Payment Reform in Massachusetts: Health Care Spending and Quality in Accountable Care Organizations Four Years into Global Payment

Abstract

Background

The United States health care system faces two fundamental challenges: a high growth rate of health care spending and deficiencies in quality of care. The growth rate of health care spending is the dominant driver of our nation's long-term federal debt, while the inconsistent quality of care hinders the ability of the health care system to maximize value for patients. To address both of these challenges, public and private payers are increasingly changing the way they pay providers—moving away from fee-for-service towards global payment contracts for groups of providers coming together as accountable care organizations. This thesis evaluates the change in health care spending and in quality of care associated with moving to global payment for accountable care organizations in Massachusetts in the first 4 years.

This thesis studies the Blue Cross Blue Shield of Massachusetts Alternative Quality Contract (AQC), a global payment contract that provider organizations in Massachusetts began to enter in 2009. The AQC pays provider organizations a risk-adjusted global budget for the entire continuum of care for a defined population of enrollees insured by Blue Cross Blue Shield of Massachusetts. It also awards substantial pay-for-performance incentives for organizations meeting performance thresholds on quality measures. This work assesses its effect on spending and quality through the first 4 years of the contract.

Methods

Enrollee-level claims data from 2006-2012 were used with a difference-in-differences design to evaluate the changes in spending and quality associated with the Alternative Quality Contract over the first 4 years. The study population consisted of enrollees in Blue Cross Blue

Shield of Massachusetts plans (intervention group) and enrollees in commercial employer-sponsored plans across 5 comparison states (control group).

Unadjusted and adjusted results are reported for each comparison between intervention and control. Changes in spending for all 4 AQC cohorts relative to control were evaluated. In adjusted analyses of spending, I used a multivariate linear model at the enrollee-quarter level, controlling for age, sex, risk score, indicators for intervention, quarters of the study period, the post-intervention period, and the appropriate interactions. For analyses of quality, an analogous model at the enrollee-year level was used. Process and outcome quality were evaluated.

Results

Seven provider organizations joined the AQC in 2009, with a total of 490,167 individuals who were enrolled for at least 1 calendar year in the study period. The control group had 966,813 unique individuals enrolled for at least 1 year during the study period. Average age, sex, and risk scores before and after the AQC were similar between the two groups.

In the 2009 cohort, claims spending grew on average $62.21 per enrollee per quarter less than control over 4 years ($p<0.001$), a 6.8% savings. Analogously, the 2010, 2011, and 2012 cohorts had average savings of 8.8% ($p<0.001$), 9.1% ($p<0.001$), and 5.8% ($p=0.04$), respectively, by the end of 2012. Savings on claims were concentrated in the outpatient facility setting, specifically procedures, imaging, and tests (8.7%, 10.9%, and 9.7%, respectively, $p<0.001$). Organizations with and without risk-contracting experience saw similar average savings of 6.3% and 7.7%, respectively, over 4 years ($p<0.001$). About 40% of savings were explained by lower volume. Pre-intervention trends were not statistically different between intervention and control (-$4.57, $p=0.86$), suggesting savings were not driven by inherently different trajectories of spending. No differences in coding intensity were found. In sensitivity

analyses, estimates were robust to alterations in the model, variables, and sample. Notably, claims savings were exceeded by incentive payments to providers (shared savings and quality bonuses) in 2009-2011, but exceeded incentives payments in 2012, generating net savings.

Improvements in quality among intervention cohorts generally exceeded New England and national comparisons. Quality performance on chronic care measures increased from 79.6% pre-intervention to 84.5% post-intervention in the 2009 cohort, compared to 79.8% to 80.8% for the HEDIS national average, a 3.9 percentage-point relative increase over the 4 years. Analogously, preventive care and pediatric care measures increased 2.7 and 2.4 percentage points relative to control, respectively. On outcome measures, achievement of hemoglobin A1c, LDL cholesterol, and blood pressure control grew by 2.1 percentage points per year in the 2009 cohort after the AQC, while HEDIS averages remained largely unchanged (Figure).

Conclusion

After 4 years, physician organizations in the AQC had lower spending growth relative to control and generally outperformed national averages on quality measures. Shared savings coupled with quality bonuses can exceed savings on claims in initial years, but over time, savings on claims may outgrow incentive payments. Incentive payments themselves may serve meaningful purposes, as quality measures may protect against stinting and shared savings may help ease providers into risk contracts. Changes in utilization suggest that this payment model can help modify underlying care patterns, a likely prerequisite for sustainable reform. The AQC experience may be useful to policymakers, insurers, and providers embarking on payment reform. Combining global budgets with pay-for- performance may encourage organizations to embark on the delivery system reforms necessary to slow spending and improve quality.

Table of Contents

Glossary of Abbreviations

ACA – Affordable Care Act

ACO – Accountable care organization

AQC – Alternative Quality Contract (by Blue Cross Blue Shield of Massachusetts)

BCBSMA – Blue Cross Blue Shield of Massachusetts

CMMI – Center for Medicare and Medicaid Innovation

CMS – Center for Medicare and Medicaid Services

DxCG – Diagnostic Cost Groups (trademark software of Verisk Health)

HbA1c – Hemoglobin A1c

HMO – Health maintenance organization (a type of insurance plan)

HEDIS – Healthcare Effectiveness Data and Information Set

LDL – Low density lipoprotein cholesterol

P4P – Pay-for-performance

PCP – Primary care physician

PCMH – Patient-centered medical home

1. INTRODUCTION

The United States health care system faces two significant challenges: an unsustainable growth rate of health care spending and deficiencies in quality of care. The growth rate of health care spending is the dominant driver of our nation's long-term federal debt.[1,2,3,4] Meanwhile, the poor quality of care in certain parts of the health care system continues to put patients at risk.

To address both of these challenges, public and private payers are changing the way that physicians and hospitals are paid. Increasingly, payers are moving away from fee-for-service and adopting global payment contracts for groups of providers working together as accountable care organizations (ACOs).[5] The ACO concept has garnered policy interest in recent years.[6,7] ACOs are discussed as a payment and delivery model with the potential to slow spending and improve quality by giving providers incentives—such as through a global payment—to reduce wasteful utilization, improve care coordination, and manage population health.[8,9] However, despite the optimism generated by ACOs, evidence of its effectiveness in lowering spending and improving quality remain limited to date.

This thesis evaluates the changes in health care spending and quality of care associated with moving to global payment for accountable care organizations in Massachusetts. It studies the Blue Cross Blue Shield of Massachusetts Alternative Quality Contract and reports the results on spending and quality in the first 4 years of the contract.

1.1. Accountable Care Organizations

An ACO is a group of providers that accepts joint accountability for health care spending and quality for a defined population of patients.[10] Three parts of this definition deserve particular

emphasis. First, the ACO is a *group* of providers, which often includes physicians, allied health professionals, hospitals, post-acute facilities, and other providers or facilities.

Second, providers in the ACO have *joint accountability* for spending and quality. Joint accountability for spending is manifested through a spending target for the entire organization for a year, whereby the ACO shares savings with the payer if spending falls below the target and may share financial risk with the payer of spending exceeding the target. Joint accountability for quality is implemented via a set of quality measures, which can include process measures (such as annual eye exams for patients with diabetes), outcome measures (hemoglobin A1c (HbA1c) in patients with diabetes), patient experience measures, and others. ACOs are rewarded for their performance on quality measures, usually measured at the end of a year. These accountabilities for spending and quality can be further tied together; for example, an ACO might only receive quality measures if a minimum savings is achieved.

Third, an ACO is responsible for the spending and quality of a *defined population*. The population may be defined prospectively, meaning that going into a year, the ACO is assigned a pre-defined population of patients in a geographic area. This enables the ACO to know exactly which patients it is responsible for during the year. Alternatively, the population may be defined retrospectively, whereby at the end of a year, the ACO assumes responsibility for spending and quality for any patients who received the plurality of their care through the ACO. Retrospective assignment does not allow an ACO to know exactly who it is responsible at the beginning of the year, although the organization may have a stable patient population through a given payer.

The Affordable Care Act (ACA) of 2010 included provisions for the creation of ACOs in the Medicare program, which began in 2012.[11] Through the Medicare Shared Savings program, provider groups can become ACOs by choosing a "one-sided" model in which they share savings with Medicare or a "two-sided" model in which they share savings and risk. Both models also

reward quality via reporting and performance on a set of 33 quality measures. In January 2012, 32 advanced organizations across the country began "Pioneer" ACO contracts through the Center for Medicare and Medicaid Innovation (CMMI), part of the Center for Medicare and Medicaid Services (CMS). Compared to ACOs in the Shared Savings Program, the Pioneer ACOs took on greater risk and greater reward in a more robust two-sided model.

Since then, 4 waves of Shared Savings ACOs have been rolled out by CMS (Figure 1). The first comprised of 27 ACOs serving about 375,000 Medicare beneficiaries starting their contracts in April, 2012; the second 89 ACOs serving 1.2 million beneficiaries starting in July, 2012; the third 106 ACOs serving 1.6 million beneficiaries starting in January, 2013; the fourth 123 ACOs serving 1.5 million beneficiaries announced most recently in December, 2013. The Shared Savings ACOs are primarily in a one-sided contract, although there are several ACOs in the Shared Savings program that are in two-sided contracts.

Outside of Medicare, provider organizations are increasing entering ACO contracts with private insurers. Early estimates in 2011 suggested that about 100 provider organizations across the country were working with private insurers to implement ACO-type contracts.[5] Most recent estimates suggest that number has increased, and that combined with public sector ACOs in the Medicare program, about 18 million covered lives in the U.S. are now under an ACO contract.[12]

1.1. Payment reform in Massachusetts

Massachusetts was an early adopter of global payment and ACOs in both the private and public sectors.[13] After expanding insurance coverage to the uninsured in 2006, Massachusetts set off to combat the unsustainable growth in health care spending.[14] A series of events provided an opportunity for stakeholders in the state to debate the merits and pitfalls of policies aimed at cost

control (Figure 2). A special state commission in 2009 voted unanimously to transition the state away from fee-for-service to global payment with 5 years. The Office of the Attorney General published several widely-read reports concerning payment differences and market power across Massachusetts providers in the ensuing years. Governor Deval Patrick made cost control a core piece of his legislative agenda; meanwhile the state legislature also took up cost control in a few prominent proposals. This effort by state government, working in conjunction with stakeholders from across the state, eventually led to a health care cost containment bill (Chapter 224 of the Acts of 2012) that Governor Patrick signed into law in August 6, 2012. The bill stipulates that health care spending cannot grow faster than the state's economy through 2017.

In the context of this broader state movement towards payment reform, Blue Cross Blue Shield of Massachusetts (BCBSMA), the state's largest private insurer, began a global payment initiative in 2009 in the form of the Alternative Quality Contract. The contract, described below in more detail, is a two-sided ACO agreement that pays physician organizations a global budget to take care of the entire continuum of care for a population of enrollees. Seven Massachusetts physician organizations joined the contract in the first year. Since then, additional organizations have entered the contract, with 4 organizations joining in 2010, 1 in 2011, and 6 in 2012 (Figure 2). By the end of 2012, approximately 85 percent of physicians in the BCBSMA network were in organizations that had joined the AQC.

In recent years, other private insurers in Massachusetts have also moved away from fee-for-service towards global payment contracts. For example, in 2011 Tufts Health Plan launched its Coordinated Care Model, which transitions physician organizations to risk-based global payment contracts with incentives for quality performance. By the end of 2012, Tufts Health Plan had approximately 90 percent of their Medicare Advantage health maintenance organization members in risk contracts and 72 percent of commercial enrollees in risk contracts.[15] Similar to

the Alternative Quality Contract, Tufts Health Plan provides physician organizations frequent and customized data analysis on their utilization rates, referral patterns, and spending.

When the Pioneer ACO program launched in January, 2012, 5 physician organizations in Massachusetts were part of the initial cohort of 32 ACOs. These organizations included Atrius Health, Beth Israel Deaconess Physician Organization, Mount Auburn Cambridge Independent Practice Association, Partners Healthcare, and Steward Health Care. Massachusetts has also seen a number of additional physician organizations join the Medicare Shared Savings Program in the subsequent years.

1.2. Alternative Quality Contract

The Alternative Quality Contract (AQC) is a global payment contract that pays physician organizations a risk-adjusted budget to take care of defined population of enrollees. The contract has several unique features. First, it is a multi-year (most often 5-year) ACO contracting model in which the global payment covers the entire continuum of care, including inpatient, outpatient, and post-acute care for patients assigned to the physician organization. Notably, the organization is accountable for all medical spending incurred by enrollees in the contract. Care received in the ACO to which their primary care physicians belong and received elsewhere are similarly counted towards the home organization's budget. The size of the budget and the growth rate of the budget are determined through negotiations between BCBSMA and the organization. In 2009, all ACOs initially entering the AQC received budgets exceeding their 2008 spending. Throughout the year, claims are added up on a fee-for-service basis using negotiated prices.

Second, the AQC was initially launched within health maintenance organization (HMO) plans in the BCBSMA network. Enrollees in HMO plans are required to designate a primary care

physician (PCP) at the beginning of each year. This allows for attribution of patients to an ACO through the patient's PCP, who effectively manages the enrollee's budget on behalf of the PCP's organization. This feature is similar to many patient-centered medical home (PCMH) models, in which each patient is linked to a PCP responsible for coordinating the patient's care.[16,17,18,19]

Third, the AQC awards quality bonuses up to 10 percent of an organization's global budget through performance on quality measures. The quality measures are shown in Appendix 1. The AQC uses 64 measures, half of them focus on ambulatory quality and the other half on inpatient quality of care. At the end of each year, an organization's performance on the measures is summed to an aggregate, which is used to determine the percent of an organization's budget it will receive as a bonus payment. BCBSMA defines 5 aggregate performance thresholds; the top threshold rewards organizations the full 10 percent bonus. Measures are weighted differently; for example, patient experience measures are triple-weighted relative to most measures (Appendix 1). The size of the AQC quality bonus is larger than typical U.S. pay-for-performance contracts.

Fourth, BCBSMA provides AQC organizations technical support to help them achieve their cost and quality goals. For example, organizations receive spending and quality reports on a regular basis, which allow them to compare performance relative to peer organizations in the state. Utilization rates compared to peer organizations for standardized patients are also reported to the AQC organizations in order to help them identify areas of potential overuse.

1.3. State of the Field

In the first year, the AQC was associated with a 1.9 percent reduction in medical spending and modest quality improvements.[20] The initial savings were largely achieved through lower prices rather than reductions in the volume of services. This finding was consistent with

AQC groups' initial focus of shifting referrals to less expensive providers.[21] Moreover, the reduction in total medical spending associated with the AQC was likely smaller than the sum of shared savings, payments for quality bonuses, and payments for infrastructure support that BCBSMA provided to organizations. In the second year, the AQC was associated with a larger reduction in medical spending of 3.3 percent relative to control and continued improvements in quality of care.[22] In addition to achieving savings through referring patients to lower-priced providers, organizations in year 2 also lowered spending through decreases in utilization, as documented in this thesis.

In the first 2 years, the AQC was also found to be associated with a decrease in spending for Medicare beneficiaries.[23] These savings were concentrated in similar services and settings as those for enrollees in the AQC, suggesting that the impact of one payment contract may lead to "spillover" effects for other populations not covered by the contract when organizations have a multi-payer patient mix. Further work on AQC effects for the pediatric population show that the contract did not meaningfully impact spending in the first two years but improved the quality of care for children, especially for children with special health care needs.[24]

Year-1 results for Medicare Pioneer ACOs were recently released by CMS on July 16, 2013. Pioneer ACOs generated a gross savings of $87.6 million largely through reductions in admissions and readmissions, of which $33 million went to the Medicare Trust Fund.[25] These savings came from 13 of the 32 organizations, while 17 organizations did not spend significantly more or less than their spending targets and 2 organizations faced financial losses totaling about $4 million. All Pioneer ACOs were rewarded for reporting on quality measures, whose bonuses were not tied to actual performance on quality in the first year. Nevertheless, Pioneer ACOs did better on blood pressure and cholesterol control for beneficiaries with diabetes than did managed-care plans, and better on readmissions relative to the Medicare fee-for-service

benchmark. After the first year, 9 Pioneer ACOs left the program while 23 remained, suggesting that while the year-1 results were encouraging, substantial challenges are faced by organizations transitioning into ACO contracts, especially two-sided contracts that include financial risk.[26]

Interim year-1 results for the Shares Savings program were released on January 30, 2014. According to CMS, 54 out of 114 Shared Savings ACOs that started in 2012 achieved spending lower than their expected levels.[27] Of these, 29 ACOs generated savings over $126 million, with a total net savings of $128 million to the Medicare Trust Fund.

A more complete analysis of the characteristics of early ACOs in the Medicare program demonstrated that patients in ACOs tended to be older, have higher incomes, were less likely to be covered by Medicaid, and less likely to be African American or disabled compared to those in non-ACOs.[28] Cost of care was lower for patients in ACOs. Moreover, hospitals that were a part of ACOs were more likely to be large academic medical centers, and hospitals in ACOs did not have measurably higher performance on quality measures than non-ACO hospitals. Additional research on the link between organizational characteristics and spending demonstrated that large independent physician organizations with a strong primary care orientation tended to have lower spending and higher quality measures for Medicare beneficiaries.[29]

1.4. Purpose of Inquiry

To date, little evidence exists regarding the effect of global payment in ACOs on health care spending and quality beyond the first year or two. This thesis evaluates changes in spending and quality associated with the AQC in Massachusetts through the first 4 years of the contract. Understanding the potential implications of moving to global payment beyond the initial years

may help inform physician organizations as they increasingly enter ACO contracts. Evidence on longer-term AQC effects may also inform policymakers and other states.

