# HSR Health Services Research

© Health Research and Educational Trust DOI: 10.1111/1475-6773.12595 BEST OF THE 2016 ACADEMYHEALTH ANNUAL RESEARCH MEETING

# Testing the Replicability of a Successful Care Management Program: Results from a Randomized Trial and Likely Explanations for Why Impacts Did Not Replicate

G. Greg Peterson, Jelena Zurovac, Randall S. Brown, Kenneth D. Coburn, Patricia A. Markovich, Sherry A. Marcantonio, William D. Clark, Anne Mutti, and Cara Stepanczuk

**Objectives.** To test whether a care management program could replicate its success in an earlier trial and determine likely explanations for why it did not.

**Data Sources/Setting.** Medicare claims and nurse contact data for Medicare fee-forservice beneficiaries with chronic illnesses enrolled in the trial in eastern Pennsylvania (N = 483).

**Study Design.** A randomized trial with half of enrollees receiving intensive care management services and half receiving usual care. We developed and tested hypotheses for why impacts declined.

**Data Extraction.** All outcomes and covariates were derived from claims and the nurse contact data.

**Principal Findings.** From 2010 to 2014, the program did not reduce hospitalizations or generate Medicare savings to offset program fees that averaged \$260 per beneficiary per month. These estimates are statistically different (p < .05) from the large reductions in hospitalizations and spending in the first trial (2002–2010). The treatment–control differences in the second trial disappeared because the control group's risk-adjusted hospitalization rate improved, not because the treatment group's outcomes worsened.

**Conclusion.** Even if demonstrated in a randomized trial, successful results from one test may not replicate in other settings or time periods. Assessing whether gaps in care that the original program filled exist in other settings can help identify where earlier success is likely to replicate.

**Key Words.** Care management, care coordination, chronic disease, Medicare Coordinated Care Demonstration, cost savings

Care management for patients with chronic conditions is central to many current health care delivery reforms that aim to improve quality of care while reducing medical expenditures. By educating patients in self-management skills, coordinating care across a range of providers, and linking patients to community and social services, care management programs seek to reduce the need for hospitalizations and, as a result, lower total cost of care. The Health Quality Partners program (HQP), which serves Medicare beneficiaries in eastern Pennsylvania, is one of a few care management interventions that, in a randomized trial, has proven to reduce hospitalizations (Counsell et al. 2007; Schore et al. 2011; Brown et al. 2012; Peikes et al. 2012; Hong, Siegel, and Ferris 2014). The HQP program has been widely recognized as a promising model for future testing and possible expansion (Klein 2013).

HQP was one of several organizations to participate in the Medicare Coordinated Care Demonstration (MCCD), which the Centers for Medicare & Medicaid Services (CMS) launched in 2002 to test a range of care management interventions within fee-for-service (FFS) Medicare (Peikes et al. 2009). In the first phase of the demonstration (2002–2010), HQP enrolled Medicare FFS beneficiaries with a wide range of illness severities, hypothesizing that program effects may vary by severity. HQP's program did not reduce hospitalizations or expenditures for its full enrolled population (Burwell 2014). However, for the 15 percent of enrollees at high risk of hospitalization, the program reduced hospitalizations by an estimated 34 percent and Medicare expenditures (including monthly program fees) by 22 percent (Brown et al. 2012; Burwell 2014). The high-risk subgroup was defined as those with coronary artery disease (CAD), congestive heart failure (CHF), or chronic obstructive pulmonary disease (COPD) and at least one hospitalization in the year before enrollment.

In 2010, CMS launched a second phase of the demonstration (2010–2014), extending HQP's program but limiting new enrollment to those who met the high-risk criteria, for whom CMS paid \$281 per beneficiary per

Address correspondence to G. Greg Peterson, Ph.D., M.P.A., Senior Health Researcher, Mathematica Policy Research, 1100 1st Street, NE, Washington, DC; e-mail: GPeterson@Mathematica-Mpr.com. Jelena Zurovac, Ph.D., M.S., and Anne Mutti, M.P.A., are with Mathematica Policy Research, Washington, DC. Randall S. Brown, Ph.D., and Cara Stepanczuk, M.P.A., are with Mathematica Policy Research, Princeton, NJ. Kenneth D. Coburn, M.D., Ph.D., F.A.C.P., and Sherry A. Marcantonio, M.S.W., A.C.S.W., are with the Health Quality Partners, Doylestown, PA. Patricia A. Markovich, Ph.D., M.P.P., and William D. Clark, M.S., are with the Research and Rapid Cycle Evaluation Group, Center for Medicare and Medicaid Innovation, Centers for Medicare and Medicaid Services, Baltimore, MD.

month (PBPM). HQP also continued to serve high-risk members who enrolled in Phase I. The goal of the extension was to test whether the successful results could be replicated, or perhaps strengthened, when the program was targeted exclusively to the high-risk population, which represents about 18 percent of all Medicare beneficiaries nationally and 37 percent of all Medicare expenditures (Brown et al. 2012).

Interim findings based on the first 3 years of the demonstration's second phase (Zurovac et al. 2014) indicated that the program had not reduced hospitalizations or expenditures. The evaluation shifted focus to assess the likely explanations for why impacts declined in the second phase. Understanding why impacts declined could help organizations adopting an intervention similar to HQP's determine how to revise the model to maximize the prospects of achieving HQP's earlier success.

In complex interventions like HQP's care management program, many factors can contribute to why a program's impacts do not replicate, including important—even if unintended—changes in the content of the intervention, the process through which the intervention is delivered, and the target population (Mahoney 2010). We developed four hypotheses for why impacts declined:

**Hypothesis 1: Shorter tenure.** If program impacts increase the longer a patient is enrolled, impacts in Phase II may be smaller because average patient tenure in the program was shorter (Phase I ran for 8 years while Phase II ran for 4 years).

**Hypothesis 2: Improvements in usual care.** Since 2010, other health care organizations in HQP's service area have introduced their own care management programs which may overlap with HQP's services, limiting the ability of HQP's program to reduce hospitalizations further. These programs, prompted in part by incentives and initiatives in the 2010 Patient Protection and Affordable Care Act (ACA), include hospital-based transitional care and care management through patient-centered medical homes (PCMHs) and Accountable Care Organizations (ACO).

**Hypothesis 3: Changes in patient population.** To ensure that Phase II enrollees met the high-risk definition, HQP began identifying prospective patients from hospital discharge records rather than by using its earlier method of reviewing billing or medical records from participating medical practices. This shift in patient identification method unintentionally led to the Phase II population being older and having more chronic conditions than the high-risk population in Phase I (Zurovac et al. 2014). The intervention may have been less effective if the more complex patient population required more frequent nurse contacts or was less able or willing to make the changes in self-management that HQP's program encouraged.

