

# A Difference-in-Difference Analysis of Changes in Quality, Utilization and Cost Following the Colorado Multi-Payer Patient-Centered Medical Home Pilot

Meredith B. Rosenthal, Ph.D.<sup>1,6</sup>, Shehnaz Alidina, M.P.H., S.D.<sup>1</sup>, Mark W. Friedberg, M.D., M.P.P.<sup>2,3,4</sup>, Sara J. Singer, M.B.A., Ph.D.<sup>1,4,5</sup>, Diana Eastman, B.A.<sup>1</sup>, Zhonghe Li, M.S.<sup>1</sup>, and Eric C. Schneider, M.D., M.Sc.<sup>2,3,4,1</sup>

<sup>1</sup>Harvard T.H. Chan School of Public Health, Boston, MA, USA; <sup>2</sup>RAND Corporation, Santa Monica, CA, USA; <sup>3</sup>Brigham and Women's Hospital, Boston, MA, USA; <sup>4</sup>Harvard Medical School, Boston, MA, USA; <sup>5</sup>Massachusetts General Hospital, Boston, MA, USA; <sup>6</sup>Department of Health Policy and Management, Harvard School of Public Health, Boston, MA, USA.

**BACKGROUND:** Research on the effects of patient-centered medical homes on quality and cost of care is mixed, so further study is needed to understand how and in what contexts they are effective.

**OBJECTIVE:** We aimed to evaluate effects of a multi-payer pilot promoting patient-centered medical home implementation in 15 small and medium-sized primary care groups in Colorado.

**DESIGN:** We conducted difference-in-difference analyses, comparing changes in utilization, costs, and quality between patients attributed to pilot and non-pilot practices.

**PARTICIPANTS:** Approximately 98,000 patients attributed to 15 pilot and 66 comparison practices 2 years before and 3 years after the pilot launch.

**MAIN MEASURES:** Healthcare Effectiveness Data and Information Set (HEDIS) derived measures of diabetes care, cancer screening, utilization, and costs to payers.

**KEY RESULTS:** At the end of two years, we found a statistically significant reduction in emergency department use by 1.4 visits per 1000 member months, or approximately 7.9 % ( $p=0.02$ ). At the end of three years, pilot practices sustained this difference with 1.6 fewer emergency department visits per 1000 member months, or a 9.3 % reduction from baseline ( $p=0.01$ ). Emergency department costs were lower in the pilot practices after two (13.9 % reduction,  $p<0.001$ ) and three years (11.8 % reduction,  $p=0.001$ ). After three years, compared to control practices, primary care visits in the pilot practices decreased significantly (1.5 % reduction,  $p=0.02$ ). The pilot was associated with increased cervical cancer screening after two (12.5 % increase,  $p<0.001$ ) and three years (9.0 % increase,  $p<0.001$ ), but lower rates of HbA1c testing in patients with diabetes (0.7 % reduction at three years,  $p=0.03$ ) and colon cancer screening (21.1 % and 18.1 % at two and three years, respectively,  $p<0.001$ ). For patients with two or more comorbidities, similar patterns of association were found, except that there was also

a reduction in ambulatory care sensitive inpatient admissions (10.3 %;  $p=0.05$ ).

**CONCLUSION:** Our findings suggest that a multi-payer, patient-centered medical home initiative that provides financial and technical support to participating practices can produce sustained reductions in utilization with mixed results on process measures of quality.

**KEY WORDS:** patient-centered care; primary care redesign; health care costs; quality improvement.

J Gen Intern Med 31(3):289–96

DOI: 10.1007/s11606-015-3521-1

© Society of General Internal Medicine 2015

Advocates for primary care reform have promoted patient-centered medical home (PCMH) models as a way to strengthen primary care. PCMH structures and processes, which are intended to help practices deliver patient-centered, proactive, coordinated care can be coupled with financial support and accountability for a defined population of patients.<sup>1</sup> PCMHs are intended to transform the business model of primary care while improving quality of care and reducing costs, largely by reducing emergency department and hospital utilization. Medicare, Medicaid and commercial payers have launched dozens of PCMH pilots across the nation.<sup>2,3</sup>

Two recent systematic reviews found mixed evidence that PCMH initiatives affect utilization, quality, and costs within two years.<sup>4,5</sup> Many studies have found evidence of improved quality based on selected process measures and reduction in emergency department visits, but the evidence for cost savings is limited. A study of Geisinger Health System's PCMH pilot found an 18 % reduction in hospital admissions after two years of a PCMH initiative.<sup>6</sup> Similarly, Group Health of Puget Sound's pilot medical home yielded 6 % fewer hospitalizations compared with other Group Health clinics in the first year and savings of \$10.30 per patient per month in the second year.<sup>7</sup>

While some recent evaluations have not identified significant effects of PCMH initiatives on health care utilization and

---

This research was financially supported by The Commonwealth Fund and The Colorado Trust (PI: Meredith Rosenthal).