In compliance with the Harvard Medical School thesis guidelines, year-1 results of the AQC are not discussed in this document because they were a part of my dissertation in the PhD Program in Health Policy. Year-2 results through the most recent year-4 results are included in this thesis for Harvard Medical School.

2. METHODS

2.1. Population

The study population consisted of enrollees in BCBSMA (intervention group) and commercial employer-sponsored plans in comparison states (control group) from January 2006 through December 2012. All enrollees were under the age of 65 and continuously enrolled in a plan for at least one calendar year during the study period. The AQC was implemented in BCBSMA health maintenance organization (HMO) plans, in which enrollees are required to designate a PCP.

The 2009 intervention cohort comprised BCBSMA enrollees whose PCPs were affiliated with an organization that entered the AQC in 2009. Within the 2009 cohort, 2 subgroups pre-specified. The "prior-risk" subgroup comprised 4 organizations that had prior experience managing risk-based contracts from BCBC, while the "no-prior-risk" subgroup comprised 3 organizations that entered the AQC without BCBS risk-contracting experience. Provider organizations in the prior-risk subgroup tended to be larger, more established delivery systems, while those in the no-prior-risk subgroup were smaller organizations, including physician groups that were newly formed prior to entering the AQC.

The control group consisted of enrollees whose PCPs belonged to an organization that did not enter the AQC. For analyses of year-1 (2009) and year-2 (2010) AQC effects in the 2009 cohort, control subjects were BCBS enrollees whose PCPs' organizations had not joined the AQC by the end of year 2 (2010). However, by year 4 (2012) about 85 percent of physicians in Massachusetts had joined the AQC, including the large majority of those serving as the control group in the year-1 and year-2 analyses. The remaining physicians not in the AQC by 2012 were

not a representative sample; they were largely very small practices and were receiving different percent payment updates from BCBS as a result of not being in an incentive contract. Thus, the year-3 and year-4 analyses required a control group from outside of Massachusetts. The solution was to obtain a pooled cohort of similar enrollees from comparison states.

The control group comprised commercially insured individuals in employer-sponsored plans across all 8 other Northeastern states (Connecticut, Maine, New Jersey, New Hampshire, New York, Pennsylvania, Rhode Island, and Vermont). All individuals were under age 65 and continuously enrolled in a health maintenance organization or point-of-service plan for at least 1 year between 2006 and 2012. Importantly, these plan types have physician networks and require enrollees to designate a PCP, similar to the AQC. All belonged to employers that continuously reported claims for the 7 years in the study period. Compared to Massachusetts, these states have similar demographic profiles; they have both large academic health systems as well as smaller community providers; and they have competition among multiple commercial insurers. The main identification strategy comes from the fact that none of these control states had a broad-scale payment reform implemented by a large commercial insurer that moved provider organizations from fee-for-service to global payment in the study period. Some states, such as Rhode Island, began piloting patient-centered medical home models in the later years of our study period. Do date, these interventions have not been shown to significantly impact spending.

While the 2009 AQC cohort comprised the main analyses of this work, the AQC continued to expand in subsequent years. Physician organizations that entered the AQC in 2010, 2011, and 2012 separately comprised different intervention cohorts in each of those years. By the end of 2012, the year-1, year-2, and year-3 changes in spending and quality associated with the AQC could be calculated for the 2010 cohort. Analogously, the year-1 and year-2 changes could be calculated for the 2011 cohort, and the lone year-1 analysis could be carried out on the 2012

cohort. I carried out these subsequent analyses for the AQC cohorts after 2009 and highlight the major findings here. However, this research focuses on the 2009 cohort, for which results up to year 4 are available.

2.2. Data

For both intervention and control groups, I used de-identified enrollee-level claims and membership data from 2006-2012. Data for BCBSMA enrollees consisted of outpatient claims, inpatient claims, and prescription pharmaceutical claims, as well as enrollment data linked to provider information. Provider data was also de-identified, and included linkages to the provider organization in the BCBSMA network. Data for control subjects were drawn from the 2006-2012 Truven (formerly Thomson Reuters MarketScan) Commercial Claims and Encounters database. This database is derived from a large convenience sample of commercially insured individuals in employer-sponsored plans offered by large firms throughout the United States. The data has been used extensively for research purposes.[30] The employers report all outpatient, inpatient, and prescription pharmaceutical claims, along with member enrollment and provider information, similar to the BCBSMA data. However, provider information does not contain linkages to any specific organizations. I also cannot identify specific insurers or link enrollees to insurers.

I used the DxCG (Diagnostic Cost Groups) software by Verisk Health to generate risk scores for each enrollee in each year among both the intervention and control group. The DxCG software uses age, sex, enrollment, and diagnostic information in the form of International Classification of Diseases (ICD) codes on the claims to generate an annual risk score at the enrollee level.[31] It is used by most private insurers in the U.S. as a risk-adjustment method and was based on foundational work using Medicare data.[32] Statistical models in the DxCG software

are capable of generating concurrent and prospective risk scores. Consistent with prior work on the AQC, I generated a concurrent risk score, which correlates a given year's total medical spending to the diagnoses and demographic information from the concurrent year. Higher risk scores represent higher expected spending on an annual basis.

Data on quality included ambulatory process measures, separated into chronic care management measures, adult preventive measures, and pediatric measures. The process measures were available on an annual enrollee level in the BCBS data from 2007-2010. Therefore, for the 2009 cohort, year-1 and year-2 changes in quality associated with the AQC could be analyzed using statistical models similar to analyses of spending, because the control group in the first 2 years was BCBS enrollees whose PCPs belonged to organizations that did not enter the AQC. In the analysis beyond 2 years, however, the control group was the comparison states as described above, drawn from the Truven data, which does not include quality measures. Thus, comparison quality data for the further longitudinal analysis into years 3 and 4 were obtained from the Healthcare Effectiveness Data and Information Set (HEDIS). I used HEDIS national averages as the comparison for enrollees in the AQC.

Quality data also included 5 outcome measures: hemoglobin A1c (HbA1c) control for patients with diabetes (≤9 percent), low density liproprotein (LDL) cholesterol control for patients with diabetes (<100 mg/dL), and blood pressure control for patients with diabetics (<140/80 mmHg), as well as blood pressure control for patients with cardiovascular disease (<140/90 mmHg), and LDL control for patients with cardiovascular disease (<100 mg/dL). Enrollee-level tracking of outcome quality measures was not implemented prior to the AQC. Therefore, the analysis of outcome measures was done by calculating average performance in each year for the entire 2009 AQC cohort. I compared outcomes among AQC enrollees to the HEDIS national averages.

2.3. Study Design

This analysis used a difference-in-difference approach to isolate the changes in health care spending and quality associated with the AQC intervention. The difference-in-differences approach is an economic method for program evaluation when data on treatment and control groups are available both before and after the intervention. It centers on calculating the change in the outcome from before to after the intervention in both the intervention and the control group. The change in the control group is then subtracted from the change in the intervention group to remove the effect of unobserved variables influencing both groups and to isolate the treatment effect associated with the intervention.[33]

The primary analysis consists of the 2009 AQC cohort and control. The pre-intervention period was 2006-2008 and the post-intervention period was 2009-2012. In the spending analysis, the dependent variable was the sum of inpatient and outpatient medical spending (including patient cost sharing). Spending was calculated from claims payments made by the insurer to providers, which reflects negotiated fee-for-service prices. Pharmaceutical claims were added to spending in the sensitivity analyses; they were excluded from the main model because not all enrollees had pharmaceutical benefits administered by their primary insurer. Pharmaceutical benefits may be carved out, in which case I did not observe drug claims or incentives associated with drug formularies or implemented by drug benefit managers. All spending was adjusted to 2012 U.S. dollars.

To analyze changes in spending by type of care (facility versus professional) and site of care (inpatient versus outpatient), I decomposed spending along these two dimensions. To also analyze results by categories of medical services, I assigned spending according to categories of

care using the Berenson-Eggers Type of Service (BETOS) classification system from the Centers for Medicare and Medicaid Services.[34] Finally, utilization of services was computed by counting actual numbers of services delivered to patients in the claims. For ease of interpretation, I scaled utilization data to volume per thousand enrollees.

2.4. Statistical Analysis

An enrollee-level linear model was used, controlling for age, sex, risk score, indicator for AQC, year indicators, and interactions between AQC and year. These interactions estimated changes attributable to the AQC, but should be interpreted while recognizing other factors in Massachusetts may have influenced spending or quality. Dollar results for each year were scaled into percentages by dividing by the current year claims costs. State and plan fixed effects were also included to account for such time-invariant characteristics between individuals. Consistent with prior work on the AQC, this model was not logarithmically transformed because the DxCG risk score is designed to predict dollar spending and linear models have been shown to better predict health spending than more complex functional forms. Especially with large sample sizes, linear models generally outperform other statistical models at estimating population averages, even though they can be less precise at estimating the tails of a spending distribution.[35 36 37 38] Results are reported with 2-tailed P values.

2.4.1. Model for Spending

The reduced-form model is shown below in equation (1). In this model, $\boldsymbol{X}_{it}$ includes a vector of enrollee characteristics, including age categories, age-sex interactions, and the DxCG

risk score. The indicator variable aqc_i denotes assignment of each individual i to the treatment or control group. The vector of year indicators is represented by $\boldsymbol{y}_t$, and the AQC-year interactions are represented by $\boldsymbol{y}_t$*aqc, which produce the coefficients of interest. To account for multiple observations within individuals, Huber-White corrections were used to adjust standard errors clustered at the plan level.[39,40]

$$Spending_{it} = \alpha_{it} + \boldsymbol{X}_{it}\delta + \beta_1 \boldsymbol{y}_t + \beta_2 aqc_i + (\boldsymbol{y}_t * aqc_i)\gamma + \varepsilon_{it} \quad (1)$$

In difference-in-differences models, identification of the policy effect on the outcome of interest relies on similar pre-intervention trends. For each AQC cohort, I tested for differences in pre-intervention trends in spending between the intervention and control groups.

Given that spending is the product of price and quantity, any policy intervention that is associated with a change in spending must be associated with a change in prices or a change in quantities. I assessed the relative contributions of price and quantity to the spending results by standardizing the prices for each service to its median price across all providers in 2006-2012. Reanalyzing the model with the standardized prices, differences in spending associated with the AQC reflect differences in utilization. Furthermore, I assessed whether the price effect was due to differential changes in negotiated fees or differential changes in referral patterns (referring patients to less expensive physicians or hospitals). I used models of utilization to directly analyze the relationship between the AQC and quantity of specific services.

2.4.2. Model for Quality

The association between the AQC and changes in quality measures in the first 2 years was studied using an analogous difference-in-difference model. I pooled BCBSMA process measures into their categories for aggregate analysis: chronic care management, adult preventive care, and pediatric care. I also analyzed separate models for each individual measure. Quality measures are calculated on an annual basis for each enrollee. Therefore, each observation in this model is at the enrollee-year level.

In the analysis of aggregate measures, I included measure-level fixed effects. Therefore, the results are interpreted as average changes *within* measures associated with the AQC. This is included because different measures have different baseline levels of achievement. In sensitivity analyses, models without measure-level fixed effects were analyzed. Process quality measures were available at the enrollee level from 2007 to 2012. Thus, this statistical model was used to analyze the year-1 and year-2 quality results for the 2009 AQC cohort. As mentioned above, the year-3 and year-4 quality analyses were done descriptively, by comparing the 2009 AQC cohort averages to the HEDIS national averages.

All analyses of outcome quality measures were conducted descriptively. The unadjusted percentage of the 2009 AQC cohort achieving quality performance on the 5 measures related to blood pressure, LDL control, and HbA1c were calculated by year. This was compared to HEDIS national averages, consistent with process quality analyses after the first 2 years.

2.4.3. Sensitivity Analyses

To test the robustness of the statistical model, I conducted a series of sensitivity analyses. These included alterations to the statistical model as well as to variables and sample. Alterations in the statistical model included omitting state or plan fixed effects, covariates, risk score, and

substituting percent cost-sharing in place of plan fixed effects. Alterations to variables or sample included analyzing only enrollees who were continuously enrolled in the study period, defining the risk score as a categorical variable using deciles, omitting cost-sharing from spending, adding pharmaceutical claims to spending, lagging the prospective risk score, using HMO controls only, and using both within-Massachusetts and national controls. For the analysis of quality, I used a logit model in sensitivity analyses in place of the linear probability model.

In a global budget payment system, another concern is the possibility of coding behavior changes that may lead to differences in spending adjusted for risk. For example, if organizations code at a higher intensity in a given year, this may garner a larger global payment in the future if spending in the given year is used to calculate future spending targets. An increase in the coding of AQC patients would make them appear sicker and make spending adjusted for risk score seem lower. Prior work showed that any risk score changes associated with the AQC explained only a small portion of spending differences. I repeated this analysis through the first 4 years by putting the risk score as the dependent variable in the model. This issue has been previously discussed in the evaluation of the Medicare Physician Group Practice Demonstration.[41]

All analyses were carried out using STATA software, version 13. The Harvard Medical School Office for Research Subject Protection approved this study protocol.

3. RESULTS

3.1. Population

Characteristics of the 4 AQC cohorts and control group are shown in Table 1. Enrollees in the AQC had an average age of approximately 35 years, and the population was about 51 percent female. Average DxCG risk scores for the cohorts ranged from 1.03 to 1.05 with similar distributions. Enrollee cost sharing average between 11 and 14 percent across cohorts, also with similar distributions. Across the study period, the 2009 AQC cohort comprised 490,167 unique enrollees who were enrolled for at least 1 calendar year. These enrollees designated one of about 1,100 PCPs practicing across 7 provider organizations, which comprised over 2,000 specialist physicians. Other cohorts varied in the number of enrollees, PCPs, and specialists (Table 1).

Characteristics of the control group were largely similar. Pre- and post-intervention comparisons between each cohort and control are shown in Table 2. There were 966,813 unique individuals enrolled for at least 1 year in the control group during the study period. Average age, sex, and risk score before and after the AQC were similar between the two groups. The control had a higher average cost-sharing percentage compared to the AQC cohorts.

3.2. Spending

3.2.1. 2009 AQC Cohort

Figure 3 shows the unadjusted spending trends for the 2009 AQC cohort and control. In unadjusted analysis, the 2009 AQC cohort spent on average $789.35 per enrollee per quarter in

the pre-intervention period (2006-2008) and $913.15 in the post-intervention period (2009-2012) for a difference of $123.80 per enrollee per quarter, while the control group spent $731.61 in the pre-intervention period and $911.40 post-intervention for a difference of $179.79 per enrollee pre quarter. The unadjusted difference between the changes (the difference-in-difference result) was -$55.99 per enrollee per quarter (Table 3). This suggests the 2009 AQC cohort experienced, on average, a decrease in spending during the 4 years after the intervention compared to before the intervention, relative to what the control group experienced.

In adjusted analysis using the multivariate regression, the AQC was associated with an average 4-year change in spending of -$62.21 per enrollee per quarter, representing a 6.8 percent decrease ($p<0.001$) in the average level of spending compared to the pre-intervention level, relative to the control group. This represents the statistical estimate of the policy effect over the first 4 years. Pre-intervention trends were not statistically different between the AQC and control group (-$4.57, $p=0.86$), suggesting that differences in post-intervention spending were not driven by inherently different trajectories of spending. This was robust to the inclusion or exclusion of covariates in the model. No significant changes in the DxCG risk score were associated with the AQC (-0.0015, $p=0.57$), suggesting that coding behavior did not meaningfully impact the results.

Figure 4 decomposes average spending by site and type of care: inpatient facility and professional as well as outpatient facility and professional. This unadjusted analysis suggests that the slowing of spending in the 2009 cohort was most pronounced in the outpatient setting rather than the inpatient setting. Within outpatient spending, facility spending accounted for the largest raw decrease relative to control, as the two trends intersect each other in late 2011. In contrast, trends in inpatient spending were similar between AQC and control in the raw plots (Figure 4).

In adjusted analysis, decomposition of average 4-year spending by site and type of care similarly showed that changes in spending were largest in the outpatient facility setting (-$48.67

per enrollee per quarter, p<0.001). The decrease in inpatient facility spending was not statistically significant (-$3.32 per enrollee per quarter, p=0.52). The decrease in outpatient professional spending was -$15.35 per enrollee per quarter, p=0.004). Inpatient professional spending did not incur a statistically significant change on average in the first 4 years of the AQC ($0.40 per enrollee per quarter, p=0.82) (Table 3).

Unadjusted decomposition of the 2009 AQC cohort by prior risk contracting experience is illustrated in Figure 5. In adjusted analysis, the Prior-Risk subgroup, which comprised about 88 percent of the cohort, had an average 4-year change in spending of -$57.61 per enrollee per quarter (-6.3 percent, p<0.001), while the No-Prior-Risk subgroup saw a change of -$68.66 (-7.7 percent, p<0.001). For both subgroups, the outpatient facility setting accounted for the largest decreases in spending (p<0.001) (Table 3). Consistent with the aggregate results above, inpatient professional spending did not change significantly for either subgroup. The Prior-Risk subgroup saw an insignificant change in inpatient facility spending of -$3.28 per enrollee per quarter over the 4 years (p=0.52), as did the No-Prior-Risk subgroup ($5.26 per enrollee per quarter, p=0.58). Both subgroups had significant decreases in outpatient professional spending of -$11.87 (p=0.02) and -$23.13 (p<0.001) over the 4 years, respectively (Table 3).