#### 2118 HSR: Health Services Research 51:6, Part I (December 2016)

**Hypothesis 4: Decline in intervention intensity.** Facing a consistently high-risk caseload, HQP's care managers may not have been able to contact high-risk enrollees as frequently in Phase II as they had in Phase I. During Phase I, care managers served patients with a wide range of risk levels and could triage their contacts to 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 the decrease may not have been sufficient.

This paper has two objectives. The first was to present final results from Phase II of the demonstration, estimating program impacts over its full 4 years (October 2010 to December 2014) on hospitalizations, outpatient emergency department (ED) visits, Medicare expenditures, and survival. The second goal was to use varied data sources and methods to assess the plausibility of each of the four hypotheses for why impacts declined and to draw conclusions about what explanation(s) are most likely.

# METHODS

#### Patient Recruitment and Randomization

For the Phase II intervention, HQP partnered with three hospital systems in eastern Pennsylvania (Doylestown Health, Crozer-Keystone Health System, and St. Mary Medical Center). Hospitals produced quarterly lists of Medicare FFS beneficiaries with a hospital stay in the previous year and a qualifying diagnosis (CAD, CHF, COPD, or diabetes). HQP reviewed the lists for patients who (i) had a primary care provider (PCP), with whom HQP's care managers could work in coordinating care, (ii) did not have one of several conditions that could limit the intervention's effectiveness,<sup>1</sup> and (iii) were age 65 or older. Care managers asked the PCPs to review lists of potential enrollees and remove those who were not eligible. Finally, care managers contacted the eligible patients to describe the HQP intervention and elicit their participation. About 30 percent of contacted patients consented to participate.

After a beneficiary enrolled, a website randomly assigned the beneficiary to a treatment or control group (with a 50 percent likelihood of assignment to either group). The treatment group received intensive care management services in addition to usual care provided through Medicare FFS, while the control group continued to receive usual care.

#### Intervention

The Phase II intervention was delivered by nurse care managers, supported by experienced supervisors and a medical director and in collaboration with patients' usual medical providers. Each care manager had a caseload of about 75 beneficiaries. Care managers delivered services to a beneficiary from enrollment until the end of the program in December 2014 or until the beneficiary died, moved out of state, or voluntarily disenrolled. The mean length of follow-up was 25.5 months, and, because disenrollment rates were very low, beneficiaries received treatment services (as proxied by CMS paying for care management services rendered each month) for, on average, 93 percent of their follow-up months. The nurses were deployed from HQP's central office in Doylestown or from participating hospitals but spent much of their time in patients' homes or meeting with patients and providers in medical offices.

After beneficiaries enrolled, nurse care managers met with them in their homes for a comprehensive assessment to identify their physical, functional, cognitive, psychological, behavioral, social, and environmental needs. Nurse care managers developed an individualized plan for each beneficiary that identified specific priority issues and interventions that aimed, first, to stabilize any new or worsening conditions and, second, to ensure that beneficiaries were receiving recommended preventive care. Nurses prioritized items for the plan based on (i) the beneficiary's articulated concerns and unmet needs, (ii) findings from risk assessments (initial and repeated), and (iii) the beneficiary's motivational readiness (Coburn et al. 2012). After the initial assessment, nurses met their patients regularly-on average twice per month, with 60 percent of the contacts in-person (primarily in the beneficiary's home, and also in the office or hospital) and the rest by telephone. During the contacts, nurses educated patients on self-management skills (tailored to a person's readiness to change; Prochaska and DiClemente 1983), reconciled and managed medications and counseled on adherence, monitored symptoms, and arranged and monitored community health and social service referrals. A small fraction (12 percent) of enrollees also participated in group exercise or weight maintenance classes led by nurse care managers.

The nurses collaborated with patients' PCPs and specialists around specific clinical issues to help beneficiaries achieve target clinical goals, receive appropriate preventive care according to guidelines, and facilitate timely interventions to prevent disease exacerbation. If a patient was admitted to the hospital (all hospitals partnering with HQP fed daily data to HQP on when enrollees were hospitalized), the nurses initiated a care transitions protocol. The protocol included coordinating with hospital and posthospital care providers and meeting with patients within 3 days of discharge to make sure discharge plans were safe, to conduct comprehensive medication review and reconciliation, and to ensure that patients followed discharge instructions, including timely follow-up visits with PCPs and specialists.

The Phase II intervention was similar to the Phase I intervention (Archibald and Schore 2003; Coburn et al. 2012), with modifications to serve the more complex needs of the consistently high-risk population in Phase II. In Phase II, the nurses had lower caseloads (75 vs. 108), met more often with patients in their home and with caregivers, and spent more time addressing complex psychosocial needs and coordinating care with PCPs and specialists. HQP also offered fewer group classes because more beneficiaries were homebound and there was no critical mass of beneficiaries able or willing to participate in some classes. HQP used program standards and analytic reports to manage and improve service performance in both phases of the demonstration.

#### Impact Estimation

We estimated impacts as the regression-adjusted difference in outcomes between the Medicare FFS beneficiaries in the treatment and control groups. The study outcomes, all derived from Medicare claims and enrollment data, are as follows: annualized number of hospitalizations, annualized number of outpatient ED visits, Medicare Part A and B expenditures (with and without monthly program fees), and 2-year mortality. The regressions adjusted for baseline characteristics, including demographics, Medicaid enrollment, and previous service use and expenditures. The regressions increased the precision of the estimates and controlled for observable, chance differences between the treatment and control groups. We used an intent-to-treat approach, keeping enrollees in the sample as long as they were observable in Medicare FFS claims, regardless of whether they continued to receive the intervention. We weighted the observations by the number of months that a beneficiary was observable from enrollment to the program's end in December 2014. Consistent with earlier studies (Peikes et al. 2009) and because CMS was particularly concerned about falsely concluding that a program had no impacts, we used a p < .10 threshold for statistical significance (two-tailed tests).

The primary sample included the 483 treatment and control beneficiaries who enrolled during Phase II. The estimates are independent of impact estimates from Phase I. To increase statistical power to detect effects, we also estimated impacts for a secondary sample that included an additional 253 beneficiaries who enrolled during Phase I, met the high-risk criteria at enrollment, and were observable for at least part of Phase II.<sup>2</sup> For both the primary and secondary samples, we examined outcomes during Phase II only (October 2010 to December 2014). The statistical power to detect effects on hospitalizations at least as large as the point estimates in Phase I (Burwell 2014) was 87 and 96 percent for the primary and secondary samples, respectively.