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s11606-015-3521-1) contains supplementary material, which is available to authorized users.

---

Published online October 8, 2015

cost,<sup>8–10</sup> the heterogeneity of PCMH interventions and contexts may lead to differential effectiveness, the patterns of which could illuminate facilitators and barriers to effective transformation.<sup>11</sup> In addition, few studies of PCMH pilots have evaluated outcomes beyond two years, even though the literature suggests that practice transformation takes time to unfold.<sup>8</sup> We examined the cumulative effects of a PCMH intervention after two and three years, and endeavored to document details of both the initiative and context to enable comparison with similar studies.

## METHODS

The institutional review board at the Harvard School of Public Health approved the study.

### Study Setting

We examined changes in patient care following initiation of a PCMH pilot in the Front Range area of Colorado, convened and supported by HealthTeamWorks, a non-profit organization that supports practices in implementing continuous quality improvement using a systems approach. The HealthTeamWorks pilot was launched in April 2009 with support from five commercial insurers (Aetna, Anthem, Cigna, Humana, and United HealthCare), one high-risk plan, which predominantly includes patients with preexisting conditions (CoverColorado), and the state Medicaid agency.

The HealthTeamWorks pilot included 15 practices involving 51 physicians, 35 allied health professionals (NPs, PAs) and 205 staff collectively serving approximately 98,000 patients. All 15 pilot sites were small to medium-sized practices (nine or fewer physicians). The National Committee for Quality Assurance (NCQA) certifies PCMHs at three accreditation levels, depending on to what degree that practice meets PCMH benchmarks and capabilities, with Level 3 being the highest level of accreditation.<sup>12</sup> At baseline, NCQA certified three pilot practices as Level-2 PCMH practices and 12 as Level-3 PCMHs.

Participating health plans paid the PCMH practices \$5.3 million over three years in per-member per-month fees based on the level of NCQA accreditation that each practice attained and \$720,000 from a related pay-for-performance program, which awarded bonuses to practices based on meeting both quality and utilization benchmarks. The average payment per practice per year was approximately \$118,000, or \$34,700 per primary care physician. For pay for performance, quality was measured using medical record and registry data for diabetes, depression screening, and tobacco counseling. Utilization targets included reductions in emergency department visits, hospital admissions and increases in generic medication use. The practices received technical assistance, including monthly in-office coaching, learning collaborative sessions three times per year, monthly webinars and innovative technology interventions from HealthTeamWorks outlined in [Appendix Figure 1](#).

One distinguishing feature of the HealthTeamWorks pilot was its deliberate focus on strategies to build “medical neighborhoods” and improve coordination of care. These include efforts such as systems for test and referral tracking to improve information flow, new staff positions to coordinate care, and care compacts with specialists to define shared responsibilities and accountability for coordinated patient care.<sup>13</sup>

### Data Collection

To evaluate changes in patient care after the HealthTeamWorks pilot, we used a difference-in-difference approach. The pre-intervention period began 1 April 2007 and ended 31 March 2009. We examined the post-intervention period at two intervals: two years after initiation of the pilot (1 April 2009 through 31 March 2011) and three years after initiation of the pilot (1 April 2009 through 31 March 2012).

The primary data source for our analysis was administrative claims data from participating health plans. Two health plans (Humana and Cigna) were unable to provide usable claims data. We excluded Colorado Medicaid because of differences in eligibility, and patient population. We hypothesized that the intervention would reduce rates of hospital and emergency department admissions and improve quality of care. We expected that better identification and management of chronically ill patients would result in increased use of prescription drugs. We did not have a fixed hypothesis about primary care and specialist visits because we expected that greater access overall and to more non-visit-based care could yield either positive or negative effects on face-to-face visits with either type of physician. Hospital admissions, emergency department visits, primary care and specialist visits, and prescriptions were measured using standard definitions.<sup>14</sup> To examine ambulatory care sensitive hospital admissions and ER visits, we used previously developed algorithms and used the Elixhauser index to stratify the population on health status.<sup>15–17</sup> We examined changes in quality on the basis of an adaptation of NCQA's Health Effectiveness Data and Information Set (HEDIS) measures for colon, breast, and cervical cancer, as well as three measures of diabetes care quality.

To quantify the change in total costs of care following the HealthTeamWorks PCMH intervention, we took the payer perspective and generated a standardized measure of cost by re-pricing the study claims according to an estimate of the average commercial fees for each service. The pricing data and algorithm were provided under license by OptumHealth.

We replicated all of our claims-data analyses for the subgroup of enrollees with two or more comorbidities identified using the Elixhauser algorithm. These patients represent approximately the top quintile of our overall sample in terms of comorbidities, have more frequent acute care utilizations, and we hypothesized that they could have more to gain from a PCMH intervention.