3.2.2. 2010 AQC Cohort

Figure 6 illustrates the unadjusted spending trends for the 2010 AQC cohort and control. In unadjusted analysis, the 2010 AQC cohort spent on average $876.42 per enrollee per quarter in the pre-intervention period (2006-2009) and $954.74 in the post-intervention period (2010-2012) for a difference of $78.32 per enrollee per quarter, while the control group spent $772.71 in the same pre-intervention period and $919.45 post-intervention for a difference of $146.74 per

enrollee pre quarter. The difference-in-difference change in spending associated with the AQC was -$68.42 per enrollee per quarter (Table 4).

Similar to the 2009 AQC cohort (Figure 3), the 2010 AQC cohort also demonstrated a large decline in spending after the intervention (in this case, 2010-2012) relative to control. This decline in spending appeared more similar to that of the No-Prior-Risk subgroup in the 2009 cohort, which is consistent with the fact that the 2010 AQC cohort is entirely comprised of physician groups that joined the AQC from fee-for-service contracts. In essence, the 2010 AQC cohort is a No-Prior-Risk cohort. Thus, the most analogous comparison between the 2009 and 2010 cohorts comes from using the No-Prior-Risk subgroup in the 2009 cohort (Figure 5).

In adjusted analysis for the 2010 cohort, the AQC was associated with an average 3-year change in spending of -$81.92 per enrollee per quarter, or a 8.8 percent decrease ($p<0.001$) in the level of spending compared to pre-intervention and relative to control (Table 4). Pre-intervention trends between AQC and control were statistically different (-$14.51, $p=0.008$). Thus, unlike in the 2009 AQC cohort, this suggests that 2010 cohort spending was growing at a slower rate prior to the intervention, compared to control. Figure 6 illustrates with this finding.

Figure 7 decomposes 2010 cohort spending by site and type of care. Consistent with the 2009 cohort findings, the slowing of spending in the 2010 cohort was driven by the outpatient setting rather than the inpatient setting. Similarly, outpatient facility spending saw the largest decline relative to control, with the two trends also intersecting each other by late 2011. Adjusted analyses supported these raw results. Decomposition of average 4-year spending by site and type of care similarly showed that changes in spending were largest in the outpatient facility setting (-$80.98 per enrollee per quarter, $p<0.001$). Decreases in outpatient professional spending were smaller but also statistically significant (-$17.86 per enrollee per quarter, $p=0.007$). The 2010

cohort did not demonstrate a statistically significant change in inpatient professional spending or in inpatient facility spending.

3.2.3. 2011 AQC Cohort

Unadjusted spending in the 2011 AQC cohort and control is shown in Figure 8. The 2011 cohort consists of a single large provider organization, whose spending trend prior to 2011 shows greater variation compared to the relatively smoother trends in the earlier cohorts. Spending in the pre-intervention period increased modestly between 2006-2008 and more so in 2009-2010. Unadjusted analysis shows that the 2011 AQC cohort spent on average $1044.91 per enrollee per quarter before the AQC (2006-2010) and $1070.56 after entering the AQC (2011-2012), with a difference of $25.65 per enrollee per quarter. Meanwhile, the control group spent $797.84 pre-intervention and $920.67 post-intervention, with the difference being $122.83 per enrollee pre quarter. The resulting difference-in-difference change in spending associated with the AQC was -$97.18 per enrollee per quarter (Table 5).

Adjusted analysis in the 2011 cohort demonstrated that the AQC was associated with an average 2-year change of -$97.10 per enrollee per quarter in spending, equivalent to -9.1 percent, $p<0.001$ (Table 5). Pre-intervention trends between intervention and control were not statistically different (-$3.70, $p=0.52$). Again, this suggests that the 2011 cohort spending was growing at a similar rate prior to the intervention as that of the control group.

The unadjusted decomposition of 2011 AQC cohort spending is summarized in Table 5. Consistent with earlier AQC cohorts, outpatient facility spending accounted for the largest share of the spending change (-$28.27 per enrollee per quarter, or -8.2 percent, $p=0.03$). The 2011 cohort also demonstrated statistically significant spending decreases in outpatient professional

services (-$22.65 per enrollee per quarter, p<0.001). Changes in inpatient professional and inpatient facility spending were not statistically significant (Table 5).

3.2.4. 2012 AQC Cohort

The 2012 AQC cohort had the longest pre-intervention period in the study (2006-2011) and 1 year of post-intervention data (2010). Its unadjusted spending along with control is shown in Figure 9. The 2012 cohort comprised 5 provider organizations, whose average spending trend prior to 2012 was increasing. Unadjusted analysis shows that the 2012 AQC cohort spent on average $981.06 per enrollee per quarter before the AQC and $1022.80 after the AQC, with a difference of $41.74. The control group spent $817.96 before the AQC and $921.01 after, with a difference of $103.05 per enrollee pre quarter. The unadjusted difference-in-difference change in spending associated with the AQC was -$61.31 per enrollee per quarter (Table 6).

In adjusted analysis, the AQC was associated with a year-1 spending change of -$59.39 per enrollee per quarter (-5.8 percent, p=0.04) (Table 6). The pre-intervention trend in the 2012 AQC cohort was modestly higher than in control ($7.93, p=0.05) on average over the 6 years. If interpreted as a meaningful difference, this suggests that the 2012 AQC cohort would have had to overcome a higher baseline growth rate to generate a spending decrease.

The adjusted decomposition of 2012 AQC cohort spending is summarized in Table 6. Outpatient facility spending again explained the largest share of the spending change (-$95.05 per enrollee per quarter, p<0.001). The 2012 cohort saw a statistically significant increase in outpatient professional spending ($14.26, p=0.049). Inpatient professional and facility spending also increased after the AQC relative to control, although estimates were insignificant (Table 6). This suggests that changes in outpatient facility spending were partly offset in this cohort.

3.2.5. Sensitivity Analyses

The changes in spending associated with the AQC for each cohort in each year are shown in Table 7. All results were derived from models using the 8 Northeastern states as controls. Thus, magnitudes for 2009 and 2010 findings (first 2 years of the contract) may differ from those of prior AQC evaluations, which used non-AQC BCBSMA enrollees as the control group.[20,22] Weighted across the cohorts, average AQC-associated savings by year were 2.4 percent in 2009, 3.1 percent in 2010, 8.4 percent in 2011, and 10.0 percent in 2012. These savings were scaled from dollar estimates into percentages by dividing by the given year's claims spending. They are compared to the aggregate magnitudes of incentive payments in a later section below.

Sensitivity analyses for these results are shown in Table 8. In section A of the table, these sensitivity analyses tested the robustness of main results against various changes in the model. Column 1 reproduces the main coefficient of interest (average quarterly change in spending associated with the AQC over the first 4 years of the contract, using the 2009 cohort vs. control comparison). The remaining columns show the same coefficient in alternative scenarios: (2) percent cost sharing in place of plan fixed effects; (3) exclusion of plan type fixed effects; (4-5) exclusion of state or plan fixed effects; (6) exclusion of state and plan fixed effects; (7) exclusion of age and sex; (8) exclusion of risk score; (9) exclusion of age, sex, and risk score; (10) and exclusion of age, sex, and risk score with inclusion of plan fixed effects. Cost sharing is derived by calculating the percent of spending paid by the enrollee out of pocket for the 10 most frequent services and then averaging those percentages by plan. This is a reflection of plan generosity.

In section B of Table 8, sensitivity analyses tested robustness against changes in the variables or sample. Column 1 is again the main coefficient of interest. The remaining columns

show the following modifications: (2) risk scores in deciles rather than a continuous variable; (3) excluding cost sharing from spending; (4) including prescription drug spending; (5) prospective risk score lagged by 1 year; (6) restricting to continuous enrollees over 7 years during the study period; (7) quarterly model at the enrollee level. Importantly, because there were some concerns that unobserved secular factors in Massachusetts could have contributed to the results, columns (8-11) tested alternative control groups that were possible to construct using the available data. These alternative control groups have drawbacks, but were nevertheless tested and compared to the main results. Column (8) uses HMO only controls from the 8 Northeastern states. This group fails to capture all enrollees in plans comparable to the AQC, which require designating a PCP and have incentives for receiving care in network. Also, this group had significant differences in pre-intervention spending trends compared to the AQC. Column (9) uses Massachusetts control subjects only from the Truven (Marketscan) dataset. This group is not ideal because it contains BCBSMA (treatment) enrollees as well; I could not separate BCBSMA enrollees from Harvard Pilgrim, Tufts, or other private payers in MA due to the absence of payer IDs in the Truven data for confidentiality. Moreover, this control group had significant differences in pre-intervention spending trends relative to the AQC. Column (10) uses non-AQC BCBSMA controls (enrollees whose providers had not joined the AQC by 2012). This is not an ideal control group because the remaining providers in non-incentive contracts were small, rural practices that received lower fee updates from BCBSMA as a consequence of remaining in fee-for-service. Moreover, this control group also had significant differences in pre-intervention spending trends relative to the AQC. The Massachusetts only control groups are also susceptible to spillover effects. Column (11) uses nationwide controls: a 10% random sample of enrollees in the 49 non-Massachusetts states in the Truven data. As with the main control group, national controls are susceptible to other factors in Massachusetts affecting the results. However, this control group does not contaminate controls

with treatment subjects and is less susceptible to AQC spillover effects within Massachusetts. Of note, similar to the baseline control group, the national control group demonstrated no significant differences in pre-intervention spending trends relative to the AQC. Overall, sensitivity analyses generally supported the main estimates.

3.3. Utilization

A decrease in spending attributable to the AQC could be driven either by a decrease in prices or a decrease in utilization (volume). Analyses on the 2009 AQC cohort showed that in year 1, this cohort achieved savings through lower prices, rather than through lowering volume. The lower prices were achieved through referring patients to lower priced providers. By the end of year 2, savings continued to be driven by lower prices through using less expensive providers, but decreases in utilization also began to surface in year 2. Roughly one-third of the savings were attributable to decreases in utilization, with about two-thirds due to lower prices.

Direct analyses of utilization are available only through the first 2 post-intervention years for the 2009 AQC cohort. These analyses were focused on several areas of technology-intensive services: cardiovascular services, imaging services, and orthopedic services.[42] They used models with the volume of services as the dependent variable. In the first two years of the contract, the 2009 cohort saw a decrease in the volume of percutaneous coronary intervention (PCI) relative to control (Figure 10A). Utilization of coronary artery bypass surgery, aneurysm repair, and carotid endarterectomy did not demonstrate statistically significant changes between the 2009 cohort and control (Figures 10A, 10B). Table 8 shows the unadjusted and adjusted changes in utilization for these services between the 2009 AQC cohort and control. Utilization of imaging services did not demonstrate statistically significant changes associated with the AQC (Figure 11, Table 9). The

volume of orthopedic services, in terms of knee replacements and hip replacements, also did not show any statistically significant changes associated with the AQC (Figure 12, Table 9).

In the analysis of average 4-year spending changes, the base model using standardized prices produced an average spending decrease of -$24.35 ($p<0.001$) associated with the AQC. Compared to the estimate above from using observed prices, this represents 49 percent of the magnitude, suggesting that just under half of the spending decrease over the first 4 years was attributable to decreases in utilization. The rest (51 percent) of the estimated policy effect is attributable to decreases in prices.

3.4. Quality

3.3.1. Process Measures

Table 10 shows the changes in performance on process quality measures for the 2009 AQC cohort in the first two years of the contract compared to control. Unadjusted results were calculated as the percent of eligible populations for a particular quality measure who met the pre-defined performance threshold for the measure (for example, annual eye exams for patients with diabetes). Difference-in-differences results are interpreted as the percentage-point change among eligible enrollees who met the performance threshold associated with the AQC. Adjusted results in Table 10 were derived using BCBSMA enrollees as controls in a linear, multivariate enrollee-level model through the first two years of the contract. Sensitivity analyses using logistic models did not meaningfully change the results. Adjusted results were decomposed into year-1 and year-2 effects to evaluate the initial trends in the AQC-associated changes. Results for the 3 aggregate measures were calculated by pooling the individual measures (see Methods).

The percent of eligible enrollees who met chronic care management quality performance increased from 79.1 percent before the AQC (2006-2008) to 83.3 percent after the AQC (2009-2010) in the 2009 AQC cohort. The percent of eligible enrollees in the BCBSMA controls saw a smaller increase from 79.7 to 80.0 percent. Adjusted results show that the AQC was associated with a 3.7 percentage-point improvement in aggregate chronic care management over the first 2 years ($p<0.001$). The year-1 effect was a 2.6 percentage-point increase ($p<0.001$), and the year-2 effect was a 4.7 percentage-point increase ($p<0.001$). This aggregate result comprised component improvements in cardiovascular LDL cholesterol screening and diabetes care (4 measures); one component that did not demonstrate a significant improvement in the first two years was short-term and maintenance prescription measures for patients with depression (Table 10).

The quality of adult preventive care improved on average 0.4 percentage points over the first two years ($p=0.004$). It did not show a statistically significant improvement in year 1 (a 0.1 percentage-point change, $p=0.67$), but improved significantly in year-2 (a 0.7 percentage-point improvement, $p<0.001$). This aggregate result was primarily driven by breast cancer screening and by withhold of antibiotics for acute bronchitis (Table 10).

Pediatric quality also improved over the first 2 years, averaging a 1.3 percentage-point increase ($p<0.001$). The year-1 improvement was 0.7 percentage points ($p=0.001$), and the year-2 improvement was 1.9 percentage points ($p<0.001$). Individual measures including well care for babies, children, and adolescents, as well as chlamydia screening for adolescents, contributed to the aggregate improvement. Appropriate testing for pharyngitis saw a decrease associated with the AQC, as a result of greater improvements in the control group. Withhold of antibiotics for acute bronchitis among children also did not contribute to the improvement (Table 10).

After 2 years, adjusted analyses using the enrollee-level model were not possible given the lack of BCBSMA controls. Thus, unadjusted weighted averages of performance on process

measures by each of the 4 AQC cohorts in each of 6 years (2007-2012) are shown in a series of figures (enrollee-level data were not available in 2006). Figure 13A shows performance on the aggregate chronic care management measure, showing a monotonic improvement for the 2009 and 2010 AQC cohorts. Without enrollee-level control data, it is not known to what degree their changes in 2011 and 2012 relative to pre-intervention are attributable to the AQC in a statistical sense. There is some variation across the AQC cohorts in their performance levels.

Figure 13B shows performance on the aggregate adult preventive care measure. Again, variation is noted across the 4 AQC cohorts, with an overall trend towards improvement. Figure 13C summarizes the aggregate pediatric care measure. With rare exception, there is also a trend towards improvement across the cohorts.

For the 2009 AQC cohort, average 4-year changes in unadjusted process quality relative to HEDIS national averages are summarized in Table 11. A continued improvement in the last 2 years of the contract is evident, although these results are not statistically adjusted. Table 11 also summarizes average changes in unadjusted process quality for the 2010, 2011, and 2012 cohorts over the duration of their contracts up through 2012, relative to HEDIS national averages.

3.3.2. Outcome Measures

Descriptive analysis of performance on outcome measures for the 2009 AQC cohort is shown in Table 12. The first 4 columns show performance on the 5 individual measures and the aggregate measure annually in 2009-2012. The right 2 columns show a comparison panel using HEDIS national averages in 2011-2012. In general, the 2009 AQC cohort performed better than national averages. This analysis comprised only unadjusted averages; differences between the 2009 AQC cohort and the HEDIS data cannot be interpreted as an AQC effect, because there

was no enrollee-level statistical analysis that could be undertaken on outcome measures. Figure 14 plots the 2009 AQC cohort against the HEDIS national average for the composite outcome score across 2006-2012. Relative to the national average, this AQC cohort experienced a steady improvement in outcomes, although the interpretation is again descriptive rather than causal.

3.5. Cumulative Payouts

An important distinction must be made between decreases in medical spending associated with the AQC, as demonstrated by the above results, and changes in cumulative payouts from the insurer. Medical spending in these analyses was calculated from actual claims filed by providers to BCBSMA. For each enrollee-quarter observation in the data, medical spending was the sum of claims filed by providers. Thus, it reflects the amount of care provided to beneficiaries, but does not include shared savings surpluses, quality bonuses, or infrastructure bonuses received by the provider organizations in the AQC. In other words, a difference-in-difference result that ties the AQC to a decrease in medical spending does not necessarily mean that overall payouts from the insurer fell in a given year.

Total payouts, including shared savings, quality bonuses, and infrastructure support, exceeded savings on claims in the first 2 years, reflecting upfront investment costs to encourage participation. This pattern continued into 2011, with a smaller gap, but reversed in the 2012, when claims savings exceeded incentive payments to generate a net savings (Table 7). By 2012, total payout growth for the AQC (claims and incentive payments combined) was below the Massachusetts state spending target of 3.6% and below the projected spending based on controls.

4. CONCLUSION AND DISCUSSION

After 4 years, the AQC was associated with decreased medical spending and improved quality of care for the 2009 AQC cohort. The growth rate of spending in this cohort slowed over the 4 years, evident in unadjusted analysis and supported by adjusted results, while performance on process and outcome quality measures steadily increased. Consistent with earlier work, AQC-associated decreases in spending were concentrated in the outpatient facility setting, and savings in the No-Prior-Risk subgroup continued to be greater than those in the Prior-Risk subgroup. The proportion of average savings attributable to decreases in utilization, as opposed to decreases in price, approached 50 percent after the first 4 years. Spending results were not due to changes in coding behavior. Results were generally robust to sensitivity analyses.