We tested whether the impact estimates for Phase II (using the primary sample) were statistically different from Phase I impacts for high-risk beneficiaries (N = 322), as reported in the evaluation's fifth report to Congress (Burwell 2014).

#### Testing the Four Hypotheses for Why Program Impacts Declined

*Shorter Tenure.* We compared Phase I and II impact estimates for high-risk beneficiaries, controlling for the length of time a beneficiary was enrolled in the program. Specifically, we estimated impacts as the regression-adjusted differences in outcomes in enrollees' first year of follow-up, second year of follow-up, and 1–3 years of follow-up, if those periods fell fully during Phase I or II. If the hypothesis is true, the patterns in impacts on hospitalizations, outpatient ED visits, and Medicare expenditures in Phases I and II should be similar after controlling for year of enrollment.

*Improvements in Usual Care.* We examined the risk-adjusted outcomes (hospitalizations and outpatient ED visits) for the treatment and control groups in Phase II of the demonstration versus Phase I. If improvements in usual care drove the decline in impacts, we would expect (i) the risk-adjusted outcomes to remain the same for the Phase I and II treatment groups and (ii) the adjusted hospitalization or ED visit rates for the control group in Phase II to be lower than the rates for the control group in Phase I (signaling improvement in usual care), erasing the difference in outcomes for the treatment and control groups in Phase II. It is important to risk-adjust the outcomes because the Phase II high-risk population was older and had more chronic conditions than the Phase I population, so comparisons of unadjusted means would confound improvements in usual care with changes in the population. The regression included 849 high-risk beneficiaries, of whom 366 enrolled during Phase I and 483 enrolled during Phase II. We measured hospitalizations through 3 years of enrollment to control for patient tenure. The regressions controlled for the same explanatory variables as in other analyses, except that we added a binary for enrollment cohort (Phase I or II) and an interaction between treatment status and enrollment cohort.

*Changes in Patient Population.* We used propensity scores to reweight the Phase II sample members so that they resembled the Phase I sample members on observable baseline characteristics. To create the propensity scores, we pooled the 366 high-risk beneficiaries in the Phase I sample with the 483 in the Phase II sample. We used all available baseline characteristics to predict whether a sample member was in the Phase I sample and then used regression coefficients to generate a propensity score for each member—where the propensity was the predicted probability of being in the Phase I sample. We reweighted the Phase II sample members with an inverse probability weight equal to the propensity score divided by one minus the propensity score (Guo and Fraser 2010). After confirming that the Phase II treatment and control groups remained balanced at baseline after the reweighting, we reestimated impacts, weighting observations by the product of their inverse probability weight, and the standard weight described earlier based on the number of observable months. If the hypothesis that changes in patient population drove the decline in impacts was true, we would expect the impact estimates to be more favorable after reweighting the Phase II sample so that the sample resembles the Phase I high-risk sample on observable baseline characteristics.

*Lower Intervention Intensity.* We analyzed HQP's data on all contacts its care managers made with or on behalf of their patients throughout both the demonstration phases. We linked these data to the Medicare claims and enrollment data to (i) limit the Phase I sample to those who met the high-risk definition and (ii) regression-adjust the results to control for changes in patient characteristics from Phase I to Phase II. Specifically, we assessed whether risk-adjusted monthly contact rates (number of contacts per enrollee per month) for treatment group members in the first year<sup>3</sup> of program enrollment differed in Phase I versus Phase II. If this hypothesis is correct, we would expect to see a decline in the intervention intensity—as measured by the contact rates—after controlling for changes in patient population. This is a necessary but not sufficient condition for explaining the decline in impacts; that is, for the hypothesis to be true, the intensity was the reason impacts declined.

# RESULTS

#### Patient Characteristics

The program enrollees at baseline were two to four times more likely to have CAD, CHF, COPD, or diabetes than the national Medicare FFS average, and their average hospitalization rate (1.7 per year) during the previous year was about six times the national average of 0.3 (Table 1). Almost all enrollees were white, non-Hispanic, and not enrolled in Medicaid. The treatment and control groups were very similar, as expected from random assignment. Although the criteria for high-risk status were the same in Phases I and II, the high-risk group in Phase II was, on average, older, had higher recent rates of hospitalization and use of home health, and had more chronic conditions—with the largest differences for COPD, kidney disease, and depression.

#### Program Impacts

There were no measurable differences between the treatment and control groups for hospitalizations, outpatient ED visits, 2-year mortality rates, or Medicare Part A and B expenditures—for either the primary or secondary sample (Table 2). Including monthly program fees that averaged \$260 PBPM, the treatment group's total Medicare expenditures were 16 percent higher than the comparison group's expenditures. Although the increase in expenditures was not statistically significant for the primary sample, it was significant for the larger secondary sample that had greater power to detect effects (p = .08).

The difference in impact estimates for Phase I and Phase II was large and statistically significant for hospitalizations and Medicare Part A and B expenditures, with and without program fees (Table 2). The Phase I–Phase II difference in the odds ratio for 2-year survival was also large but not statistically significant (p = .13). Because Phase I lasted much longer than Phase II (8 vs. 4 years), the mean patient exposure to the intervention was higher in Phase I than in Phase II (49.4 vs. 25.5 months).

### Tests of the Four Hypotheses for Why Impacts Declined

*Shorter Tenure.* In Phase I of the demonstration, the program reduced hospitalizations by 0.21 per person per year (p = .04) and outpatient ED visits by 0.25 per person per year (p = .01) in the first three follow-up years. It reduced Medicare Part A and B expenditures without program fees by \$360 PBPM

inrollees
Ε
n-Risl
Higł
iase I
Ч
s and
ollee
Enro
Ξ
of Phase ]
laracteristics o
C
Baseline
1:
Table