## Quantitative Analysis

We attributed each patient to the practice that provided the plurality of their primary care and repeated this attribution annually. If a patient was enrolled in a health plan but had no primary care visits in a given year, we used their previous year's primary care practice for attribution. We excluded patients older than 65 years since Medicare coverage, out of concern that we would not observe complete data where Medicare was a secondary payer. To be included in any year of the analysis, patients were required to have a minimum of six months of insurance eligibility in the year.

For each practice in the PCMH pilot, we identified comparison practices in the same geographic region through propensity score matching using the claims data from the baseline period (1 April 1 2007 through 31 March 2009).<sup>17</sup> In the propensity score model, we included the Elixhauser comorbidity index, the number of patients from the high-risk CoverColorado plan, and practice-level average rates of inpatient admissions, emergency department visits, and primary care visits. Matching was based on a caliper width of 0.6 standard deviations of the logit transformation of the propensity score, with a maximum of ten matches for each pilot practice. We selected the caliper width to balance bias reduction through selecting closer matches against the desirability of increasing the number of matches to increase power.<sup>18</sup>

To validate our comparison cohort, we examined differences in practice-level patient characteristics and utilization patterns using Wilcoxon-Mann-Whitney statistics. Our difference-in-difference approach relies on the assumption that comparison practices' performance is on a common trend with the pilot practices; hence, any post-intervention differences in pilot practices' performance relative to that pre-intervention trend and concurrent comparison practice performance are attributable to the pilot program. A fundamental assumption is that pilot and comparison practices' performance would exhibit a similar trend absent the pilot program. Therefore, we further tested for differences in the quarterly trends in our utilization and quality measures before the pilot program.

To adjust for residual confounding after matching, we estimated the effect of the intervention on utilization using generalized estimating equations (GEE) models assuming negative binomial or logistic distributions with a log link, and then adjusted for patient level clustering. The dependent variables in these models were constructed as counts per quarter (per year for quality measures), and the independent variables were patient age, sex, the Elixhauser comorbidity index, an indicator for whether the patient's primary care physician was in the pilot program, number of months of eligibility, a linear time trend to account for secular changes in utilization, an indicator for the post-intervention period, and an interaction between the pilot and post-intervention practices. Using model predictions based on simulating alternative scenarios, we calculated difference-in-difference estimates. We then bootstrapped the difference-in-difference to obtain standard errors.

Bootstrapping was required due to the non-linearity of the model and our interest in the magnitude of the transformed interaction effect.<sup>19</sup> We used a standard two-part GEE model for cost analyses.<sup>20</sup> The two-part model breaks down the estimation into the probability of having any cost and, for those with costs greater than zero, a model of the level of costs. In the first part where the dependent variable is whether there is any cost greater than zero, we used a binomial distribution with a logit link. For the second part, where the dependent variable is the level of cost for those with non-zero values, we used a gamma distribution with log link. For the subgroup of patients with two or more comorbidities, one-part linear models were estimated because the vast majority of observations had costs greater than zero and two-part models failed to converge. All estimates are reported both in terms of the absolute and relative (i.e., percentage) changes in the original units of measurement from pilot baseline levels for all patients and for those with two or more comorbidities separately.  $p < 0.05$  was considered statistically significant.

## RESULTS

### Practice Characteristics

There were no differences between the pilot and the 66 comparison practices on most baseline characteristics with the exception of percent female, emergency department costs, and Hemoglobin A1C (HbA1c) testing for patients with diabetes (Tables 1 & 2). There were no differences in quarterly trends on utilization measures prior to the intervention (data not shown). For lipid testing and breast cancer screening, however, pilot practices were improving less rapidly than comparison practices (data not shown).

### Utilization

Table 3 provides difference-in-difference estimates for utilization and costs for each time period relative to the pre-intervention levels, first for all patients and then for the sub-sample of patients with two or more comorbidities. All estimates should be interpreted as the change in pilot practices' performance relative to the change in comparison practices' performance. For ease of exposition, we describe these differential changes as reductions or increases, but they should be interpreted as changes relative to the trend in comparison practices over the same period.

At the end of two years, HealthTeamWorks pilot practices reduced their patients' use of the emergency department by 1.4 visits per 1000 member months, or approximately 7.9 % ( $p=0.02$ ). At the end of three years, pilot practices sustained this difference with 1.6 fewer emergency department visits per 1000 member months, or a 9.3 % reduction from baseline ( $p=0.01$ ). At three years, there was a reduction of 4.2 primary

**Table 1 Sample Characteristics of Pilot and Comparison Practices in the Baseline Period for the Full Analytic Sample (1 April 2007 – 31 March 2009)**

