As shown in Appendix Table B.1, the treatment and control groups remained well balanced after re-weighting the sample to resemble the pre-extension group. None of the differences between the two groups exceeded 0.15 standardized differences, and none was statistically significant at the $p < 0.05$ level.

Appendix Table B.1. Baseline characteristics of high-risk beneficiaries who enrolled after the extension with weighting to make them look like the pre-extension enrollees, by treatment status (percentages unless otherwise noted)

| | | Treatment enrollees (n = 241) | Control enrollees (n = 242) | Difference | *p*-value | Standardized difference |
|---|---|---|---|---|---|---|
| Age | < 65 | 0.0 | 0.0 | 0.0 | 1.0 | n.a. |
| | 65–74 | 40.6 | 32.1 | 8.5 | 0.20 | 0.12 |
| | 75–84 | 48.1 | 55.2 | -7.1 | 0.41 | -0.07 |
| | > or = 85 | 11.3 | 12.7 | -1.4 | 0.56 | -0.05 |
| Male | | 49.0 | 50.2 | -1.2 | 0.88 | -0.01 |
| Race and ethnicity | Black, non-Hispanic | 0.4 | 1.2 | -0.8 | 0.16 | -0.13 |
| | Hispanic | 0.0 | 0.4 | -0.4 | 0.32 | -0.09 |
| Medicaid buy-in[a] | | 4.8 | 2.9 | 1.9 | 0.56 | 0.05 |
| Original reason for Medicare entitlement | Disability | 5.8 | 3.3 | 2.5 | 0.15 | 0.13 |
| | ESRD | 0.0 | 0.0 | 0.0 | 0.32 | -0.09 |
| Diagnosis[b] | CAD | 83.6 | 82.0 | 1.6 | 0.86 | 0.02 |
| | CHF | 39.7 | 35.0 | 4.7 | 0.44 | 0.07 |
| | Diabetes | 39.2 | 44.8 | -5.6 | 0.43 | -0.07 |
| | COPD | 25.5 | 23.9 | 1.6 | 0.68 | 0.04 |
| | Cancer[c] | 14.6 | 15.0 | -0.5 | 0.93 | -0.01 |
| | Stroke | 13.5 | 13.5 | 0.0 | 0.98 | 0.00 |
| | Depression | 14.3 | 13.8 | 0.5 | 0.87 | 0.01 |
| | Dementia and Alzheimer's | 4.9 | 4.0 | 0.8 | 0.65 | 0.04 |
| | Osteoporosis | 22.9 | 19.9 | 3.0 | 0.61 | 0.05 |
| | Rheumatoid arthritis | 34.2 | 29.2 | 5.0 | 0.40 | 0.08 |
| | CKD | 12.8 | 15.6 | -2.8 | 0.28 | -0.10 |
| | Atrial fibrillation | 34.2 | 33.0 | 1.2 | 0.85 | 0.02 |
| Hospitalized for the following conditions in the year before enrollment | CAD | 24.9 | 15.1 | 9.8 | 0.13 | 0.14 |
| | CHF | 7.2 | 6.2 | 1.0 | 0.74 | 0.03 |
| | Diabetes | 0.1 | 0.8 | -0.7 | 0.09 | -0.15 |
| | COPD | 4.7 | 4.6 | 0.1 | 0.97 | 0.00 |
| | Cancer[c] | 0.0 | 0.0 | 0.0 | 0.32 | 0.09 |
| | Depression | 0.2 | 0.2 | 0.0 | 0.89 | -0.01 |
| | Dementia and Alzheimer's | 0.0 | 0.0 | 0.0 | 1.00 | n.a. |
| | Osteoporosis | 0.0 | 0.0 | 0.0 | 1.00 | n.a. |
| | Rheumatoid arthritis | 5.3 | 4.3 | 1.0 | 0.62 | 0.05 |
| | CKD | 1.5 | 1.0 | 0.6 | 0.45 | 0.07 |
| | Atrial fibrillation | 8.7 | 7.2 | 1.5 | 0.69 | 0.04 |
| | Hip fracture | 2.1 | 0.6 | 1.5 | 0.17 | 0.13 |
| In the year before enrollment | Annualized hospitalizations (number) | 1.44 | 1.45 | -0.01 | 0.94 | -0.01 |
| | Any use of home health | 34.6 | 29.9 | 4.7 | 0.33 | 0.09 |
| | Any use of a SNF | 7.3 | 7.9 | -0.6 | 0.75 | -0.03 |

| | | Treatment enrollees (n = 241) | Control enrollees (n = 242) | Difference | *p*-value | Standardized difference |
|---|---|---|---|---|---|---|
| Two years before enrollment | Annualized hospitalizations (number) | 1.04 | 0.96 | 0.07 | 0.51 | 0.06 |
| | Number of conditions treated per person[d] | 3.39 | 3.29 | 0.10 | 0.73 | 0.03 |
| Characteristics of enrollees' zip code of residence, mean | Median household income ($) | 82,661 | 83,730 | -1,069 | 0.89 | -0.01 |
| | College degree or more | 41.9 | 41.3 | 0.6 | 0.89 | 0.01 |
| | Unemployment rate | 6.8 | 7.1 | -0.3 | 0.58 | -0.05 |
| **Omnibus test for balance on baseline characteristics[e]** | | | | | | |
| *p*-value | | n.a. | n.a. | n.a. | 0.38 | n.a. |