The 2010 AQC cohort, comprised of organizations entering from fee-for-service, also experienced a continued decrease in medical spending following from earlier work, although its pre-intervention spending trend was slower than control.[20,22] Year-1 and year-2 results from the 2011 and 2012 cohort are largely consistent with those of the initial cohorts. In general, these results compare favorably with initial reports on ACO performance in the Medicare program and other ongoing ACO-type evaluations. Meanwhile, quality of care in the AQC cohorts largely improved across the years. Process measures improved in a statistically robust manner compared to control enrollees in the first two years, and continued to increase in later years as shown by unadjusted analysis. Unadjusted outcome measures improved relative to national averages.

The spending and quality results observed among AQC groups as they progressed in the contract may serve as a useful benchmark for policymakers and organizations working towards moving the payment system away from fee-for-service. These results from the AQC, however, are still early, and are only representative of one payment model in one state. Nevertheless, they

suggest that global payment implemented effectively within accountable care organizations may serve as a foundation for providers to begin slowing medical spending. The relationship between payers and provider organizations in the ACO paradigm will be crucial for success. Alignment of the incentives to control spending and improve quality will likely be important for collaboration between these parties. For example, the exchange of claims data and progress reports in real time showing spending and quality trends for organizations compared to peers may allow insurers and providers to work together on targeting areas of overuse and low-value care.

4.1. Limitations

The main concern is that other factors in Massachusetts could have influenced spending and quality during the study period. The 2012 Massachusetts payment reform legislation created the state Health Policy Commission and broadly encouraged ACO adoption. Also, global budget contracts with other payers may have spillover effects on the BCBSMA population. However, reforms in Massachusetts mostly postdate the study period. Moreover, Medicare's Pioneer ACO program was launched in 2012; Tufts Health Plan and Harvard Pilgrim Health Plan began global payment contracts in 2012-2013. Therefore, although the findings for 2012 may be susceptible to spillovers, and anticipatory effects from other contracts may also play a role, prior analyses using internal controls, consistency of the sensitivity analyses, and qualitative findings from provider interviews suggest that the AQC played a meaningful role.[20-22]

There are a number of other limitations. First, selection bias is a concern as participation in the AQC was voluntary. The lack of differences in pre-intervention trends between AQC and control attenuates this concern, suggesting that spending trajectories were not already diverging prior to the AQC. That most provider organizations in Massachusetts entered the AQC by year-4

further attenuates this concern. Nevertheless, a potential selection bias cannot be eliminated, as there remain unobserved factors that may have influenced participation as well as spending.

Second, internal validity is threatened if AQC organizations also entered global payment contracts with other payers, which may have spillover effects on the care of BCBSMA patients.[43] Medicare's Pioneer ACO program was launched in 2012; Tufts Health Plan and Harvard Pilgrim Health Plan began global payment contracts in 2012-2013. Therefore, our findings for 2012 may be susceptible to spillovers. Anticipatory effects from these other contracts may also play a role.

Internal validity is also threatened if control states underwent payment reform. However, we know of no broad-scale payment reforms among large private insurers in these states. Some states, such as Rhode Island, piloted medical home interventions, but thus far they have not been shown to significantly affected spending.[44,45] Nevertheless, payment reform was an active issue in many states, especially during the later years of our study period. We cannot identify specific providers or insurers in the Truven data, preventing us from rigorously testing these concerns. However, to the extent that any payment reforms occurred in control states, their effects would be minimized by pooling all these states. To the extent that payment reforms might have slowed spending in control states, our estimated AQC-associated savings would be conservative.

The key question is whether our control group serves as a good counterfactual. We believe the lack of differences in pre-intervention trends and pooling of control states boost the fidelity of the control group. Moreover, this control group, which differed from that of prior AQC evaluations which used non-AQC BCBSMA enrollees, generated similar year-1 and year-2 savings in the 2009 and 2010 cohorts compared to results using those prior controls.[20,22]

Third, results may not generalize to ACOs in Medicare. Most Medicare ACO contracts are 1-sided with shared savings only. Moreover, prices in Medicare are largely uniform rather than negotiated, so savings for Medicare would require reductions in utilization or shifts to less

expensive settings (rather than referrals to less expensive providers). Similarly, results may not generalize to other states, which face different constraints and challenges.[46,47,48,49,50]

Fourth, our quality analyses were descriptive, rather than derived from a statistical model. Earlier work using models analogous to our spending analysis showed significant improvements in all 3 dimensions of process quality, consistent with our descriptive results. Our measures also do not capture all dimensions of quality. Process measures are primary care-centered, while the 5 outcome measures leave numerous important outcomes unmeasured.

Finally, the distinction between decreases in medical spending and changes in cumulative payouts deserves emphasis. As described above, the decrease in medical spending in early years of the AQC were likely exceeded by shared savings, quality bonuses, and infrastructure payouts combined. This was not inconsistent with the design of the AQC, which sets budgets based on actuarial projections to lower spending over the multi-year contract, taking anticipated quality bonuses and other payments into account. These different payments can be viewed as the initial investments by BCBSMA to help motivate provider organizations to move away from pure fee-for-service and embark on delivery system changes to improve the value of care. Obviously, the long-term success of the model depends on how the budget and its growth rate are set, but it also depends on how well organizations can reduce waste within the budgets they take on.

4.2. ACOs Going Forward

In the ACO paradigm, physician organizations face the challenge of changing practice patterns on the ground. After insurance expansion and payment reform from insurers, changing the practice of medicine to control spending and improve quality may appropriately be thought of as the third phase of health care reform.[51] Under global payment, ACOs are asked to manage

population health, coordinate care among providers of different specialties, and function as a medical home for its patients. These are substantial challenges for even large organizations with experience in these domains, not to mention smaller physician groups joining together to become new ACOs. From a scientific standpoint, little is known about how ACOs can teach teamwork to its physicians, about how they can institute joint accountability across specialties, and about how they can become organizations that focus on value rather than volume.

Little is systematically known about how to change the culture of medicine in a palatable way for physicians in an ACO. Organizations such as the Mayo Clinic, Geisinger Health System, Kaiser Permanente, Intermountain Healthcare, and Virginia Mason have implemented innovative payment and delivery systems. Other provider organizations such as the Southcentral Foundation in Alaska have been able to produce impressive results on utilization and quality. Together, these organizations' experiences in recent decades suggest that changing the culture of medicine is key for achieving cost and quality goals. Each has approached cultural change differently, but stories from these organizations have several common themes.

First is leadership. ACOs that succeed on cost and quality tend to have leaders who can motivate an organizational ethos that complements the professional ethos of medicine. Under a global payment contract, clinicians in an organization are truly in it together. When a physician does not order an unnecessary test, savings accrue to the organization. When a provider calls a patient and works with him or her to prevent an unnecessary visit to the emergency department, the entire organization benefits. When patients are satisfied with their care, the organization is rewarded together. Therefore, successful leadership seems to motivate members in an ACO to feel invested in one another. It is able to unite providers in a shared vision and keep them going forward through difficult tradeoffs. For example, if an organization decides to invest more of its

resources under global payment to population health and prevention, leaders will need to secure buy-in from physicians across the organization.

In addition to leadership, incentives are certainly important. Innovations in payment are a theme among physician organizations that have successfully lowered spending and improved the quality of care in certain clinical contexts. A focus on the collective value of care for patients is a helpful foundation for the ACO. It encourages clinicians to think about the cost of care, how they coordinate care with one another, consult each another, and refer patients to each another, all of which affect resource utilization. Both financial and nonfinancial incentives that reward value, particularly through teamwork, could be flexibly designed by an organization to suit its culture under a global payment. Several of the organizations mentioned above have improved the value of their care with physicians on salary, with creative incentives to motivate physicians to care about their colleagues' patients as well as their own.[52] Other organizations have found ways to motivate team performance around common clinical scenarios.[53] As the ACO paradigm moves forward, a greater understanding of behavioral economics and the sociology of physician referral networks by ACO leaders may enable them to creatively design additional incentives.[54,55,56]

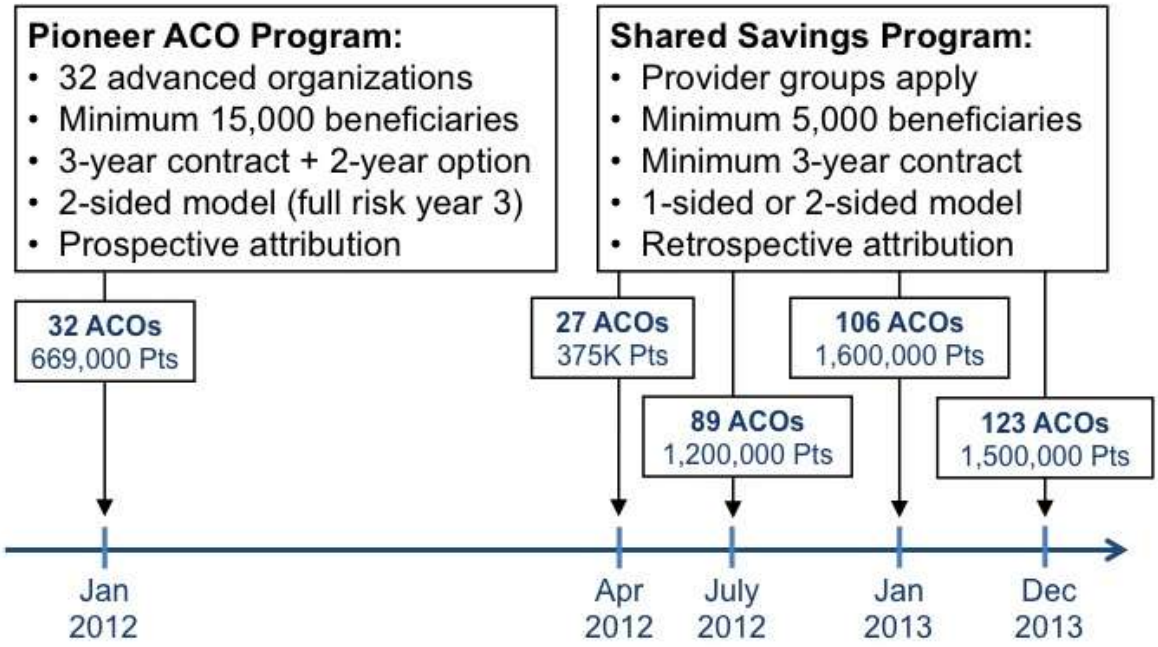
Furthermore, engaging patients in the clinical decision-making process and in practicing prevention outside the doctor's office can be a part of successful cultural change. For example, the Mayo clinic uses a number of patient family advisory committees, which listen to patient and family concerns and involves patients in establishing practice guidelines. ACOs put physicians and patients on the same team. Reducing the supply of unnecessary care and reducing demand for it are equally beneficial for an organization's global budget. However, for population health management to work, the population likely needs to feel empowered and connected to providers in the organization.

4.3. Suggestions for Future Work and Summary

Research on the AQC may be informative for the physician and health policy community by providing an example of changes in spending and quality associated with a broad-scale global payment initiative. Future work on the AQC should further explore changes in utilization in the later years of the contract, which seem to explain an increasing share of the savings. Changes in volume in specific service lines, following on the work with the first 2 years for the 2009 cohort and describing what happens in the other cohorts, would be a meaningful extension to this work.

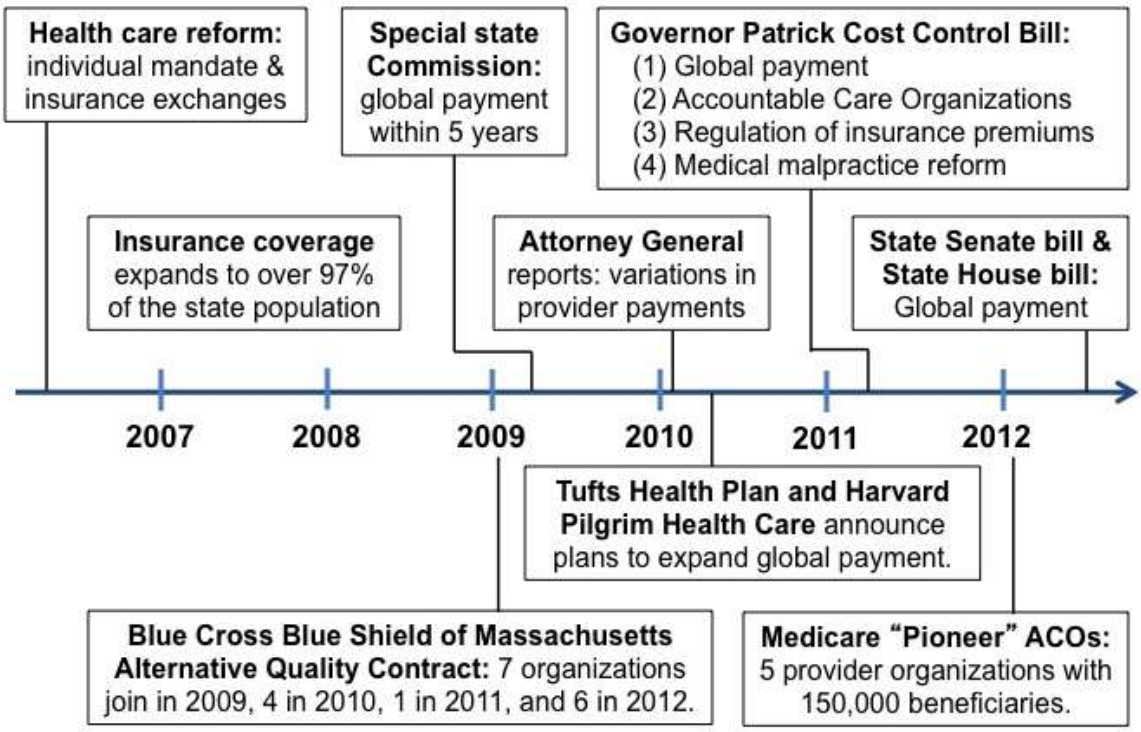
Results from the first 4 years of the AQC suggest that global payment within accountable care organizations may be an effective tool for slowing the growth rate of health care spending and improving the quality of care. A multi-year global budget with shared savings and shared risk may provide physician organizations an incentive to embark on delivery system reforms to improve the value of care. Robust quality measures tied to substantial quality incentives could serve as an effective buffer against stinting, at least in areas that quality measures target. Despite the promise of global payment, challenges remain for physician organizations across the country adopting this type of payment model. The ability of payment reform to improve the value of U.S. health care depends on whether provider organizations can successfully change practice patterns and the culture of medicine in an increasingly constrained health care economy.

Figure 1. Accountable Care Organizations in the Medicare Program*



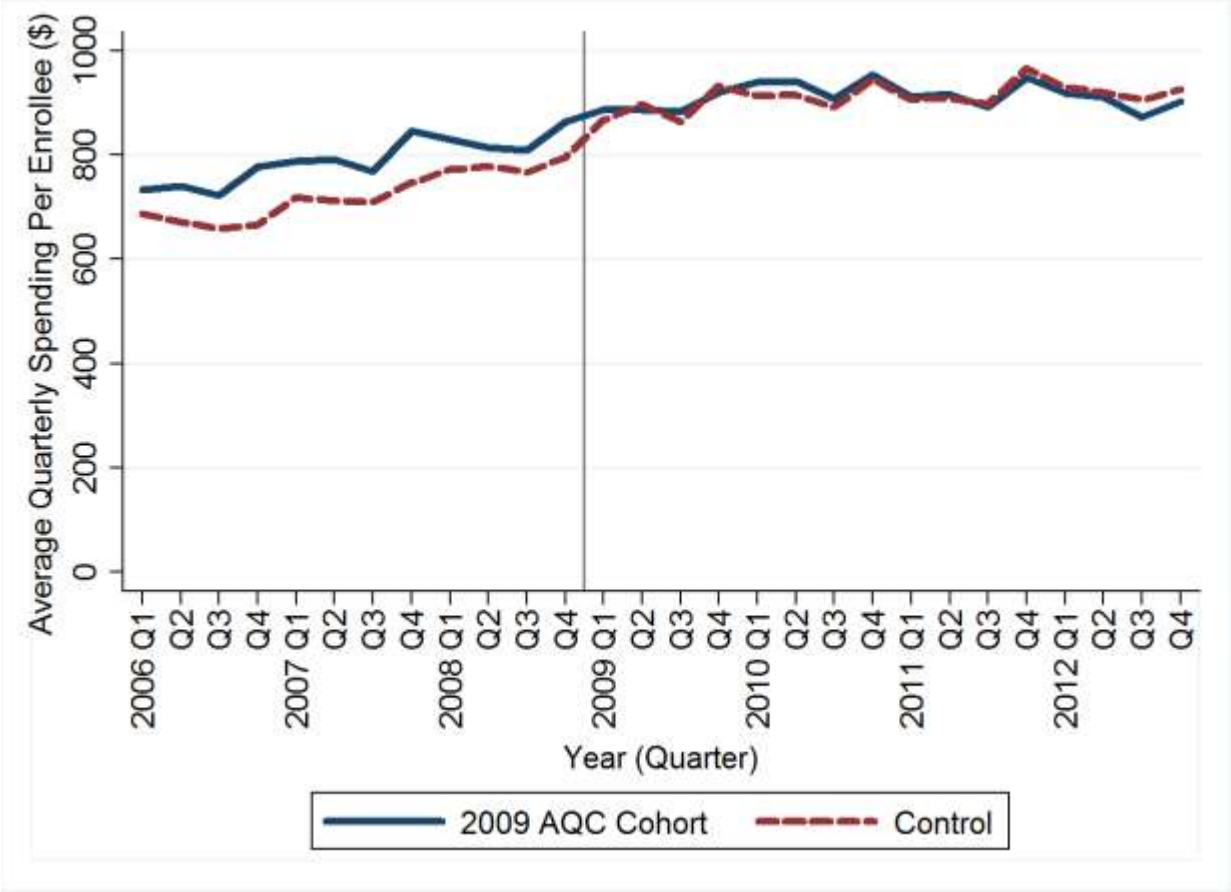
* The Affordable Care Act authorized the creation of accountable care organizations (ACOs) in Medicare. The first ACOs were launched in January, 2012, comprising 32 advanced or "Pioneer" physician organizations that took on a 2-sided ACO contract with shared savings and shared risk for large populations of Medicare beneficiaries. Since then, 4 waves of Shared Savings Program ACOs have been launched, consisting of organizations in 1-sided contracts with shared savings but no shared risk during the initial contracting period. In total, as of January, 2014, the Center for Medicare and Medicaid Services estimates that there are 360 ACOs in the U.S. serving about 5.3 million Medicare beneficiaries.