	Pilot Practices	Comparison Practices	<i>p</i> value of difference
Number of practices	15	66	—
Attributed member months per practice	18,312	13,938	0.31
Average age	44.1	42.4	0.09
% female	51.7	66.5	< 0.001*
Average Elixhauser Index	1.4	1.3	0.23
Average baseline cost per member per month in U.S. dollars			
Total cost of care	436.05	415.02	0.64
Inpatient cost	80.65	68.43	0.36
Emergency department cost	24.11	29.82	0.01*
Average baseline utilization per member per month			
Primary care physician visits	0.27	0.26	0.92
Specialist visits	0.15	0.16	0.38
Inpatient (IP) admissions	0.02	0.02	0.53
Emergency room (ER) visits	0.02	0.01	0.10
Ambulatory care sensitive ER visits	0.002	0.002	0.92
Ambulatory care IP admissions	0.001	0.002	0.55
Prescriptions	1.24	1.14	0.12
Prescription days supply	39.52	36.09	0.21
Average baseline quality (% screened)			
HbA1c testing in patients with diabetes	82.6	85.7	0.09
Lipid testing in patients with diabetes	82.4	74.8	0.12
Dilated eye exams in patients with diabetes	17.5	18.5	0.83
Colon cancer screening	25.7	25.9	0.86
Breast cancer screening	53.9	55.4	0.43
Cervical cancer screening	38.1	47.6	0.13

Note: Measures are summarized to the practice level and *t*-tests are used to test the significance between the comparison and the pilot practices; \*statistically significant at the 0.05 level

care visits per 1000 member months in the pilot practices, or approximately 1.5 % ( $p=0.02$ ). Among patients with two or more comorbidities, there was a reduction from baseline in primary care visits at both two years (a net effect of 12 visits per 1000 member months or a 2.7 % reduction,  $p=0.006$ ) and at three years (a net effect of eight visits per 1000 member months or a 1.8 % reduction,  $p=0.04$ ). For patients with two or more comorbidities, at three years there was a reduction of 0.9 ambulatory care sensitive inpatient admissions per 1000 member months (10.3 %;  $p=0.05$ ).

## Cost

Among all patients, after two years we found a reduction in emergency department costs of \$4.11 per member per month (a 13.9 % reduction,  $p<0.0001$ ). After three years, the reduction from baseline in emergency department costs was sustained at \$3.50 per member per month (a 11.8 % reduction,  $p=0.001$ ) (Table 3). For patients with two or more comorbidities, we found a reduction in emergency department costs of \$11.54 per member per month after two years and \$6.61 per member per month after three years (25.2 % and 14.5 %

**Table 2 Sample Characteristics of Pilot and Comparison Practices in the Baseline Period for Patients with Two or More Comorbidities (1 April 2007 – 31 March 2009)**

	Pilot Practices	Comparison Practices	<i>p</i> value of difference
Number of practices	15	66	—
Attributed member months per practice	3,921	2,397	0.12
Average age	47.0	44.7	0.09
% female	59.2	70.9	0.01*
Average Elixhauser Index	3.0	2.6	0.33
Average baseline cost per member per month in U.S. dollars			
Total cost of care	781.58	791.82	0.88
Inpatient cost	152.29	142.75	0.36
Emergency department cost	47.00	40.08	0.16
Average baseline utilization per member per month			
Primary care physician visits	0.34	0.35	0.96
Specialist visits	0.27	0.31	0.28
Inpatient (IP) admissions	0.03	0.03	0.97
Emergency room (ER) visits	0.03	0.02	0.38
Ambulatory care sensitive ER visits	0.002	0.003	0.24
Ambulatory care sensitive IP admissions	0.003	0.004	0.81
Prescriptions	1.79	1.87	0.47
Prescription days supply	59.19	58.16	0.81
Average baseline quality (% screened)			
HbA1c testing in patients with diabetes	87.9	84.3	0.08
Lipid testing in patients with diabetes	84.8	76.2	0.11
Dilated eye exams in patients with diabetes	19.0	19.8	0.86
Colon cancer screening	19.6	19.5	0.88
Breast cancer screening	45.3	45.8	0.72
Cervical cancer screening	28.4	33.6	0.17

Note: Measures are summarized to the practice level and *t*-tests are used to test the significance between the comparison and the pilot practices; \*statistically significant at the 0.05 level

reduction from baseline, respectively, after two and three years,  $p<0.0001$  and  $p=0.003$ ).

## Clinical Quality

Table 4 presents the results for clinical quality indicators for both the full sample and for patients with two or more comorbidities. Participation in the HealthTeamWorks pilot was associated with improved cervical cancer screening after two years by 4.7 % (a 12.5 % relative improvement;  $p<0.001$ ) and after three years by 3.3 % (a 9.0 % relative improvement,  $p<0.001$ ). Conversely, colon cancer screening declined in intervention practices (a 21.1 % decrease,  $p=0.01$ ). For patients with two or more comorbidities, at the end of two years, there were two improvements: a 6.2 % increase in cervical cancer screening (a 16.2 % relative improvement,  $p<0.001$ ) and a 3.5 % increase in breast cancer screening (a 6.8 % relative improvement,  $p=0.01$ ). After three years, the significant improvements in both areas were smaller but sustained (cervical cancer screening: an increase of 4.4 %, an 11.5 % relative improvement,  $p=0.001$ ; breast cancer: an increase of 2.6 %, a 5.1 % relative improvement,  $p=0.03$ ). In contrast,