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

Notes: The sample includes beneficiaries enrolled from October 1, 2010, to June 30, 2014. All beneficiaries met the "high-risk" criteria—that is, they had CAD, CHF, COPD, or diabetes and one or more hospitalizations in the year before enrollment. This analysis uses additional covariates not used in other impact regressions—for example, hospitalization for each of several conditions in the year before enrollment and sociodemographic makeup of enrollees' residence zip codes. However, it excludes the variable denoting whether the enrollee is a resident of the original service area because this variable was not measured before enrollment.

[a]Medicaid buy-in indicates that the beneficiary is eligible for both Medicare and Medicaid. The Medicare FFS average was approximated using the percentage of Medicare beneficiaries who were dual eligibles in 2010. See http://kff.org/medicaid/state-indicator/duals-as-a-of-medicare-beneficiaries.

[b]Diagnoses are based on the 2010 version of CCW definitions. The definitions use a look-back period of one year before enrollment for atrial fibrillation, cancer, COPD, depression, osteoporosis, and stroke and two years for CAD, CHF, CKD, diabetes, and rheumatoid arthritis. The evaluation used a two-year look-back period for dementia and Alzheimer's, rather than the three years used by CCW, because of the limits of the Medicare claims data extracted for the analysis.

[c]This category excludes skin cancer.

[d]Number of conditions for which one is treated is counted over a one- or two-year period, depending on the look-back period.

[e]Results from an overall chi-squared test indicate the likelihood of observing differences in the explanatory variables as large as the differences we observed if, in fact, the treatment and control populations were perfectly balanced. The value of $p = 0.38$ for the omnibus test suggests that the two groups are well-balanced because we cannot reject the null hypothesis that their characteristics are identical.

CAD = coronary artery disease; CCW = CMS Chronic Conditions Data Warehouse; CHF = congestive heart failure; CKD = chronic kidney disease; COPD = chronic obstructive pulmonary disease; ESRD = end-stage renal disease; FFS = fee-for-service; SNF = skilled nursing facility.

n.a. = not applicable.