Figure 2. Timeline of Health Care Reform in Massachusetts*



* In 2006, Massachusetts embarked on a coverage expansion that increased the rate of insurance in the state to over 97 percent. The ensuing years saw continued growth in health care spending, prompting state lawmakers, the governor, private insurers, and other stakeholders to engage in an effort to slow the growth of health care spending. The Alternative Quality Contract (AQC) was implemented in 2009, with 7 physician organizations entering the contract in the first year. By 2012, about 85 percent of the physicians in the state who work with Blue Cross Blue Shield of Massachusetts had entered the AQC. Importantly, the AQC took place in this broader context of state efforts to slow health care spending growth.

Figure 3. Unadjusted Spending: 2009 AQC Cohort vs. Control*



* Unadjusted spending per enrollee per quarter. The control group comprises commercial enrollees in employer-sponsored HMO and POS plans across 8 Northeastern states: CT, ME, NH, NJ, NY, PA, RI, and VT.

Figure 4. 2009 AQC Cohort vs. Control: Decomposition of Unadjusted Average Medical Spending By Type and Site of Care.*

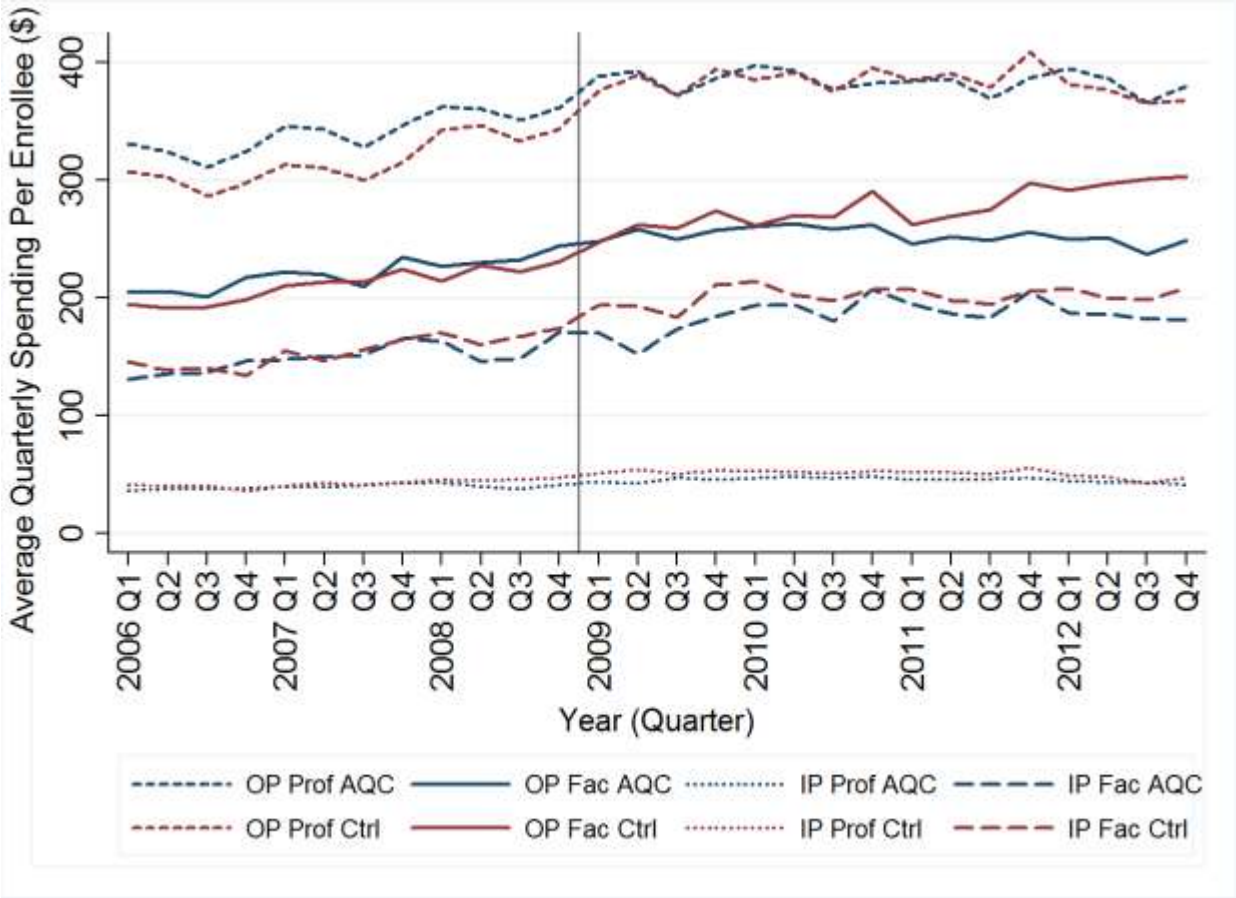


Figure 5. 2009 AQC Cohort vs. Control: Decomposition of Unadjusted Average Medical Spending By Organizational Prior Risk Contracting Experience.*

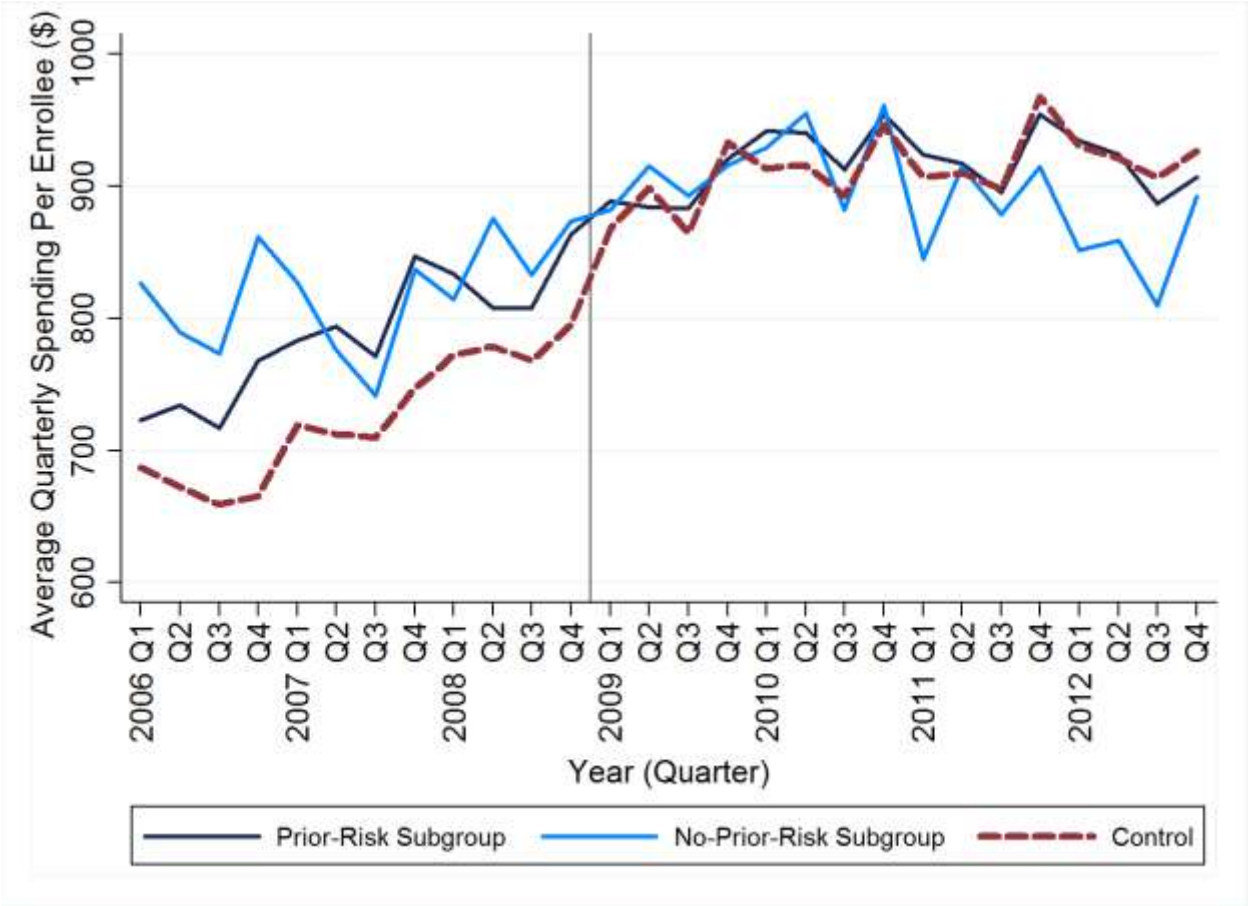


Figure 6. 2010 AQC Cohort vs. Control: Unadjusted Average Medical Spending Per Enrollee Per Quarter.*

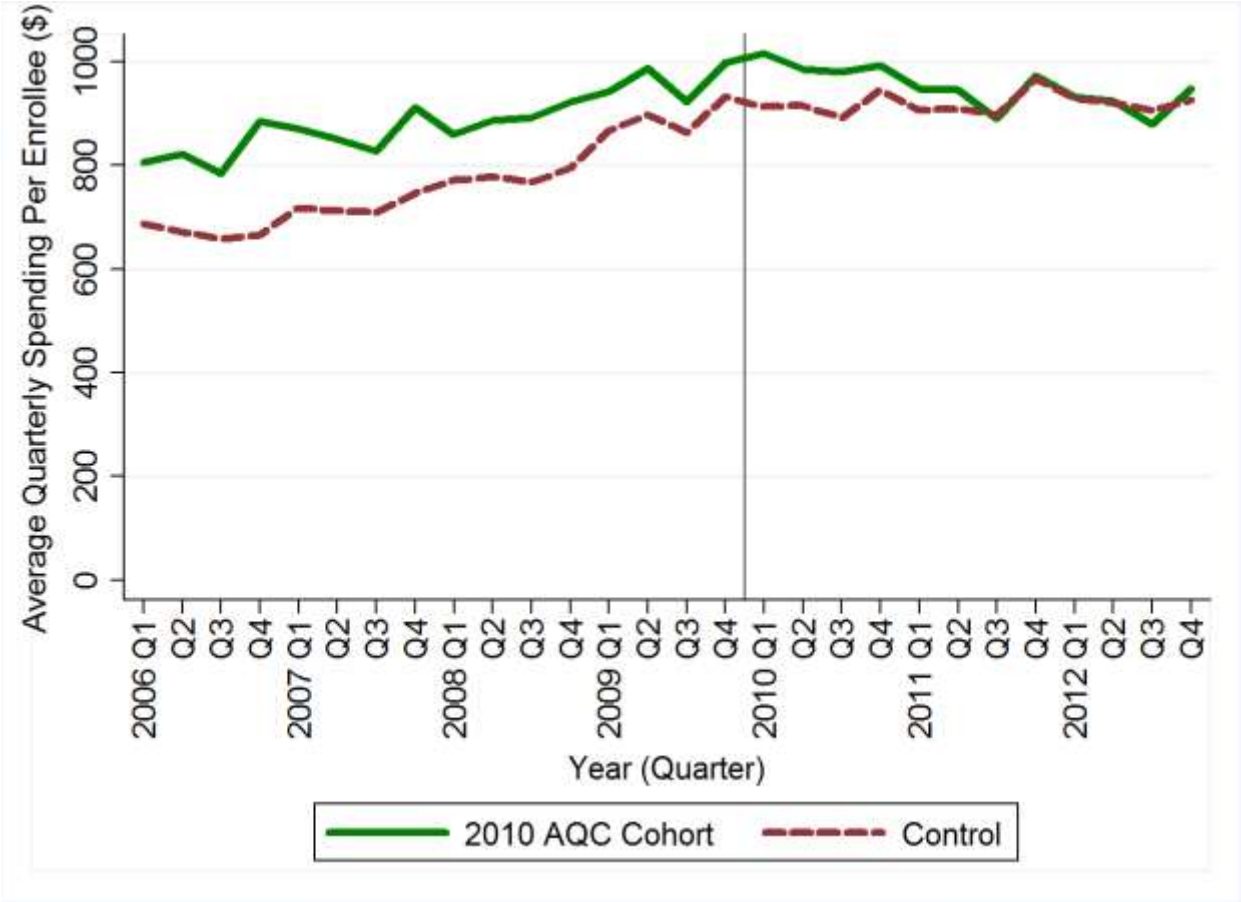


Figure 7. 2010 AQC Cohort vs. Control: Decomposition of Unadjusted Average Medical Spending By Type and Site of Care.*

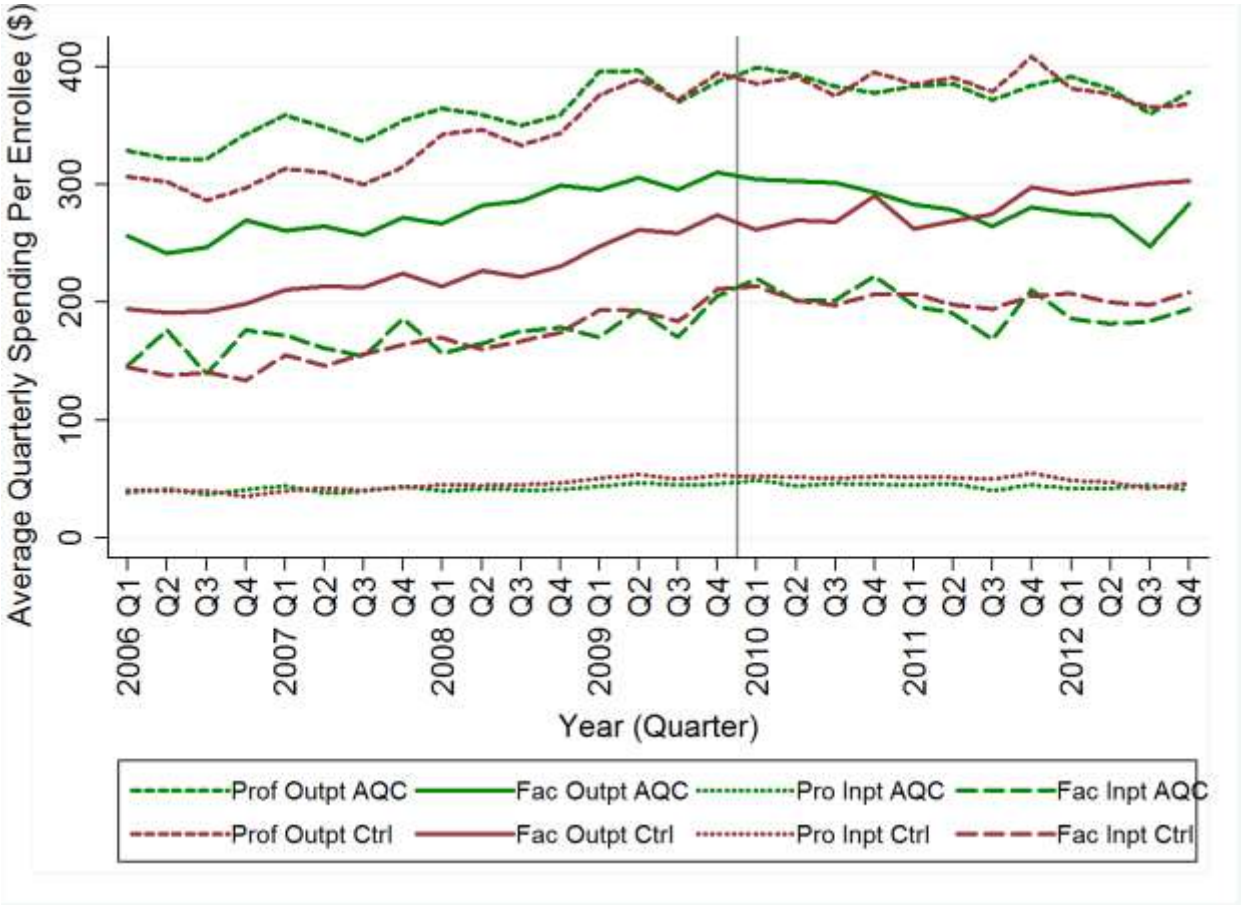


Figure 8. 2011 AQC Cohort vs. Control: Unadjusted Average Medical Spending Per Enrollee Per Quarter.*