Table 3 Changes in Health Care Utilization and Cost in Pilot Practices Relative to Comparison Practices

	Two Years Post-Intervention			Three Years Post-Intervention		
General Population	Effect in Units of Utilization per Member Month (95 % CI)	Effect as a % of Pilot Baseline Relative to Comparison Group	<i>p</i> value	Effect in Units of Utilization per Member Month (95 % CI)	Effect as a % of Pilot Baseline Relative to Comparison Group	<i>p</i> value
<b>Utilization</b>						
Primary care visits	-0.001 (-0.005, 0.003)	-0.4 %	0.62	-0.0042 (-0.008, -0.0006)	-1.5 %	0.02*
Specialist visits	0.0008 (-0.005, 0.007)	0.6 %	0.8	-0.003 (-0.009, 0.004)	-2.0 %	0.38
Emergency room visits	-0.0014 (-0.003, -0.0002)	-7.9 %	0.02*	-0.0016 (-0.003, -0.0004)	-9.3 %	0.01*
Inpatient admissions	-0.0007 (-0.003, 0.002)	-2.6 %	0.94	-0.0008 (-0.003, 0.002)	-3.1 %	0.48
Ambulatory care sensitive emergency room visits	0.00001 (-0.0002, 0.0003)	0.7 %	0.91	0.0001 (-0.0002, 0.0004)	4.6 %	0.51
Ambulatory care sensitive inpatient admissions	-0.0004 (-0.001, 0.0005)	-18.3 %	0.38	-0.0004 (-0.001, 0.0+003)	-18.9 %	0.23
Number of prescriptions	-0.024 (-0.05, 0.002)	-2.3 %	0.07	-0.026 (-0.053, 0.0013)	-2.5 %	0.06
Prescription days of treatment supplied	-0.68 (-1.5, 0.2)	-2.1 %	0.12	-0.7104 (-1.58, 0.16)	-2.2 %	0.11
<b>Cost</b>						
Total cost of care (\$ PMPM)	-4.1 (-19.5, 11.3)	-1.1 %	0.6	-8.03 (-20.7, 5.4)	-2.1 %	0.24
Inpatient costs (\$PMPM)	-2.79 (-11.0, 5.4)	-4.2 %	0.5	-5.05 (-16.3, 4.9)	-7.7 %	0.37
Emergency department cost (\$PMPM)	-4.11 (-6.1, -2.1)	-13.9 %	<.001*	-3.49 (-5.6, -1.4)	-11.8 %	.001*
<b>Patients with two or more comorbidities</b>						
	Effect in Units of Utilization per Member Month (95 % CI)	Effect as a % of Pilot Baseline Relative to Comparison Group	<i>p</i> value	Effect in Units of Utilization per Member Month (95 % CI)	Effect as a % of Pilot Baseline Relative to Comparison Group	<i>p</i> value
<b>Utilization</b>						
Primary care visits	-0.012 (-0.02, -0.003)	-2.7 %	0.006*	-0.008 (-0.015, -0.001)	-1.8 %	0.04*
Specialist visits	-0.005 (-0.01, 0.003)	-1.6 %	0.22	-0.002 (-0.01, 0.009)	-0.8 %	0.68
Emergency room visits	-0.001 (-0.003, 0.0003)	-3.0 %	0.12	-0.001 (-0.003, 0.0004)	-2.6 %	0.14
Inpatient admissions	0.0003 (-0.003, 0.004)	0.3 %	0.88	0.0003 (-0.003, 0.004)	0.3 %	0.85
Ambulatory care sensitive emergency room visits	0.00001 (-0.0004, 0.0005)	0.9 %	0.08	0.0001 (-0.0004, 0.0005)	1.1 %	0.80
Ambulatory care sensitive inpatient admissions	-0.0008 (-0.002, -0.0001)	-8.9 %	0.60	-0.0009 (-0.002, -0.00003)	-10.3 %	0.05*
Number of Prescriptions	-0.01 (-0.05, 0.03)	-0.6 %	0.52	-0.02 (-0.07, 0.02)	-1.2 %	0.28
Prescription days of treatment supplied	0.17 (-0.99, 1.30)	0.3 %	0.78	-0.12 (-1.50, 1.30)	-0.2 %	0.86
<b>Cost</b>						
Total cost of care (\$ PMPM)	-35.49 (-105.97, 35.00)	-5.2 %	0.32	-13.83 (-51.62, 23.91)	-2.0 %	0.47
Inpatient costs (\$ PMPM)	-16.12 (-44.04, 11.79)	-10.4 %	0.26	-9.44 (-27.32, 5.50)	-6.1 %	0.26
Emergency department costs (\$ PMPM)	-11.54 (-17.10, -5.98)	-25.2 %	<.0001*	-6.61 (-11.01, -2.32)	-14.5 %	0.003*

colon cancer screening in patients with two or more comorbidities decreased after two years (a 23.0 % reduction,  $p=0.005$ ) and three years (a 20.3 % reduction,  $p=0.02$ ).