			2			
		d	hase II Enrollees		Phase I High-Risk Enrollees	Phase II Enrollees, Weighted*
Characteristic	Medicare FFS Average (2012) (n = 32 million)	Treatment Group Mean (n = 241)	Control Group Mean (n = 242)	Difference	Treatment and Control (n = 366)	Treatment and Control (n = 483)
Age, %						
Under 65	16.7	0.0	0.0	0.0	0.0	0.0
65-74	45.5	40.2	32.6	7.6	36.9	36.4
75-84	25.4	35.7	40.5	-4.8	48.6	51.6
85 and older	12.4	24.1	26.9	-2.8	14.5	12.0
Male, %	44.7	46.1	42.1	3.9	51.1	49.6
Race and $\operatorname{ethnicity}^{\dagger}, \%$						
Black, non-Hispanic	10.4	1.7	3.3	-1.6	1.4	0.8
Hispanic	2.6	0.0	0.4	-0.4	0.3	0.2
Eligible for both Medicare	$21.0^{\ddagger}$	2.5	2.5	0.0	2.7	3.9
and Medicaid, %						
$Diagnosis^{\$}, 0,0$						
Coronary artery disease	29.8	79.3	74.0	5.3	82.8	82.8
Congestive heart failure	15.3	51.0	45.9	5.2	38.0	37.3
Diabetes	28.0	43.6	48.8	-5.2	42.6	42.0
Chronic obstructive pulmonary	11.8	42.3	41.7	0.6	26.0	24.7
disease						
Cancer, excluding skin cancer	NA	16.6	14.5	2.1	13.1	14.8
Stroke	4.0	13.7	12.0	1.7	12.0	13.5
Depression	15.9	24.9	25.2	-0.3	14.5	14.0
Dementia and Alzheimer's	11.1	9.5	8.3	1.3	4.9	4.5
Osteoporosis	6.7	19.9	20.7	-0.7	18.0	21.4
Rheumatoid arthritis	30.3	41.5	41.0	0.0	33.3	31.7
						Continued

2124

# HSR: Health Services Research 51:6, Part I (December 2016)

		P	hase II Enrollees		Phase I High-Risk Enrollees	Phase II Enrollees, Weighted*
Characteristic	Medicare FFS Average (2012) (n = 32 million)	Treatment Group Mean (n = 241)	Control Group Mean (n = 242)	Difference	Treatment and Control (n = 366)	Treatment and Control (n = 483)
Chronic kidney disease Atrial fibrillation Number of chronic conditions (of 12 above)	16.2 8.2 1.5	34.0 32.4 4.1	32.6 31.0 4.0	$1.4 \\ 1.4 \\ 0.1$	14.5 33.3 3.3	14.2 33.6 3.3
In the year before enrollment Annualized hospitalizations, number Medicare Part A and B expenditures,	0.295 860	1.604 2,390	1.726 2,521	-0.123 -131	1.45 NA <sup>¶</sup>	1.44 NA <sup>¶</sup>
doutars per peneticiary per monun Any use of home health, % Any use of a skilled nursing facility, % Characteristics of enrollees' residence	NA NA	51.9 18.7	52.1 23.6	-0.2 -4.9	35.0 7.9	32.2 7.6
ZIP code, mean Median household income, dollars College degree or more, % Unemployment rate, %	51,371 28.5 8.1	79,715 38.8 7.5	79,151 37.9 7.6	$564 \\ 1.0 \\ -0.1$	81,776 40.8 7.1	83,197 41.6 7.0
Notes. For each sample for which impacts v were no statistically significant differences l *The Phase II enrollees are weighted to ress <sup>†</sup> Includes all (not only fee-for-service [FFS] <sup>‡</sup> The Medicare FFS average was approxim medicaid/state-indicator/duals-as-a-for-fmed <sup>§</sup> Diagnoses were based on the 2010 version for dementia and Alzheimer's rather than tl <sup>fi</sup> We did not use Medicare expenditures in NA, not available.	were analyzed (Pha between treatment: semble the Phase I e []) Medicare benefic nated by using the dicare-beneficaries on of the Chronic C he 3 years used by generating the weig	use II enrollees, Ph. and control groups mrollees on baselin iaries who were em percentage of Med Orditions Warehou CCW because of th to make the two	ase I high-risk enror $(p > .15$ for all varie e characteristics. e characteristics. colled on or after Ja icare beneficiaries ise (CCW) definiti ase (CCW) definiti ne limits of the Med groups similar bec	illees, and Phi ables). nuary 1, 2012. who were dui ons. The eval licare claims d ause of inflati	use II enrollees) afte ul eligibles in 2011. 1ation used a 2-yeau ata extracted for the m in Medicare expe	sr weighting, there See http://kff.org/ r look-back period : analysis. inditures.

Table 1. Continued

2125

Table 2: Outcom	es during Phase II ar	nd Comparison to Pł	iase I Outcomes		
		Phase II Outcomes			
Outcome	Primary Sample- Beneficiaries Enrolled During Phase II (n = 483)	Secondary Sample- Beneficiaries Enrolled During Phuse II and Phase I High-risk Enrollees Who Remained Enrolled During Phase II (n = 736)	Primary Sample After Weighting <sup>+</sup> (n = 483)	$Phase I Outcomes^{2}$ $(n = 322)$	Difference in Outcomes for Phase II (Primary Sample) and Phase I <sup>§</sup>
Annualized hospitalizations Control group mean Regression-adjusted treatment-control	number per person per yea 0.743 0.065 (-0.086, 0.216)	. $0.723$ $0.039$ $(-0.080, 0.157)$	0.524 0.106 (-0.025, 0.237)	0.872 $-0.293^{***}(-0.458, -0.129)$	-0.129 $0.359^{***}$ $(0.136, 0.581)$
difference (90% confidence interval) Annualized outpatient energency department visits, number per person per year Control group mean Regression-adjusted treatment-control difference (90% confidence interval) Medicare Part A and B expenditures, dollars per beneficiary per month	0.747 0.010 ( $-0.135$ , $0.156$ )	$\begin{array}{c} 0.729 \\ -0.059 \left( -0.174, 0.056  ight) \end{array}$	0.655 -0.017 (-0.145, 0.111)	A N A N	NA NA

2126

Continued

		Phase II Outcomes			
Outcome	Primary Sample- Beneficiaries Enrolled During Phase II (n = 483)	Secondary Sample- Beneficiaries Enrolled During Phase II and Phase I High-risk Enrollees Who Remained Enrolled During Phase II (n = 736)	Primary Sample After Weighting <sup>*</sup> (n = 483)	$Phase I \ Outcomes^{4}$ (n = 322)	Difference in Outcomes for Phase II (Primary Sample) and Phase I <sup>§</sup>
Without program fees Control group mean Regression-adjusted treatment-control difference (90%	1,748 16 ( $-315,347$ )	1,694 29 (–218, 277)	1,284 118 (-151, 386)	1,415 -425** (-698, -152)	333 441* (13, 689)
conndence mterval) With program fees Regression-adjusted treatment-control difference (10,0%	276(-55,607)	$264^{*}(17,512)$	$380^{**}(112, 648)$	$-313^{*}(-587, -40)$	589** (161, 1,017)
confidence interval) confidence interval) Mean program fees paid for beneficiaries in treatment group,	\$260	\$235	\$2.62	\$112	\$148
dollars per beneficiary per month Died within 2 years of enrollment <sup>1</sup> , % Control group mean, %	ان ۲.	NA	VN	13.8	1.7 Continued