## D. Meeting the assumptions behind propensity score weighting

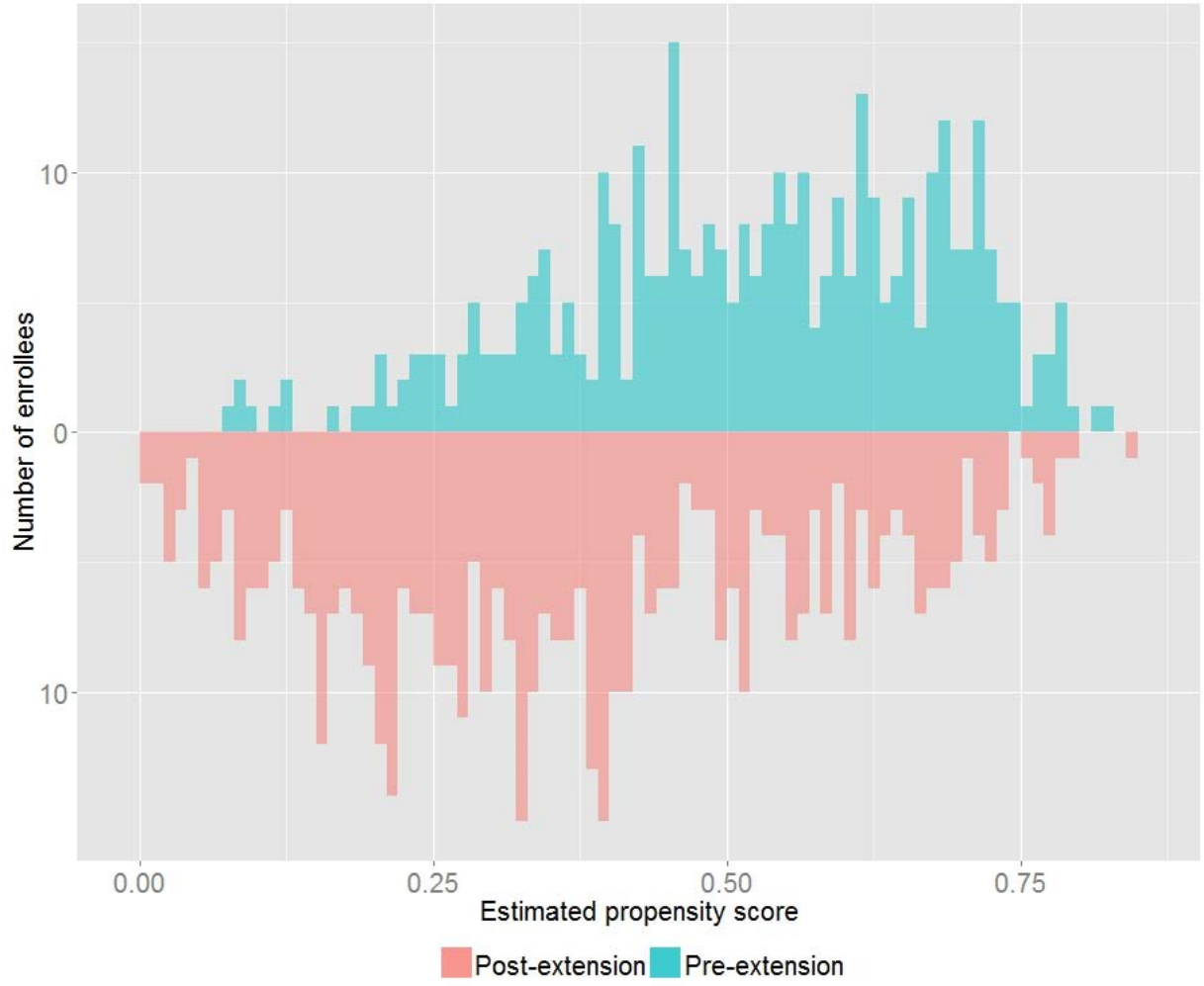
Literature shows that, in their traditional application, propensity scores can yield unbiased impact estimates when the likelihood of belonging to the treatment group is modeled appropriately and, after controlling for observable characteristics, selection into the treatment group is exogenous (Busso, DiNardo, and McCrary 2014; Guo and Fraser 2010). In our application, these conditions do not apply directly because we rely on random assignment to generate impact estimates. However, it is important to meet analogous conditions in our context to demonstrate that, after re-weighting based on propensity scores, the post-extension group is, in fact, similar to the pre-extension sample. Therefore, in our context, the analogous conditions are that (1) we have modeled selection into the pre-extension sample correctly and (2) conditional on observable characteristics, selection into the pre-extension sample is exogenous.

There are two reasons we believe that both conditions generally hold in our context. First, even though the outreach method for identifying patients changed after the extension, the underlying eligibility criteria are the same in the two periods, meaning that the two populations are not wildly different from one another without weighting. Second, the propensity model includes a rich set of covariates to control for the patient differences that results from the differences in the outreach method. However, there may be some factors, such as functional limitations, that influence selection into the pre-extension sample. We cannot capture those in our propensity models, and this remains a limitation of the analysis.

## E. Rationale for selecting propensity score weighting over matching

An alternative to propensity weighting would have been to match post-extension enrollees to pre-extension enrollees with similar propensity scores, and to drop from the impact analysis anyone in the post-extension sample who did not have a match. We chose the re-weighting approach instead because it retains all post-extension enrollees in the impact analysis (albeit some with small weights), maximizing statistical power to detect program effects. Re-weighting may perform worse than propensity score matching in creating well-balanced groups when the two groups differ greatly (Busso, DiNardo, and McCrary 2014). As discussed earlier, even though the pre- and post-extension samples differ considerably for the purpose of comparing impacts for the two populations, they were quite similar for purposes of propensity score weighting. We confirmed that there was abundant overlap between the two groups before matching by examining the distribution of estimated propensity score for the pre- and post-extension enrollees (see Figure B.1).

Figure B.1. Estimated propensity scores (predicted likelihood of belonging to the pre-extension group) for pre- and post-extension enrollees




Sources: Medicare Enrollment Database, National Claims History File, Standard Analytic File, and Mathematica randomization file.

Notes: The sample includes high-risk beneficiaries enrolled before the extension period (from April 1, 2002, to March 31, 2010) and after the extension (October 1, 2010, to June 30, 2014). All beneficiaries met the high-risk criteria, that is, they have CAD, CHF, COPD, or diabetes and one or more hospitalizations in the year before enrollment).

The x-axis shows the number of enrollees who have each of the values of the propensity scores, shown on the y-axis.

This page has been left blank for double-sided copying.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research