## DISCUSSION

The HealthTeamWorks PCMH pilot was one of the first, and at the time largest in terms of participating plans and practices, multi-payer medical home initiatives in the United States. Our analysis suggests that the HealthTeamWorks PCMH pilot was associated with meaningful reductions in emergency department utilization that were sustained into the third year of the pilot. Holding all other costs equal, the reductions in

emergency department costs we identified translate into nearly \$5 million per year in savings for the approximately 100,000 patients touched by the pilot. Notably, however, we did not find overall cost savings, perhaps because of offsetting increases in other categories of spending. We also found a reduction in ambulatory care sensitive inpatient hospital admissions for patients with two or more comorbidities, suggesting that some PCMH interventions may be able to deliver on the promise to reduce hospital use by patients with chronic illness.

Alongside acute care reductions, we found small but significant reductions in primary care visits. The mechanism that produced these reductions cannot be determined from our data. One possibility that is consistent with our logic model

Table 4 Changes in the Quality of Care in Pilot Practices Relative to Comparison Practices

	Two Years Post-Intervention				Three Years Post-Intervention			
	Odds Ratio	Absolute Change from Baseline (95 % CI)	Effect as a % of Pilot Baseline Relative to Comparison Group	p value	Odds Ratio	Absolute Change from Baseline (95 % CI)	Effect as a % of Pilot Baseline Relative to Comparison Group	p value
General Population								
HbA1c testing for patients with diabetes	0.94	-0.4 % (-0.8 %, 0.05 %)	-0.4 %	0.09	0.92	-0.6 % (-1.1 %, -0.05 %)	-0.7 %	0.03*
Lipid testing for patients with diabetes	1.01	0.4 % (-0.7 %, 1.6 %)	0.5 %	0.44	1	-0.02 % (-1 %, 1 %)	0.0 %	0.97
Dilated eye exams for patients with diabetes	0.99	0.4 % (-0.5 %, 1.3 %)	2.2 %	0.41	1	0.2 % (-0.7 %, 1 %)	0.9 %	0.71
Colon cancer screening	0.85	-4.7 % (-6.5 %, -2.9 %)	-21.1 %	<.001*	0.88	-4.0 % (-5.7 %, -2.4 %)	-18.1 %	< 0.001*
Breast cancer screening	1.02	1.03 % (-0.5 %, 2.5 %)	2.1 %	0.17	1.01	0.6 % (-0.7 %, 1.9 %)	1.3 %	0.35
Cervical cancer screening	1.10	4.7 % (3.5 %, 5.8 %)	12.5 %	<.001*	1.09	3.3 % (2.2 %, 4.5 %)	9.0 %	< 0.001*
Patients with 2 or more comorbidities								
HbA1c testing for patients with diabetes	1.00	-0.2 % (-1.0 %, 0.6 %)	-0.2 %	0.64	0.96	-0.4 % (-1.2 %, 0.5 %)	-0.4 %	0.39
Lipid testing for patients with diabetes	1.00	0.11 % (-2.0 %, 2.2 %)	0.1 %	0.91	0.98	-0.7 % (-2.4 %, 1.1 %)	-0.8 %	0.45
Dilated eye exams for patients with diabetes	1.02	0.1 % (-0.7 %, 1.0 %)	0.5 %	0.81	1.04	0.2 % (-0.6 %, 0.9 %)	0.9 %	0.65
Colon cancer screening	0.88	-4.1 % (-7.0 %, -1.2 %)	-23.0 %	.005*	0.88	-3.6 (-6.7 %, -0.6 %)	-20.3 %	.02*
Breast cancer screening	1.06	3.5 % (0.9 %, 6.2 %)	6.8 %	0.01*	1.05	2.6 % (0.3 %, 4.9 %)	5.1 %	.03*
Cervical cancer screening	1.12	6.2 % (3.3 %, 9.0 %)	16.2 %	<.001*	1.09	4.4 % (1.8 %, 7.0 %)	11.5 %	.001*

Notes: HbA1c is hemoglobin A1c. Quality measures are measured per patient on an annual basis. Only patients with diabetes were included in the measures of HbA1c, lipid testing and eye exams. The denominator for colon cancer screening is limited to patients over 50 years of age. The denominator for breast cancer screening is limited to women over 40 years of age. The denominator for cervical cancer screening is limited to women between the ages of 21 and 64 years of age. For details of colon, breast and cervical cancer screening measures, please see Methods section. Odds ratios measure the increase post-intervention in the odds that a pilot practice patient as compared to a comparison practice patient received the recommended test. Odds ratios greater than one indicate improved quality. \*denotes statistically significant at the 0.05 level

for the PCMH is that telephone consultations, email communication and group visits were provided as substitute for primary care visits.<sup>21</sup> Another element of the PCMH involves more systematic care planning, which could improve the effectiveness of visits, reducing the need for return visits. Future studies should be designed to examine these substitution possibilities, which might offer some hope for more efficient use of an already strained workforce.