Replicability of a Successful Care Management Program

2127

Table 2. Continued

		Phase II Outcomes			
Outcome	Primary Sample- Beneficiaries Erwolled During Phase II (n = 483)	Secondary Sample- Beneficiaries Enrolled During Phase II and Phase I High-risk Enrollees Who Remained Enrolled During Phase II (n = 736)	Primary Sample After Weighting <sup>+</sup> (n = 483)	$Phase I Outcomes^{\pm}$ (n = 322)	Difference in Outcomes for Phase II (Primary Sample) and Phase I <sup>§</sup>
Regression-adjusted odds ratio (90% confidence interval)	$0.88\ (0.49-1.59)$	NA	NA	$0.33(0.13-0.81)^{**}$	$0.55^{\dagger \dagger}$
Mean number of follow- up months (treatment and control)	25.5	30.8	24.5	49.4	-23.9
Note: Sample sizes are tr * $p < .10$ , ** $p < .05$ , *** †The Phase II enrollees i *The results from Phase size in Table 1 ( $n = 366$ ) %The $p$ -values are for test "The sample is limited to sample for Phase II and it* $p = .13$ for the test that NA, not available.	eatment and control gro > < .01. ure weighted to resemble t come from Burwell (20 , which includes benefic s of the null hypothesist n beneficiaries who enror n = 248 for Phase 1). the odds ratios in the tw	ups combined. The contro is the Phase I enrollees on by 114), who used high-risk em itaries who enrolled through itaries who enrolled through in the impact estimates in blled early enough to be fo o phases are the same.	l group mean is weight iseline characteristics. ollees through June 2( March 2010. the two phases are the llowed up for 2 years l	ed but not regression-adjusted 009. The sample size $(n = 322]$ same. before the end of the period (;	l. ) differs from the sample n = 369 for the primary

2128

Table 2. Continued

(p = .048) (Table 3). The estimated effect on expenditures with program fees was not statistically significant, but the point estimate indicated a large decrease (\$244 PBPM, p = .18). In Phase II, 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 = .09).

Improvements in Usual Care. The annualized hospitalization rate was essentially the same for the Phase II treatment group as for the Phase I treatment group, after controlling for differences in patient characteristics (Table 4). In contrast, the Phase II control group's hospitalization rate was 34 percent lower than the Phase I control group's rate (0.544 vs. 0.825 hospitalizations per beneficiary per year), a difference that was highly significant (p = .009). This pattern means that the large impact estimate seen in Phase I disappeared during Phase II because the control group's risk-adjusted outcomes improved. The result is robust to sensitivity tests that reduced the influence of outliers by trimming the observations with rates above the 99th (or 98th) percentile to the 99th (or 98th) percentile. In contrast, the outpatient ED visit rate for the Phase II treatment group was substantially higher than Phase I treatment group's rate (0.766 vs. 0.415, p < .001), whereas the rate for the two comparison groups was essentially unchanged (p = .45).

*Changes in Patient Population.* Reweighting the Phase II sample with propensity scores made the sample similar to the Phase I high-risk group on all baseline characteristics (final column in Table 1). After reweighting, impact estimates did not improve (Table 2). The point estimates remained essentially the same for outpatient ED visits and became slightly less favorable for hospitalizations and Medicare Part A and B expenditures, but none was statistically significant.

*Decrease in Intervention Intensity.* The nurse contact rates with or on behalf of enrollees were substantially *higher* during Phase II than during Phase I before and after adjusting for differences in the patient population (Table 5). The differences were largest for in-person contacts, which more than doubled (from 0.68 to 1.41 per person per month). However, the mean number of group

Table 3:Regression-Adjusted DiffPatient Follow-Up and Demonstration	ferences in Mean on Phase	n Outcomes b	etween Treatme	int and Con	trol Groups	by Year of
	Phase I (1	4pril 2 002–Septemb	er 2010)	Phase II (O	ctober 2010–Dece	mber 2014)
	First Year of Follow-up (n = 366)	Second Year of Follow-up (n = 348)	First Through Third Years of Follow-up (n = 366)	First Year of Follow-up (n = 483)	Second Year of Follow-up (n = 386)	First Through Third Years of Follow-up (n = 483)
Regression-adjusted difference in mean outco Annualized hospitalizations, number per	omes between treatme -0.119 (.37)	ent and control gro -0.227 (.11)	ups ( <i>p</i> -value) -0.206** (.04)	0.074~(.54)	0.077 (.59)	0.069~(.46)
person per year ( <i>p</i> -value) Annualized outpatient emergency department visits, number per	$-0.210^{**}$ (.0495)	$-0.312^{*}$ (.053)	$-0.254^{***}$ (.007)	(0.059 (.60)	-0.027 (.85)	0.014 (.88)
person per year ( $p$ -value) Medicare Part A and B expenditures, dollars $_{ m I}$	per beneficiary per m	onth ( <i>p</i> -value)				
Without program fees With program fees	-379 (.14) -963 (.30)	-245 $(.27)-131$ $(.56)$	$-360^{**}$ (.048) -944 (18)	-196 $(.49)76 (.79)$	293 (.36) 549* (.09)	19(.93)
Average length of follow-up Mean number of eligible follow-up months (freatment and control)	11.7	10.2	29.2	11.4	10.3	24.5

Notes. Sample sizes are treatment and control groups combined. \* p < .10, \*\* p < .05, \*\*\* p < .01.

2130

Table 4: Hospitalizations and Outpatient Emergency Department Visits for High-Risk Beneficiaries in Phases I and II, by Treatment Status, Controlling for Changes in Patient Population

	ז ומ		Adjusted Difference
	Phase I	Phase II	(p-value)
Annualized hospitalizations, num	ber per person per year		
Treatment	0.632	0.614	-0.018 (.86)
Control	0.825	0.544	-0.281 (.009***)
Adjusted difference (p-value)	-0.193(0.06*)	0.069(0.44)	0.262 (.05*)
Annualized outpatient emergency	v department visits, per p	person per year	
Treatment	0.415	0.766	0.351 (<.001***)
Control	0.663	0.739	0.075(.45)
Adjusted difference (p-value)	$-0.248(0.009^{***})$	0.027(0.74)	0.276 (.03**)
Sample sizes			
Treatment	188	241	
Control	178	242	
Combined	366	483	

Notes. The control group means in Phase I of the demonstration are weighted but not regressionadjusted. The means for the treatment group in Phase I and the means for both the treatment and control groups in Phase II are calculated by adding appropriate coefficient(s) from the regression model (described in text) to the Phase I control group mean.

p < .10, p < .05, p < .05, p < .01.