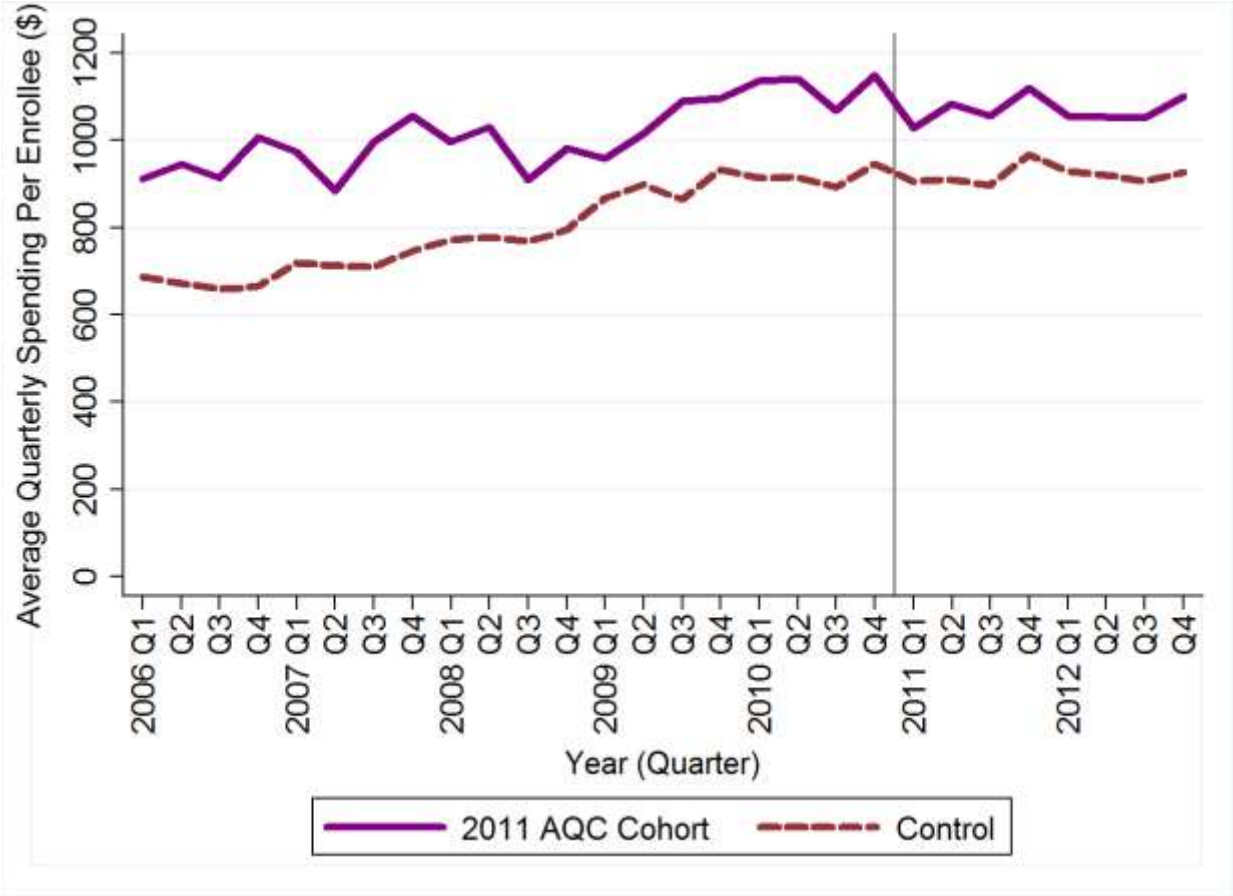


Figure 9. 2012 AQC Cohort vs. Control: Unadjusted Average Medical Spending Per Enrollee Per Quarter.*

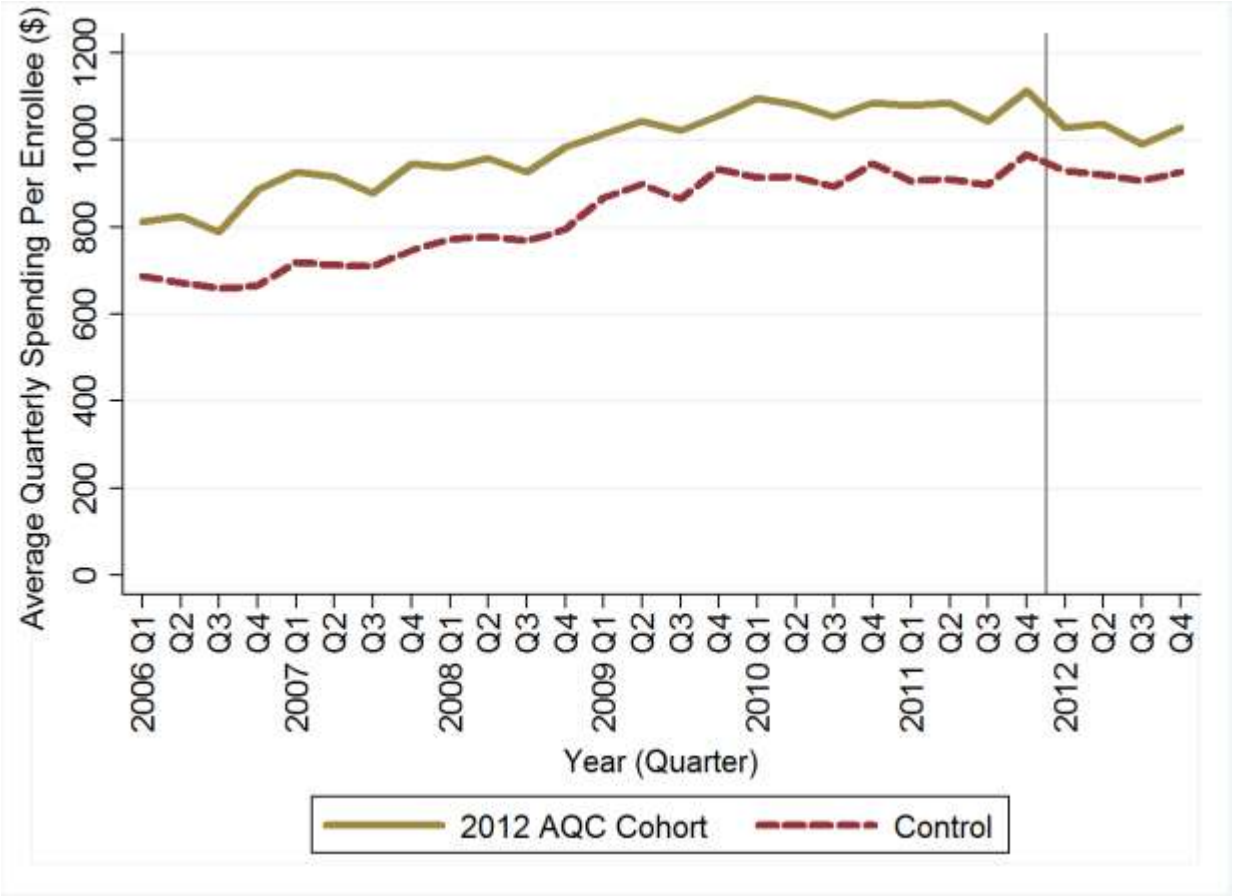
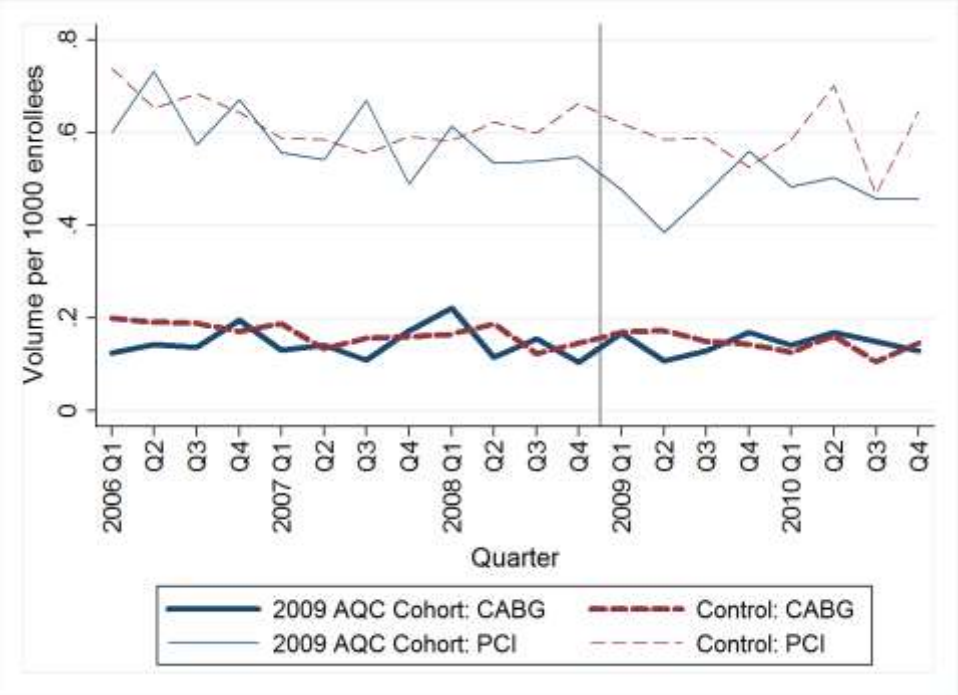


Figure 10. Utilization of Cardiovascular Services, 2009 AQC Cohort vs. Control*

A. Coronary Artery Bypass Surgery and Percutaneous Coronary Intervention



B. Aneurysm Repair and Carotid Endarterectomy, 2009 AQC Cohort vs. Control*

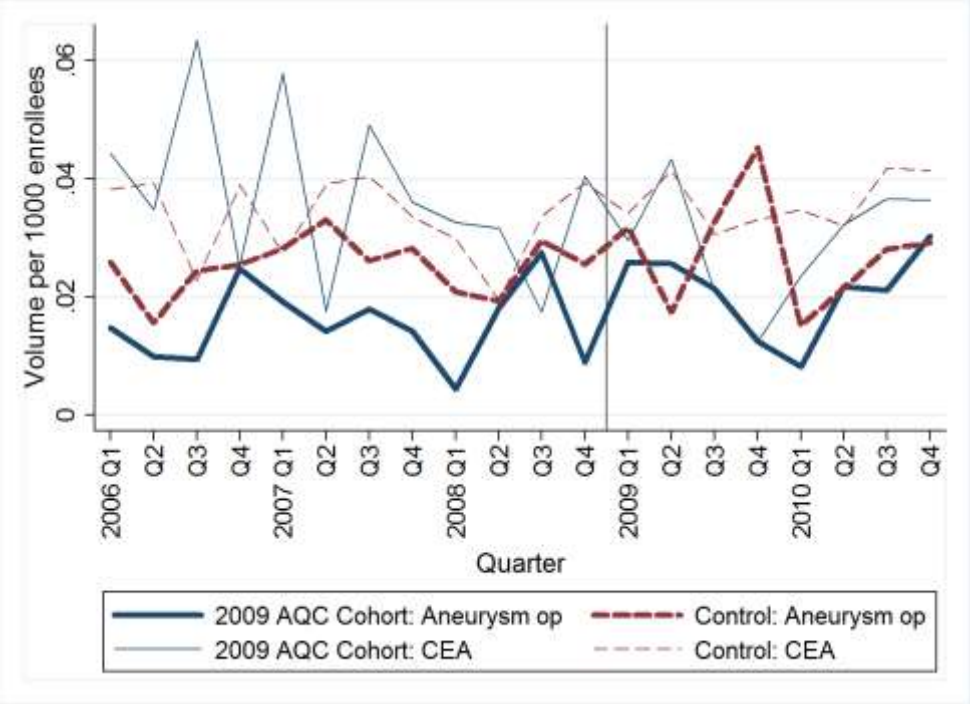
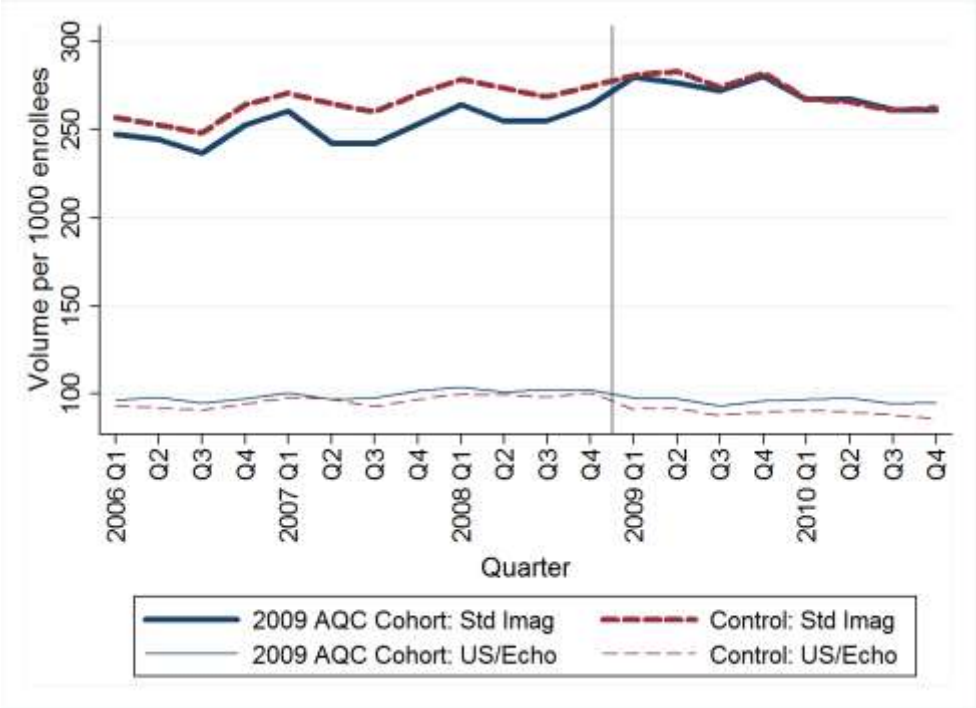