This PCMH intervention produced variable effects on measures of clinical quality; however, our study cannot definitively detect what mechanism may have led to this change. It is possible that the effort practices invested to make the changes promoted by the PCMH intervention distracted them from some screening programs. The reductions in the number of face-to-face primary care visits we observed may also have led to fewer opportunities to counsel patients and order screening tests. If a PCMH intervention reduces visits, practices may need to develop other methods for prompting screening.

It is notable that reductions in utilization and costs were observed for the attributed patient population as a whole, as well as for patients with two or more comorbidities. The fact that these benefits were sustained into the third year of the initiative and a new reduction in ambulatory care sensitive admissions accrued at this stage bodes well for the long-term prospects of PCMH interventions in Colorado and elsewhere. With a sample of one PCMH pilot in this study, we cannot tease out why the HealthTeamWorks was able to produce reductions in acute care use and costs when other PCMH pilots have failed to do so. As experience with PCMH initiatives accumulates researchers can contribute to policy and practice by identifying which design features (e.g., on-site coaching) or market characteristics are associated with greater improvements in patient care.

Our analysis is subject to several limitations. First, the HealthTeamWorks intervention was introduced in a small number of practices, limiting our statistical power. We were

unable to obtain complete data from all payers involved, which reduced our sample size and thus power. Second, practices volunteered and were subsequently selected for the pilot, and despite statistical matching, we cannot exclude the possibility of selection bias in our results. Third, our analysis of quality effects is limited in that we did not have clinical data or measures of health outcomes. Fourth, we rely on a quasi-experimental approach to identifying the potential effects of the HealthTeamWorks pilot, and our findings may be sensitive to untestable assumptions about the comparability of our control practices to the HealthTeamWorks practices. Comparisons of baseline demographic and utilization means and pre-intervention trends revealed few significant differences, but our results should be interpreted in light of those differences. Finally, the HealthTeamWorks PCMH pilot was a complex intervention involving new payments, quality incentives, technical assistance and a collaborative approach to quality improvement. Our analysis is unable to identify the extent to which those components or other factors contributed the outcomes we observed.

Interpretation of our results should take into account differences in pilot and matched comparison practices. In the pre-intervention period, pilot practices had significantly lower percentages of female patients in both samples, and for the overall sample, higher per-member per-month emergency department costs. Moreover, prior to the intervention trend tests indicated that lipid testing and breast cancer testing were not improving as rapidly in the pilot practices as in the comparison practices. While levels at baseline were not different, these trend differences could have biased our findings towards the null.

Evaluations of PCMH interventions have produced mixed results on cost and quality measures. Possible explanations for these results include the lack of comparison groups and small sample sizes in some studies, and differences between the design features of PCMH interventions may have contributed to both positive and negative findings in this literature. Our study suggests that a PCMH intervention can produce reductions in acute care utilization and enhance some aspects of the quality of care, while potentially reducing others.

Both in Colorado and nationally, PCMH recognition continues to grow, and a wide variety of payers use PCMH status as a criterion for payment.<sup>22</sup> In addition, some Accountable Care Organizations consider PCMH-model primary care as a foundation for their efforts to manage population health under a global budget.<sup>23</sup> Given the momentum to enhance and extend PCMH models and interventions, evaluating whether populations are also healthier as a result will be an important future endeavor.

**Acknowledgements: Contributors:** Thanks to Marjie Harbrecht, Kari Loken, and the participating practices for providing access and information without which the evaluation would not have been possible. Thanks to the participating health plans for their data.

**Funding:** Funding for this research was provided by the Commonwealth Fund and the Colorado Trust.

**Prior Presentations:** Professor Rosenthal presented preliminary results to pilot stakeholders in November of 2011 and June 2012. Professor Rosenthal presented final results and participated in a policy roundtable at the Colorado Trust in November 2012.

**Conflict of Interest:** Dr. Friedberg has received compensation from the United States Department of Veterans Affairs for consultation related to medical home implementation and research support from the Patient-Centered Outcomes Research Institute via subcontract to the National Committee for Quality Assurance. Dr. Schneider serves as co-chair of NCQA's Committee on Performance Measurement, which reviews and approves quality measures used in PCMH recognition programs. All other authors declare that they do not have a conflict of interest.

**Corresponding Author:** Meredith B. Rosenthal, Ph.D.; Department of Health Policy and Management Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA (e-mail: meredith\_rosenthal@harvard.edu).