Table 5: Program Contacts with Patients in the First Year of Enrollment in Phase I and II, with and without Adjusting for Changes in Patient Population

Contact with	Mean Contact Rate (per Beneficiary per Month)		t males for	Mean C Benefic	ontact Rate (per iary per Month)	p-value for	
or on Behalf of a Beneficiary	Phase I	Phase II, Unadjusted	Difference (percent)	p-value for Unadjusted Difference	Phase II, Adjusted	Difference <sup>†</sup> (percent)	p-value for Adjusted Difference
Individual, any	2.13	3.22	1.10 (51.5)	<.001***	2.77	0.64 (30.2)	.004***
Individual, in-person	0.68	1.60	0.92 (135.1)	<.001***	1.41	0.73 (107.7)	<.001***
Individual, with provider	0.20	0.49	0.29 (143.3)	<.001***	0.35	0.15 (73.1)	.048**
Group (as part of participation in a group class)	0.44	0.08	-0.36 (-81.3)	<.001***	0.10	-0.35 (-78.5)	<.001***
Individual (any) or group	2.57	3.30	0.73 (28.5)	.001***	2.87	0.29 (11.4)	.190

\*\*p < .05, \*\*\*p < .01.

<sup>†</sup>Phase II adjusted rate minus the Phase I rate.

contacts declined substantially (from 0.44 to 0.10). As a result, the total contact rate (individual or group) was only modestly higher in the Phase II sample than in the Phase I sample (2.87 vs. 2.57 contacts per person per month), and this difference was not statistically significant (p = .19).

# DISCUSSION

Over the 4 years of the demonstration's second phase, the HQP program did not achieve its earlier success. The program did not reduce hospitalizations, outpatient ED visits, 2-year mortality rates, or Medicare expenditures relative to the control group. Therefore, there were no Medicare savings that could offset the program fees that averaged \$260 PBPM. The Phase II findings were substantively and, for most outcomes, statistically different from the Phase I findings, which showed that the program reduced hospitalizations and ED visits, improved survival, and decreased expenditures net of program fees for high-risk enrollees.

#### Likely Explanations for Why Impacts Declined

Based on the tests of the four hypotheses, the most plausible explanation for why impacts declined is that improvements in usual care made it more difficult for HQP's services to reduce hospitalizations further. The main evidence supporting this hypothesis is that, after controlling for differences in patient population, the difference in impacts disappeared because the hospitalization rate for the control group in Phase II improved (relative to the rate for the control group in Phase I)—not because the rate for the treatment group worsened. Because decreases in hospitalization expenditures are the main mechanism by which HQP can reduce total Medicare expenditures (as was the case in Phase I),<sup>4</sup> the improvements in usual care would also limit the program's ability to reduce total expenditures. However, improvements in usual care do not appear to drive the reduction in program impacts on outpatient ED visits because the ED rate increased for the treatment group (while the outcomes for the control group remained unchanged).

The substantial decrease in participation in group classes, which persisted after controlling for differences in patient risk between the two phases, may have contributed to the decline in impacts. HQP stopped offering many of its group classes in Phase II. After disenrolling the low-risk beneficiaries (who tended to be more mobile), HQP no longer had a critical mass of beneficiaries willing and able to participate in the classes. Previous studies have found that self-management classes offered in group settings can reduce hospitalization rates for people with chronic illnesses (Lorig et al. 1999).

The evidence does not support the other hypotheses. The program impacts in Phase I appeared within the first 3 years of patient enrollment while no such impacts appeared within the first 3 years of enrollment in Phase II, meaning that differences in patient tenure cannot explain the decline in impacts. Weighting the Phase II population to resemble the Phase I population did not improve the point estimates for program impacts in Phase II. This suggests that the changes in the patient population did not drive the change in impacts. Finally, there is no evidence that—aside from the decline in group classes—the intensity of the intervention declined. Indeed, the number of nurse contacts with or on behalf of patients—particularly in-person contacts were substantially higher for Phase II enrollees than for Phase I high-risk enrollees.

The idea that usual care has improved is consistent with national estimates showing that the total number of hospitalizations declined by nearly 24 percent from the start of Phase I in 2002 through the end of 2013 (Krumholz et al. 2015). Hospitalizations steadily declined during Phase I, with an average of 1.7 percent per year, but much more quickly during Phase II, averaging 4.2 percent per year. Daughtridge, Archibald, and Conway (2014) show that, even though some of the declines from 2009 to 2013 are attributable to shifts in site of care from inpatient to observational stays, much of the decline appears to result from genuine improvements in care.

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 during Phase II. The ACA's Hospital Readmissions Reduction Program has incentivized hospitals in the region—including the three HQP partnered with in Phase II—to initiate transitional care programs to reduce readmission rates (Zuckerman et al. 2016). Care management is a central component of the Medicare ACO models, also launched as part of ACA. In HQP's service area, the Renaissance ACO was formed during Phase II and was associated with lowering hospitalization rates, although not Medicare spending (L&M Policy Research 2015; Nyweide et al. 2015). Thus, the control group might have received services similar to those that HQP provided to its enrollees.

The growth in PCMHs (Edwards et al. 2014) may also have improved usual care. One study (Friedberg et al. 2014) found that a PCMH pilot in

eastern Pennsylvania that spanned the two phases of the MCCD (2008–2011) did not reduce hospitalizations, but many practices in the region continued PCMH efforts beyond 2011 and studies from other parts of Pennsylvania indicate that these efforts have reduced hospitalizations and ED visits (Friedberg et al. 2015). Finally, one or more of the following factors cited by Krumholz et al. (2015) as possibly driving the national decline in Medicare FFS hospitalization rates might also have contributed to the declines for the HQP control group: (i) new medications or treatment regimens for CAD, CHF, or COPD (such as greater use of statins or increased implantation of stents); (ii) increases in exercise, decreases in smoking, and better risk-factor management; (iii) greater use of postacute care (rehabilitation, nursing facilities, and home health care) that could reduce the likelihood of readmission; and (iv) decreased use of hospital care at the end of life.