Figure 11. Utilization of Imaging Services, 2009 AQC Cohort vs. Control*

A. Utilization of Standard Imaging and Ultrasound, 2009 AQC Cohort vs. Control*



B. Computed Tomography and Magnetic Resonance Imaging

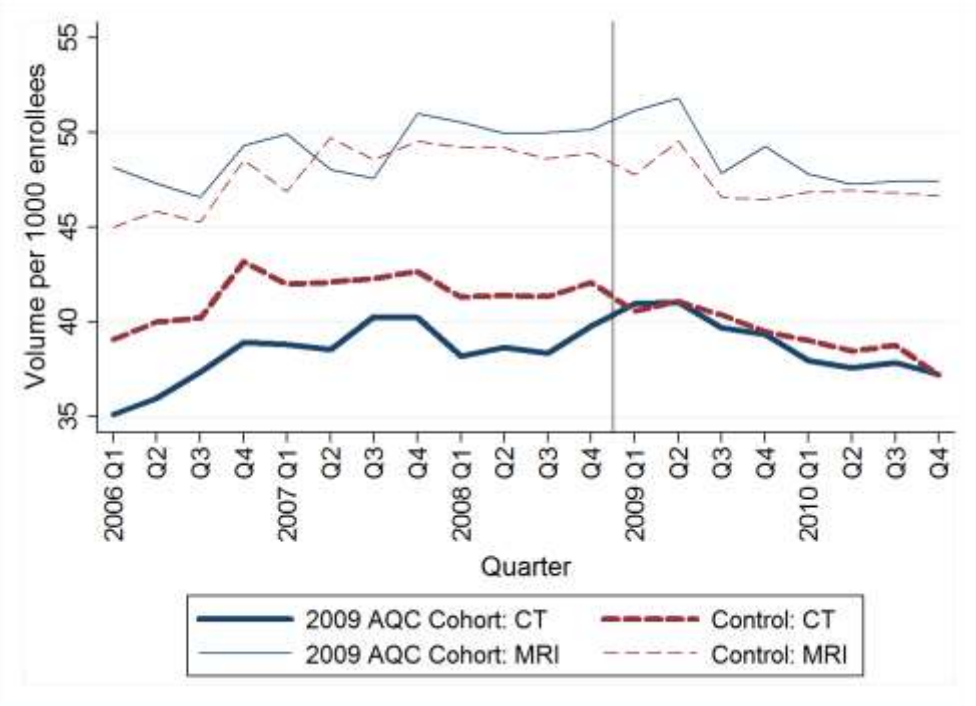


Figure 12. Utilization of Orthopedic Services, 2009 AQC Cohort vs. Control*

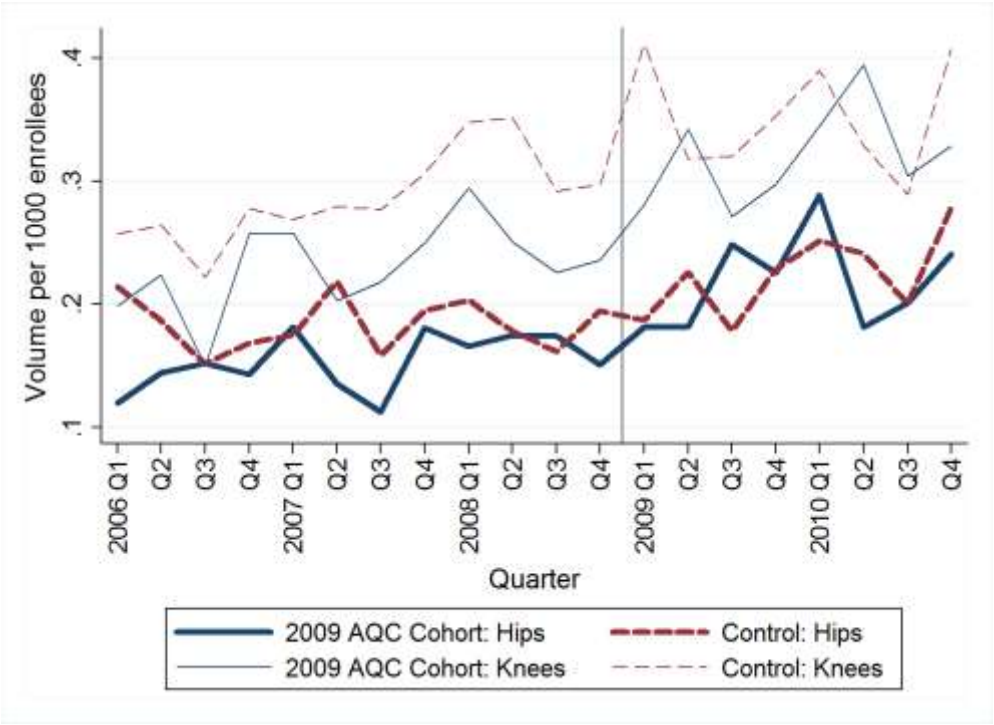
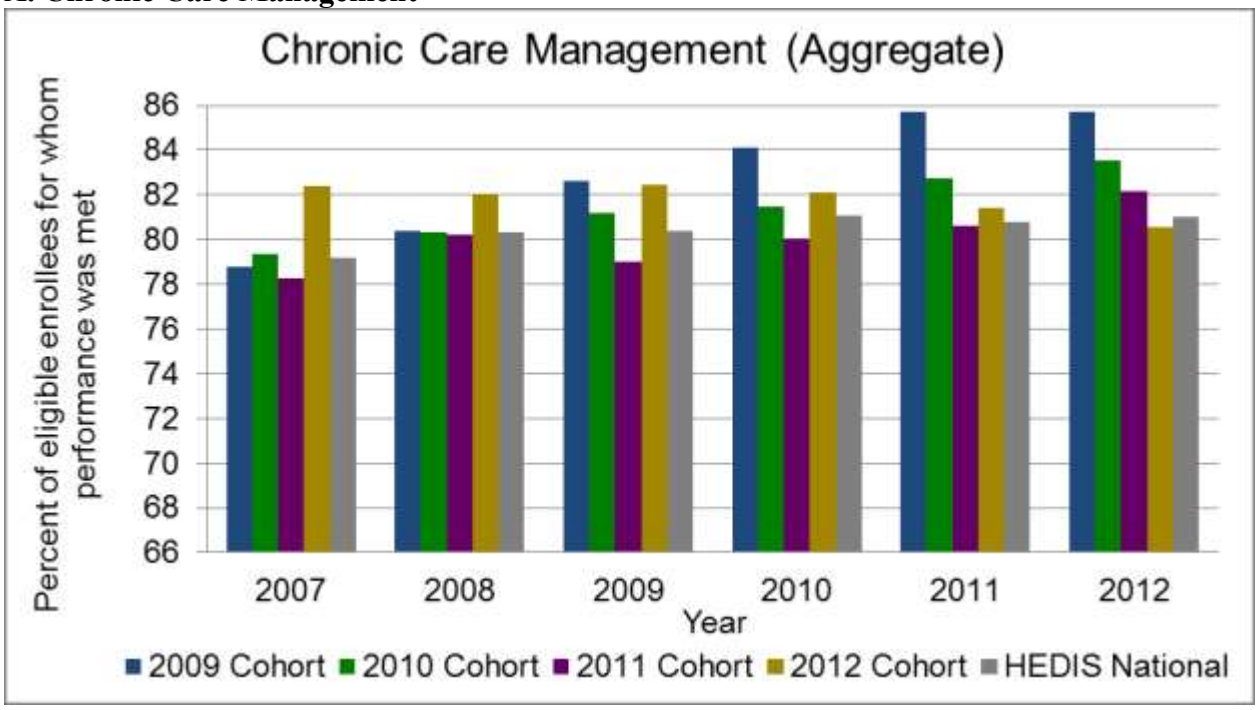
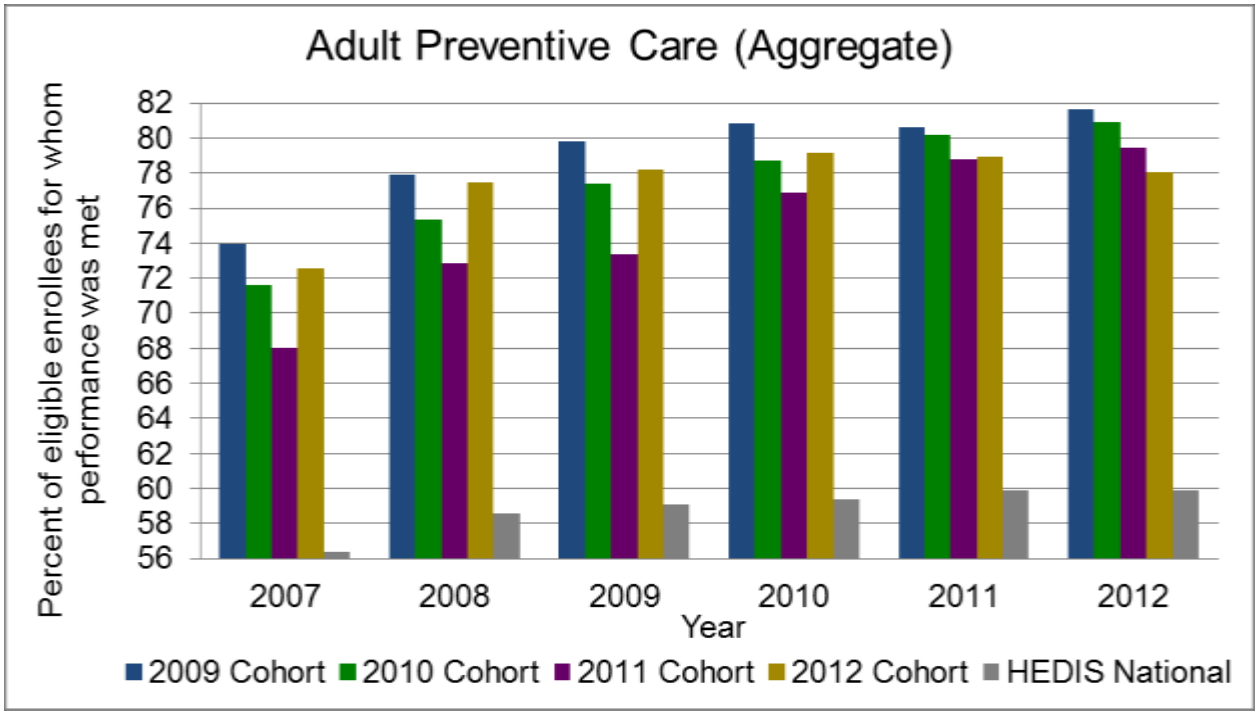


Figure 13. Process Quality by AQC Cohort, Aggregate Results 2007-2012
A. Chronic Care Management*



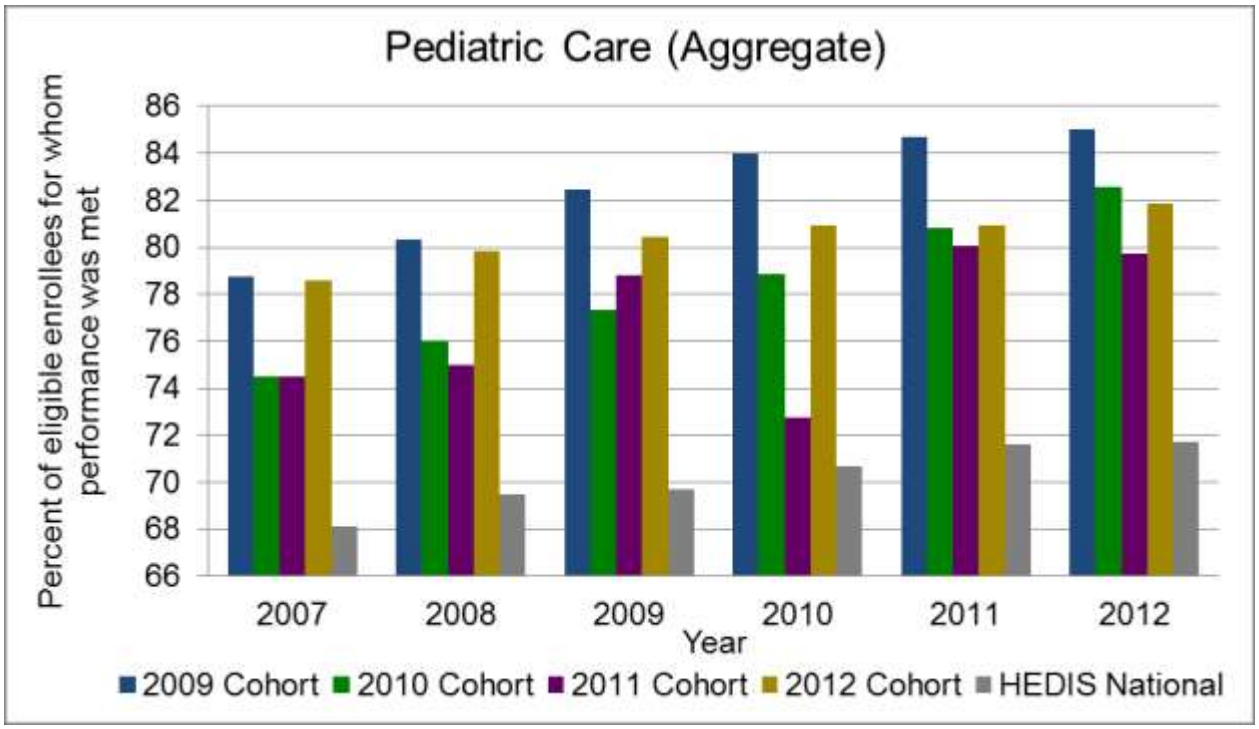
* Unadjusted performance on chronic care management quality measures for all AQC cohorts and control. This aggregate measure is a weighted average of 7 individual process measures: cardiovascular low-density lipoprotein (LDL) cholesterol screening, 4 measures for enrollees with diabetes (glycated hemoglobin testing, eye exam, LDL cholesterol screening, and nephrology screening), and 2 measures for depression care (short-term prescription and maintenance prescription).

B. Adult Preventive Care*



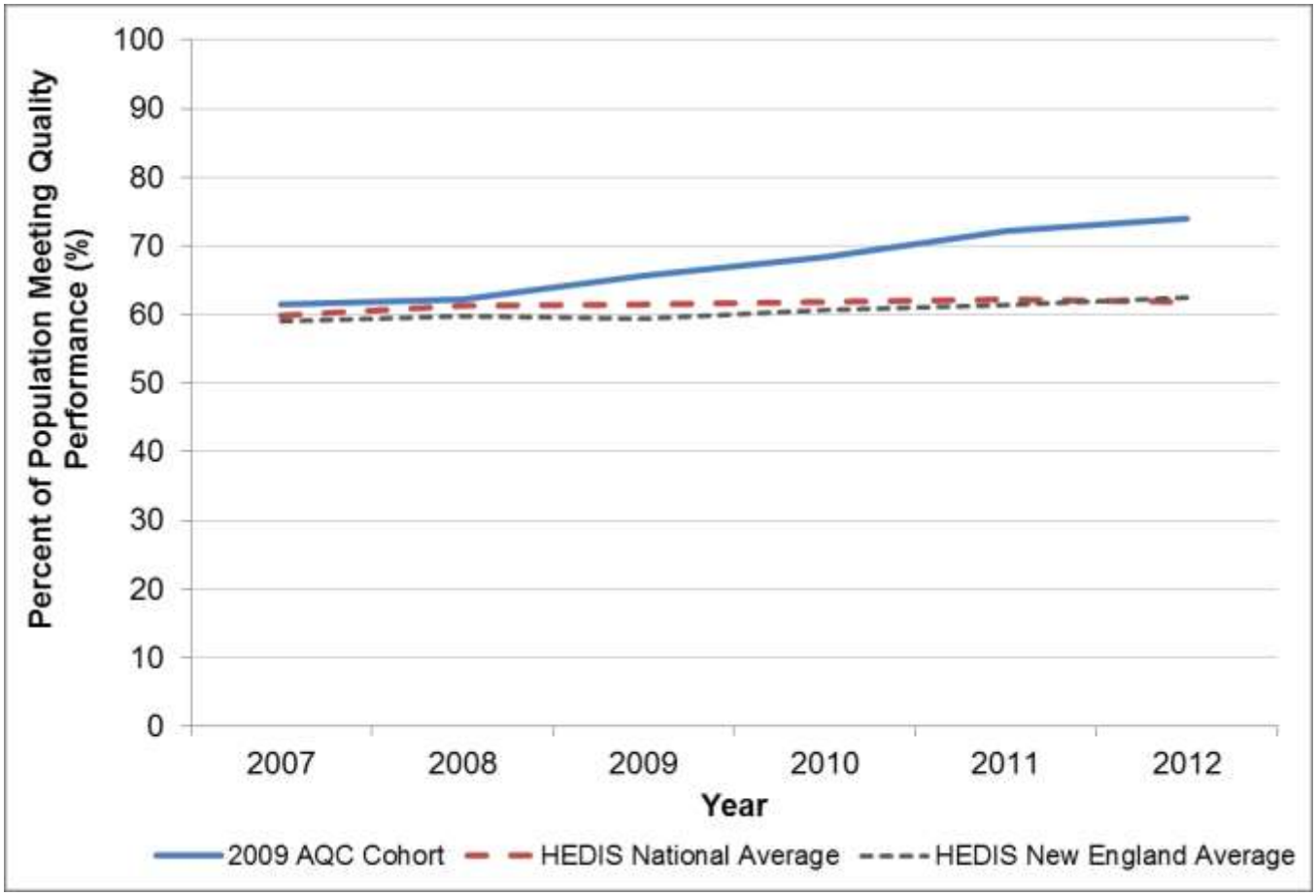
* This aggregate measure is a weighted average of 5 individual measures: breast cancer screening, cervical cancer screening, colorectal cancer screening, chlamydia screening for enrollees 21–24 years of age, and no antibiotics for acute bronchitis.

C. Pediatric Care*



* This aggregate measure is a weighted average of 6 individual measures: Appropriate testing for pharyngitis, chlamydia screening for enrollees 16–20 years of age, no antibiotics for upper respiratory infection, and 3 measures for well child visits (babies <15 months of age, children 3–6 years of age, and adolescents).

Figure 14. Outcome Quality, 2009 AQC Cohort vs. HEDIS (2007-2012)*



* Outcome quality consisted of 5 measures. For patients with diabetes: (1) hemoglobin A1c control (≤9 percent), (2) low density liproprotein (LDL) cholesterol control (<100 mg/dL), (3) blood pressure control (<140/80 mmHg); for patients with cardiovascular diseases (4) blood pressure control (<140/90 mmHg), and (5) LDL control (<100 mg/dL). HEDIS is the Healthcare Effectiveness Data and Information Set.

Table 1. Characteristics of AQC Cohorts*

AQC Cohort	2009 Cohort	2010 Cohort	2011 Cohort	2012 Cohort	Control Group
Enrollees*	490,167	177,312	97,754	583,002	966,813
Age (yr)	34.1 ± 18.2	35.6 ± 17.9	41.1 ± 14.7	31.8 ± 19.1	33.7 ± 18.3
Female sex (%)	52.2	52.1	52.3	51.9	50.0
DxCG risk score†					
Mean	1.03	1.09	1.26	1.03	1.00
Median	0.48	0.51	0.61	0.46	0.41
(Interquartile range)	(0.19–1.07)	(0.21–1.15)	(0.25–1.35)	(0.19–1.05)	(0.16–1.04)
Cost sharing (%)					
Mean	12.1	11.8	12.7	10.4	18.5
Median	8.6	8.4	8.3	7.2	14.4
(Interquartile range)	(4.3–15.6)	(4.3–14.9)	(4.1–16.0)	(3.6–12.9)	(8.2–24.2)
Provider Organizations	7	4	1	5	--
Primary care physicians	1,151	469	420	2,115	--
Specialists	2,197	1,010	1,319	7,260	--
Affiliated hospitals	15	13	2	10	--

* Number of unique individuals enrolled for at least 1 year in the study period. Enrollees in AQC cohorts designated primary care physicians who practice in an organization that joined the AQC. The control group comprises commercially insured individuals in employer-sponsored plans across the 8 other Northeastern states (CT, ME, NH, NJ, NY, PA, RI, VT). No data on provider organizations were available for controls. Age, sex, health risk score, and cost sharing are pooled across all enrollees in the entire study period.

† The DxCG risk score is a measure of enrollee health status, calculated using coefficients from a statistical model from a national claims database that relates spending to ICD-9 diagnoses and demographic information. The DxCG method is similar to Medicare's Hierarchical Condition Category risk scores system and is commonly used for risk adjustment purposes. It is a product of Verisk Health and a proprietary software. Across all enrollees in the study data, the average risk score was 1.03, and it ranged from 0.18 at the 25th percentile to 1.07 at the 75th percentile. Higher values mean higher expected spending.