## REFERENCES

1. Starfield B, Leiyu S, Macinko J. Contribution of primary care to health systems and health. *Milbank Q*. 2005;83(3):457-502.
2. Bitton A, Martin C, Landon BE. A nationwide survey of patient-centered medical home demonstration projects. *J Gen Intern Med*. 2010;25(6):584-592. doi:10.1007/s11606-010-1262-8.
3. Federal PCMH Activities. <http://www.pcmh.ahrq.gov/page/federal-pcmh-activities>. Accessed September 1, 2015.
4. Jackson GL, Powers BJ, Chatterjee R, et al. The patient-centered medical home: a systematic review. *Ann Intern Med*. 2013;158(3):169-178. doi:10.7326/0003-4819-158-3-201302050-00579.
5. Peikes D, Zutshi A, Genevro JL, Parchman ML, Meyers DS. Early evaluations of the medical home: building on a promising start. *Am J Manag Care*. 2012;18(2):105-116.
6. Gilfillan RJ, Tomcavage J, Rosenthal MB, et al. Value and the medical home: effects of transformed primary care. *Am J Manag Care*. 2010;16(8):607-614.
7. Reid RJ, Coleman K, Johnson EA, et al. The group health medical home at year two: cost savings, higher patient satisfaction, and less burnout for providers. *Health Aff (Millwood)*. 2010;29(5):835-843. doi:10.1377/hlthaff.2010.0158.
8. Rosenthal MB, Friedberg MW, Singer SJ, Eastman D, Li Z, Schneider EC. Effect of a multipayer patient-centered medical home on health care utilization and quality: The Rhode Island chronic care sustainability initiative pilot program. *JAMA Intern Med*. 2013;173(20):1907-1913. doi:10.1001/jamainternmed.2013.10063.
9. Werner RM, Duggan M, Duey K, Zhu J, Stuart EA. The Patient-centered Medical Home: an evaluation of a single private payer demonstration in New Jersey. *Med Care*. 2013;51(6):487-493. doi:10.1097/MLR.0b013e31828d4d29.
10. Friedberg MW, Schneider EC, Rosenthal MB, Volpp KG, Werner RM. Association between participation in a multipayer medical home intervention and changes in quality, utilization, and costs of care. *JAMA*. 2014;311(8):815-825. doi:10.1001/jama.2014.353.
11. Tomoaia-Cotisel A, Scammon DL, Waitzman NJ, et al. Context matters: the experience of 14 research teams in systematically reporting contextual factors important for practice change. *Ann Fam Med*. 2013;11(Suppl 1):S115-S123. doi:10.1370/afm.1549.
12. NCQA (National Committee for Quality Assurance). Standards and Guidelines for NCQA's Patient-Centered Medical Home (PCMH) 2014.
13. Alidina S, Rosenthal M, Schneider E. Collaborating within Medical Neighborhoods: Insights from Patient-Centered Medical Homes in Colorado.
14. National Committee for Quality Assurance. Healthcare Effectiveness Data and Information Set. 2011.
15. Agency for Healthcare Research and Quality. Prevention Quality Indicators. [http://www.qualityindicators.ahrq.gov/Modules/pqi\\_resources.aspx](http://www.qualityindicators.ahrq.gov/Modules/pqi_resources.aspx). Accessed September 1, 2015.
16. The Center for Health and Public Service Research Website. The Center for Health and Public Service Research website. [http://wagner.nyu.edu/files/admissions/acs\\_codes.pdf](http://wagner.nyu.edu/files/admissions/acs_codes.pdf) Accessed September 1, 2015.

17. **Elixhauser A, Steiner C, Harris DR, Coffey RM.** Comorbidity measures for use with administrative data. *Med Care.* 1998;36(1):8–27.
18. **Austin PA.** Comparison of Twelve Algorithms for Matching on the Propensity Score. *Statist Med.* 2014;33:1057–1069.
19. **Karaca-Mandic P, Norton EC, Dowd B.** Interaction terms in non-linear models. *HSR* 47(1), Part I: 255–274.
20. **Buntin M, Zaslavsky A.** Too Much Ado about Two-part Models and Transformation? Comparing Methods of Modeling Medicare Expenditures. *J Health Econ.* 2004;23(3):525–542.
21. **Reid RJ, Fishman PA, Yu O, et al.** Patient-centered medical home demonstration: a prospective, quasi-experimental, before and after evaluation. *Am J Manag Care.* 2009;15(9):e71–e87.
22. **NCQA.** The Future of Patient-Centered Medical Homes.; 2014. Available at: [http://www.ncqa.org/Portals/0/Public%20Policy/2014%20Comment%20Letters/The\\_Future\\_of\\_PCMH.pdf](http://www.ncqa.org/Portals/0/Public%20Policy/2014%20Comment%20Letters/The_Future_of_PCMH.pdf).
23. **Colla CH, Fisher ES.** Beyond PCMHs and Accountable Care Organizations: Payment Reform That Encourages Customized Care. *J Gen Intern Med.* 2014:1–3. doi:10.1007/s11606-014-2928-4.