#### Other Possible Explanations

Although the tests suggest the most plausible explanations, we cannot rule out two alternative explanations. First, the Phase II enrollees and high-risk Phase I enrollees may have differed in important ways that were not captured in the claims data, and so the propensity weighting results may not fully reflect the impacts HQP's program would have had if two populations were more alike. The ability of care management programs to reduce expenditures may be sensitive to the intensity, range, and nature of the risk and complexity of the target population; the specific services provided; and the duration and continuity of care. Overall, enrollees in Phase II represented a significantly higher risk group. It is possible that unobserved differences in factors known to increase risk substantially among high-need older adults such as physical (e.g., frailty), emotional, and cognitive deficits contributed to the reductions in the intervention's effects on resource utilization. Incorporating data on such factors, which were collected by HQP on all study participants before randomization, was beyond the scope of this analysis. Findings from such an analysis could further inform our understanding of baseline differences between intervention groups in Phases I and II and could lead to different results in our simulation of how different impacts would have been had the two cohorts been more similar.

Second, the results may have been driven by chance—Phase I's favorable results may not have been real (false positives), and, given the modest statistical power, Phase II's results may have been a false negative (Ioannidis 2005; Nosek 2015). The concern that the original findings may have been a false positive is reasonable because they are based on a subgroup with a small sample. However, several factors that Sun et al. (2014) listed as important for validating subgroup findings suggest that the results were real. First, the impacts for the high-risk group on hospitalizations and expenditures were statistically different (p < .01) from the impacts for those not in the high-risk group. Second, the finding of larger effects for the high-risk subgroup was consistent across several of the randomized trials in the MCCD (Brown et al. 2012). Finally, larger impacts for this high-risk group are clinically plausible because these beneficiaries are more likely to have preventable hospitalizations and the targeted conditions (CAD, CHF, COPD, and diabetes) are, according to earlier research, particularly amenable to care management (Lewis 2010).

# CONCLUSIONS

These findings have several implications for future care management interventions and research. First, the ability to demonstrate, in a rigorous randomized trial, that a program was effective in one setting does not guarantee that it will be so again in another setting or time period. Since the passage of the ACA, the rate of change in the U.S. health care system has accelerated dramatically. This trend will only continue as CMS pursues its goal of tying 50 percent of Medicare FFS payments to quality or value through alternative payment models by 2018 (Health and Human Services, 2015). Because usual care is evolving rapidly, for a program to be effective, it needs to address gaps in care that exist in the current environment. One way that research can help inform whether a successful care management program in a given setting will add value relative to the standard of care in a different setting is by developing a detailed understanding of the mechanisms for how successful programs work, including the key gaps in care that the program addressed and were critical to its success. Researchers and program designers can then assess whether those gaps exist in a different setting or time period. Our results also suggest that care management programs may work best in places where risk-adjusted hospitalization rates for the target population are high (compared to other regions) and have not shown the same steady decline as observed nationally in the past decade.

These results also highlight the importance of aligning care management activities sparked by the ACA to ensure that they are complementary. This will only become more important as usual care continues to evolve with the expansion of new models of care, including PCMHs, ACOs, and bundled payments—along with FFS Medicare paying directly for some of the transitional care and other care management services for chronically ill beneficiaries that previously were only funded under pilots like the MCCD.

### ACKNOWLEDGMENT

Joint Acknowledgment/Disclosure Statement: The analyses upon which this publication is based were performed under contract HHSM-500-2014-00429G, funded by the Centers for Medicare & Medicaid Services (CMS), Department of Health and Human Services. The views expressed in this article are solely those of the authors and do not necessarily represent the policy or views of the Centers for Medicare and Medicaid Services (CMS), or does the mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. CMS also funded both phases of the Medicare Coordinated Care Demonstration (MCCD).

Kenneth Coburn is President of Health Quality Partners (HQP) and Sherry Marcantonio is Senior Vice President and, as such, both have a financial interest in the success of HQP, which implemented the care management intervention described in this manuscript. This manuscript is based on analyses that were conducted by Mathematica Policy Research (MPR) in an independent evaluation of the MCCD, which culminated in a separate evaluation report authored only by MPR staff. HQP contributed to the independent evaluation by providing nurse contact data and by meeting with MPR staff frequently to describe the intervention and discuss possible reasons for the decline in impacts between the demonstration's two phases. Dr. Coburn and Ms. Marcantonio are coauthors of this publication because of these critical contributions to this paper and because, after the independent evaluation was complete, they contributed to the further analysis of the results and their policy implications.

We thank the many people who contributed to this manuscript, including the following Mathematica Policy Research employees: Carol Razafindrakoto, Xiaofan Sun, and Huihua Lu for statistical analyses; Mark Flick and Lisa Shang for programming and data management; and Deborah Peikes, Arnold Chen, and Jennifer Schore (former Mathematica employee) for their leadership of the evaluation during the MCCD's first phase. We also thank Robert Lazansky at HQP for data aggregation and analyses; Renee Mentnech at CMS for overall project guidance; HQP staff, including nurse care managers, who met with MPR throughout the demonstration; and the manuscript's two anonymous reviewers. Finally, we thank the many Medicare beneficiaries who participated in the MCCD.

*Disclosures*: No other disclosures. *Disclaimers*: None.

# NOTES

- 1. HQP excluded patients with amyotrophic lateral sclerosis, Alzheimer's disease, dementia, cancer, end-stage renal disease, HIV/AIDS, Huntington's disease, psy-choses or schizophrenia, and those who resided in long-term care facilities.
- 2. The 253 beneficiaries were still balanced on observable factors at the start of Phase II, meaning that the treatment during Phase I did not create an imbalance that could bias Phase II impact estimates.
- 3. We assessed contact rates in the first year of enrollment to avoid confounding that would occur if contact rates varied by year of follow-up (given that length of follow-up varied between Phases I and II). We also assessed Phase I–Phase II differences in contact rates in Years 2 and 3 of enrollment and findings were substantively similar.
- 4. In Phase I, reductions in expenditures for inpatient stays and postacute care accounted for 83 percent of the total decline in Part A and B expenditures.

# REFERENCES

- Archibald, N., and J. Schore. 2003. The Early Experience of the Health Quality Partners Case Management Program. Princeton, NJ: Mathematica Policy Research, Inc. MPR reference number 8756-310.
- Brown, R., D. Peikes, G. Peterson, J. Schore, and C. Razafindrakoto. 2012. "Six Features of Medicare Coordinated Care Demonstration Programs That Cut Hospital Admissions of High-Risk Patients." *Health Affairs (Project Hope)* 31 (6): 1156–66.
- Burwell, S. 2014. "Fifth Report to Congress on the Evaluation of the Medicare Coordinated Care Demonstration: Findings over 10 Years." Washington, DC: U.S. Department of Health and Human Services [accessed on September 13, 2015]. Available at http://innovation.cms.gov/Files/reports/MedicareCoordi natedCareDemoRTC.pdf
- Coburn, K. D., S. Marcantonio, R. Lazansky, M. Keller, and N. Davis. 2012. "Effect of a Community-Based Nursing Intervention on Mortality in Chronically Ill Older Adults: A Randomized Controlled Trial." *PLoS Medicine* 9 (7): e1001265.
- Counsell, S. R., C. M. Callahan, D. O. Clark, W. Tu, A. B. Buttar, T. E. Stump, and G. D. Ricketts. 2007. "Geriatric Care Management for Low-Income Seniors: A Randomized Controlled Trial." *Journal of the American Medical Association* 298 (22): 2623–33.

- Daughtridge, G. W., T. Archibald, and P. H. Conway. 2014. "Quality Improvement of Care Transitions and the Trend of Composite Hospital Care." *Journal of the American Medical Association* 311 (10): 1013–4.
- Edwards, S. T., A. Bitton, J. Hong, and B. E. Landon. 2014. "Patient-Centered Medical Home Initiatives Expanded in 2009-13: Providers, Patients, and Payment Incentives Increased." *Health Affairs (Millwood)* 33 (10): 1823–31.
- Friedberg, M. W., E. C. Schneider, M. B. Rosenthal, K. G. Volpp, and R. M. Werner. 2014. "Association between Participation in a Multipayer Medical Home Intervention and Changes in Quality, Utilization, and Costs of Care." *Journal of the American Medical Association* 311 (8): 815–25.
- Friedberg, M. W., M. Rosenthal, R. Werner, K. Volpp, and E. Schneider. 2015. "Effects of a Medical Home and Shared Savings Intervention on Quality and Utilization of Care." *Journal of the American Medical Association Internal Medicine* 175 (8): 1362–8.
- Guo, S., and M. W. Fraser. 2010. "Propensity Score Analysis." In *Statistical Methods and Applications*. Los Angeles: SAGE.
- Health and Human Services. 2015. "Better, Smarter, Healthier: In Historic Announcement, HHS Sets Clear Goals and Timeline for Shifting Medicare Reimbursements from Volume to Value" [accessed on August 3, 2016]. Available at http://www.hhs.gov/about/news/2015/01/26/better-smarter-healthier-in-historicannouncement-hhs-sets-clear-goals-and-timeline-for-shifting-medicare-reimbur sements-from-volume-to-value.html
- Hong, C. S., A. L. Siegel, and T. G. Ferris. 2014. "Caring for High-Need, High-Cost Patients: What Makes for a Successful Care Management Program?" *The Commonwealth Fund Issue Brief* 19: 1–19.
- Ioannidis, J. 2005. "Why Most Published Research Findings Are False." *PLoS Medicine* 2 (8): 697–701.
- Klein, E. 2013. "If This Was a Pill, You'd Do Anything to Get It." *The Washington Post* [accessed on April 17, 2016]. Available at https://www.washingtonpost.com/ news/wonk/wp/2013/04/28/if-this-was-a-pill-youd-do-anything-to-get-it
- Krumholz, H. M., S. V. Nuti, N. S. Downing, S. T. Normand, and Y. Wang. 2015. "Mortality, Hospitalizations, and Expenditures for the Medicare Population Aged 65 Years or Older, 1999–2013." *Journal of the American Medical Association* 314 (4): 355–65.
- L&M Policy Research. 2015. "Evaluation of CMMI Accountable Care Organization Initiatives: Pioneer ACO Evaluation Findings from Performance Years One and Two." Washington, DC: L&M Policy Research [accessed on April 17, 2016]. Available at https://innovation.cms.gov/Files/reports/PioneerACOEvalRpt2.pdf
- Lewis, G. H. 2010. "Impactibility Models: Identifying the Subgroup of High-Risk Patients Most Amenable to Hospital-Avoidance Programs." *The Milbank Quarterly* 88 (2): 240–55.
- Lorig, K. R., D. S. Sobel, A. L. Stewart, B. W. Brown Jr, A. Bandura, P. Ritter, V. M. Gonzalez, D. D. Laurent, and H. R. Holman. 1999. "Evidence Suggesting That a Chronic Disease Self-Management Program Can Improve Health Status While Reducing Hospitalization: A Randomized Trial." *Medical Care* 37 (1): 5–14.

- Mahoney, J. E. 2010. "Why Multifactorial Fall-Prevention Interventions May Not Work: Comment on 'Multifactorial Intervention to Reduce Falls in Older People at High Risk of Recurrent Falls'." Archives of Internal Medicine 170 (13): 1117.
- Nosek, B. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349 (6251): aac4716.
- Nyweide, D., W. Lee, T. Cuerdon, H. Pham, M. Cox, R. Rajkumar, and P. Conway. 2015. "Association of Pioneer Accountable Care Organizations vs. Traditional Medicare Fee for Service with Spending, Utilization, and Patient Experience." *Journal of the American Medical Association* 313 (21): 2152–61.
- Peikes, D., A. Chen, J. Schore, and R. S. Brown. 2009. "Effects of Care Coordination on Hospitalization, Quality of Care, and Health Care Expenditures among Medicare Beneficiaries: 15 Randomized Trials." *Journal of the American Medical Association* 301 (6): 603–18.
- Peikes, D., G. Peterson, R. S. Brown, S. Graff, and J. P. Lynch. 2012. "How Changes in Washington University's Medicare Coordinated Care Demonstration Pilot Ultimately Achieved Savings." *Health Affairs* 31 (6): 1216–26.
- Prochaska, J. O., and C. C. DiClemente. 1983. "Stages and Processes of Self-Change of Smoking: Toward an Integrative Model of Change." *Journal of Consulting and Clinical Psychology* 51 (3): 390–5.
- Schore, J., D. Peikes, G. Peterson, A. Gerolamo, and R. S. Brown. 2011. "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.
- Sun, X., J. P. A. Ioannidis, T. Agoritsas, A. C. Alba, and G. Guyatt. 2014. "How to Use a Subgroup Analysis: Users' Guide to the Medical Literature." *Journal of the American Medical Association* 311 (4): 405–11.
- Zuckerman, R. B., S. H. Sheingold, E. J. Orav, J. Ruhter, and A. M. Epstein. 2016. "Readmissions, Observation, and the Hospital Readmissions Reduction Program." *The New England Journal of Medicine* 374: 1543–51.
- Zurovac, J., G. Peterson, C. Stepanczuk, and R. S. Brown. 2014. "Evaluation of the Medicare Coordinated Care Demonstration: Interim Impact Estimates for the Health Quality." Washington, DC: U.S. Department of Health and Human Services [accessed on April 17, 2016]. Available at https://innovation.cms.gov/files/ reports/medicarecoordinatedcaredemohqp.pdf

# SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Copyright of Health Services Research is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.