Table 2. Characteristics of the Population: AQC Cohorts vs. Control

A. 2009 AQC Cohort vs. Control*

	2009 Cohort (N=490,167)		Control (N=966,813)	
	Pre (2006-2008)	**Post (2009-2012)**	**Pre (2006-2008)**	**Post (2009-2012)**
Age (yr)	33.5 ± 18.2	34.6 ± 18.1	33.4 ± 18.2	33.9 ± 18.3
Female sex (%)	52.3	52.1	50.1	50.0
DxCG risk score				
Mean	0.97	1.07	0.94	1.05
Median	0.45	0.50	0.39	0.43
(Interquartile range)	(0.19–1.00)	(0.21–1.11)	(0.14–0.97)	(0.16–1.09)
Cost-sharing (%)				
Mean	10.9	13.0	17.5	19.3
Median	8.3	8.8	13.9	14.9
(Interquartile range)	(4.1–14.5)	(4.4–16.6)	(7.9–22.8)	(8.4–25.4)

B. 2010 AQC Cohort vs. Control*

	2010 Cohort (N=177,312)		Control (N=966,813)	
	Pre (2006-2009)	**Post (2010-2012)**	**Pre (2006-2009)**	**Post (2010-2012)**
Age (yr)	34.5 ± 18.2	37.3 ± 17.5	33.5 ± 18.3	34.0 ± 18.3
Female sex (%)	52.1	52.0	50.2	49.8
DxCG risk score				
Mean	1.04	1.15	0.97	1.06
Median	0.49	0.55	0.40	0.43
(Interquartile range)	(0.19–1.10)	(0.23–1.22)	(0.15–1.00)	(0.16–1.09)
Cost-sharing (%)				
Mean	10.8	13.2	17.5	20.1
Median	8.1	8.8	13.7	15.7
(Interquartile range)	(4.2–14.0)	(4.6–16.6)	(7.8–22.7)	(8.8–26.5)

C. 2011 AQC Cohort vs. Control*

	2011 Cohort (N=97,754)		Control (N=966,813)	
	Pre (2006-2010)	**Post (2011-2012)**	**Pre (2006-2010)**	**Post (2011-2012)**
Age (yr)	41.1 ± 14.8	41.1 ± 14.7	33.6 ± 18.3	33.9 ± 18.2
Female sex (%)	51.4	53.5	50.2	49.5
DxCG risk score				
Mean	1.24	1.28	0.98	1.05
Median	0.60	0.63	0.41	0.43
(Interquartile range)	(0.24–1.32)	(0.26–1.38)	(0.15–1.02)	(0.16–1.08)
Cost-sharing (%)				
Mean	11.9	14.0	17.7	20.7
Median	7.9	9.0	13.8	16.3
(Interquartile range)	(3.9–14.8)	(4.5–18.0)	(7.9–23.1)	(9.2–27.4)

D. 2012 AQC Cohort vs. Control*

	2012 Cohort (N=583,002)		Control (N=966,813)	
	Pre (2006-2011)	**Post (2012)**	**Pre (2006-2011)**	**Post (2012)**
Age (yr)	31.8 ± 19.1	31.9 ± 18.1	33.6 ± 18.0	33.9 ± 18.2
Female sex (%)	52.0	51.4	50.1	49.5
DxCG risk score				
Mean	1.03	1.06	1.00	1.03
Median	0.46	0.48	0.41	0.42
(Interquartile range)	(0.19–1.05)	(0.20–1.09)	(0.16–1.03)	(0.16–1.06)
Cost-sharing (%)				
Mean	10.0	13.0	18.2	20.8
Median	7.1	8.4	14.2	16.4
(Interquartile range)	(3.6–12.5)	(4.3–16.5)	(8.1–23.7)	(9.2–27.5)

Table 3. Change in Average Spending per Enrollee per Quarter, 2009 AQC Cohort vs. Control*

	AQC Enrollees in MA			Individuals in Control States			Between-Group Change		
	Pre 2006-08	Post 2009-12	Change	Pre 2006-08	Post 2009-12	Change	Unadjusted	Adjusted	P
2009 Cohort	789.35	913.15	123.8	731.61	911.40	179.79	-55.99	-62.21	<0.001
Inpt Professional	39.15	45.16	6.01	42.54	50.72	8.18	-2.17	0.40	0.82
Inpt Facility	149.05	185.58	36.53	156.33	201.10	44.76	-8.24	-3.32	0.524
Outpt Professional	340.67	384.23	43.56	319.74	383.7	63.97	-20.4	-15.35	0.004
Outpt Facility	219.91	252.65	32.74	213.00	275.88	62.88	-30.13	-48.67	<0.001
Prior Risk Subgroup	786.11	916.77	130.66	731.61	911.40	179.79	-49.13	-57.61	<0.001
Inpt Professional	39.15	45.65	6.49	42.54	50.72	8.18	-1.69	0.40	0.815
Inpt Facility	148.04	184.25	36.21	156.33	201.10	44.76	-8.56	-3.28	0.522
Outpt Professional	343.41	392.76	49.35	319.74	383.70	63.97	-14.61	-11.87	0.019
Outpt Facility	214.94	249.21	34.27	213.00	275.88	62.88	-28.61	-46.69	<0.001

No Prior Risk Subgroup	817.97	890.51	72.53	731.61	911.40	179.79	-107.26	-68.66	<0.001
Inpt Professional	39.10	42.10	3.00	42.54	50.72	8.18	-5.19	2.44	0.26
Inpt Facility	157.92	193.90	35.98	156.33	201.10	44.76	-8.79	5.26	0.577
Outpt Professional	316.57	330.87	14.30	319.74	383.70	63.97	-49.66	-23.13	<0.001
Outpt Facility	263.76	274.21	10.45	213.00	275.88	62.88	-52.43	-64.25	<0.001

All values are in units of dollars per quarter per enrollee. Adjusted results are derived from the statistical model as described in the Methods section. All values are inflation-adjusted to 2012 U.S. dollars.

Table 4. Change in Average Spending per Enrollee per Quarter, 2010 AQC Cohort vs. Control *

	AQC Enrollees in MA			Individuals in Control States			Between-Group Change		
	Pre 2006-09	Post 2010-12	Change	Pre 2006-09	Post 2010-12	Change	Unadjusted	Adjusted	P
2010 Cohort	876.42	954.74	78.32	772.69	919.43	146.74	-68.42	-81.92	<0.001
Inpt Professional	41.02	44.5	3.48	44.96	50.24	5.28	-1.81	1.23	0.583
Inpt Facility	166.53	197.54	31.01	166.42	203.31	36.88	-5.88	6.21	0.409
Outpt Professional	354.02	383.15	29.13	336.05	383.98	47.92	-18.79	-17.86	0.007
Outpt Facility	274.18	283.81	9.63	225.25	281.9	56.65	-47.02	-80.98	<0.001

* The 2010 AQC cohort comprises 4 physician organizations that entered the AQC from prior fee-for-service contracts. Thus, this cohort is analogous in its absence of prior risk contracting experience as the No-Prior-Risk subgroup of the 2009 AQC cohort.

All values are in units of dollars per quarter per enrollee. Adjusted results are derived from the statistical model as described in the Methods section. All values are inflation-adjusted to 2012 U.S. dollars.

Table 5. Change in Average Spending per Enrollee per Quarter, 2011 AQC Cohort vs. Control *

	AQC Enrollees in MA			Individuals in Control States			Between-Group Change		
	Pre 2006-10	Post 2011-12	Change	Pre 2006-10	Post 2011-12	Change	Unadjusted	Adjusted	P
2011 Cohort	1,044.91	1,070.56	25.65	797.82	920.64	122.83	-97.18	-97.10	<0.001
Inpt Professional	51.26	50.3	-0.96	46.17	49.44	3.27	-4.23	-4.49	0.225
Inpt Facility	215.15	211.39	-3.76	173.19	202.4	29.21	-32.97	-30.95	0.07
Outpt Professional	399.85	414.1	14.25	344.94	382.54	37.6	-23.35	-22.65	0.013
Outpt Facility	326.77	344.63	17.86	233.52	286.26	52.75	-34.89	-28.27	0.028

All values are in units of dollars per quarter per enrollee. Adjusted results are derived from the statistical model as described in the Methods section. All values are inflation-adjusted to 2012 U.S. dollars.

Table 6. Change in Average Spending per Enrollee per Quarter, 2012 AQC Cohort vs. Control *

	AQC Enrollees in MA			Individuals in Control States			Between-Group Change		
	Pre 2006-11	Post 2012	Change	Pre 2006-11	Post 2012	Change	Unadjusted	Adjusted	P
2012 Cohort	981.06	1,022.80	41.74	817.93	921.01	103.08	-61.33	-59.39	0.035
Inpt Professional	45.11	43.46	-1.65	47.16	46.32	-0.84	-0.81	5.40	0.245
Inpt Facility	173.87	183.26	9.39	177.81	203.58	25.77	-16.37	10.65	0.464
Outpt Professional	386.48	417.73	31.25	352.48	373.01	20.53	10.72	14.26	0.049
Outpt Facility	331.64	332.86	1.22	240.48	298.1	57.62	-56.4	-95.05	<0.001

All values are in units of dollars per quarter per enrollee. Adjusted results are derived from the statistical model as described in the Methods section. All values are inflation-adjusted to 2012 U.S. dollars.

Table 7. Changes in Medical Spending and Total Payments Associated with the AQC by Cohort by Year

Changes in Medical Spending on Claims*	2009		2010		2011		2012		Cohort Average		
	$	P	$	P	$	P	$	P	$	P	%
2009 Cohort	-20.95	0.02	-30.06	0.02	-77.07	<0.001	-120.78	<0.001	-62.21	<0.001	-6.8
2010 Cohort			-29.06	0.03	-85.49	<0.001	-131.21	<0.001	-81.92	<0.001	-8.8
2011 Cohort					-76.96	0.001	-117.24	0.001	-97.10	<0.001	-9.1
2012 Cohort							-59.39	0.04	-59.39	0.04	-5.8
Weighted Average Savings on Claims†	2.4% of current year FFS claims		3.1% of current year FFS claims		8.4% of current year FFS claims		10.0% of current year FFS claims				
Incentive Payments to Providers‡	6% to 9% of current year FFS claims		9% to 12% of current year FFS claims		10% to 13% of current year FFS claims		6% to 9% of current year FFS claims				
Implication	BCBSMA payments to providers, including shared savings and bonuses for quality and infrastructure, exceeded savings on claims.				Payments exceeded savings on claims, but by a smaller amount.		Savings on claims exceeded payments, rendering net savings				
Scope of Adoption in Massachusetts	About 20% of providers in the BCBSMA network had entered the AQC by 2010.				33% of providers in the AQC by 2011		75% of providers in the AQC by 2012				

* All values are per enrollee per quarter. Changes in spending on claims are from a difference-in-differences regression adjusted for covariates. Negative values represent savings. Cohort averages (right columns) are scaled into a percent by dividing a cohort's average savings in the AQC by its average pre-AQC spending levels. Dollars are inflation-adjusted to 2012 U.S. dollars.

† Average savings on claims weighted across cohorts in each year, scaled into percentages by dividing into the average fee-for-service (FFS) claims costs weighted across cohorts in each year. This percentage is directly comparable to incentive payments.

‡ Incentive payments are the sum of shared savings under the budget, quality bonuses, and infrastructure bonuses. These values are expressed in percentage ranges due to the confidentiality of contracts between BCBSMA and provider organizations.

Table 8. Sensitivity Analyses, 2009 AQC Cohort vs. Control*

A. Alterations to the Statistical Model[†]

	(1) Base Model	(2) Percent Cost-sharing	(3) Omit Plan Type FE	(4) Omit State FE	(5) Omit Plan FE	(6) Omit State & Plan FE	(7) Omit Age & Sex	(8) Omit Risk Score	(9) Omit Age & Risk	(10) Prior with Plan FE
AQC*Post	-62.21***	-63.64***	-62.16***	-62.03***	-66.78***	-62.55***	-65.22***	-63.36***	-49.16***	-59.13***
	(11.12)	(11.96)	(11.12)	(11.14)	(12.53)	(12.96)	(13.43)	(16.82)	(18.06)	(15.45)
AQC	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Years	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Age	Y	Y	Y	Y	Y	Y		Y		
Sex	Y	Y	Y	Y	Y	Y		Y		
Risk	Y	Y	Y	Y	Y	Y	Y			
State FE	Y	Y	Y		Y					
Plan type	Y	Y								
Plan FE	Y		Y	Y						Y
% CS		Y								
Observations	3,715,260	3,715,048	3,715,260	3,715,260	3,715,260	3,715,260	3,715,260	3,729,885	3,729,885	3,729,885
R-squared	0.529	0.528	0.529	0.529	0.528	0.528	0.527	0.015	0.001	0.005

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

[†] These sensitivity analyses test the robustness of our main results against various changes in the model. Column 1 reproduces the main coefficient of interest (average quarterly change in spending associated with the AQC over the first 4 years of the contract, using the 2009 cohort vs. control comparison). The remaining columns show the same coefficient in alternative scenarios: (2) percent cost sharing in place of plan fixed effects; (3) exclusion of plan type fixed effects; (4-5) exclusion of state or plan fixed effects; (6) exclusion of state and plan fixed effects; (7) exclusion of age and sex; (8) exclusion of risk score; (9) exclusion of age, sex, and risk score; (10) exclusion of age, sex, and risk score with inclusion of plan fixed effects. CS is cost sharing; it is derived by calculating the percent of spending paid by the enrollee out of pocket for the 10 most frequent services and then averaging those percentages by plan. This is a reflection of plan generosity. State FE are state fixed effects. Plan FE are plan fixed effects, where the plan is a unique plan number or benefit design issued by a given insurer, rather than a unique insurer. The statistical model is described in the text of the paper.

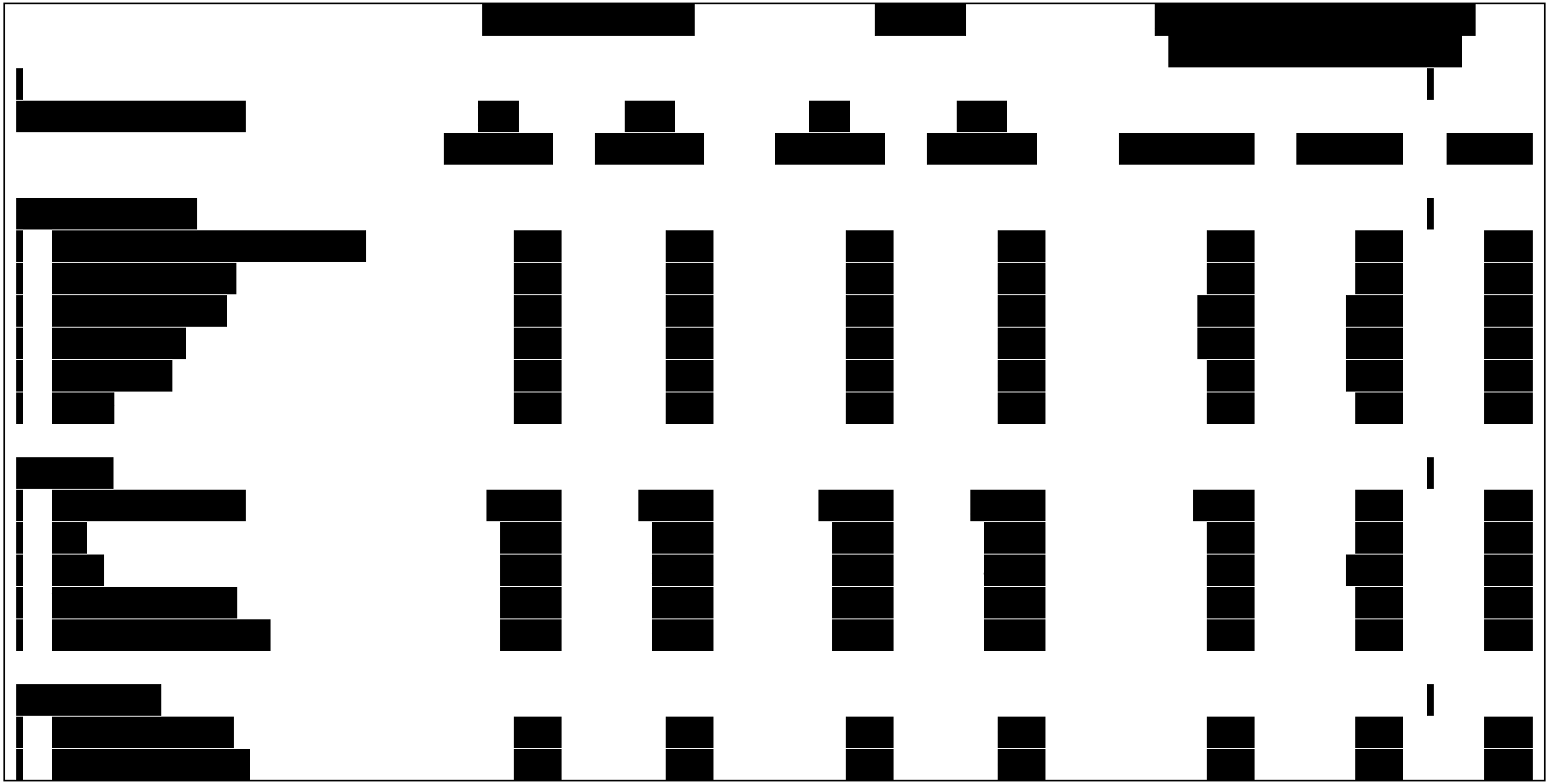
B. Alterations in Variables or Sample[††]

	(1) Base Model	(2) Risk Score Deciles	(3) Net of Cost Sharing	(4) With Rx Drugs	(5) Risk Prosp Lagged	(6) Seven-Year Continu	(7) Enrollee-Quarter Model	(8) HMO Controls Only	(9) MA Only Controls	(10) Non-AQC BCBSMA Controls	(11) Non-MA USA Controls
AQC*Post	-62.21***	-91.12***	-56.65***	-82.12***	-48.46**	-84.14**	-54.51***	-47.65***	-32.47**	-17.82***	-66.29**
	(11.12)	(11.89)	(10.84)	(14.15)	(18.52)	(17.24)	(13.11)	(11.14)	(13.41)	(6.19)	(28.66)
AQC	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Years	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Age	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sex	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Risk	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Plan type	Y	Y	Y	Y	Y	Y	Y		Y	Y	Y
Plan FE	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y
Observations	3,715,260	3,729,885	3,677,771	3,715,260-	2,483,522	599,247	14,053,576	2,330,377	2,218,738	3,514,405	7,838,725
R-squared	0.529	0.160	0.523	0.543	0.145	0.499	0.240	0.545	0.527	0.536	0.015

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

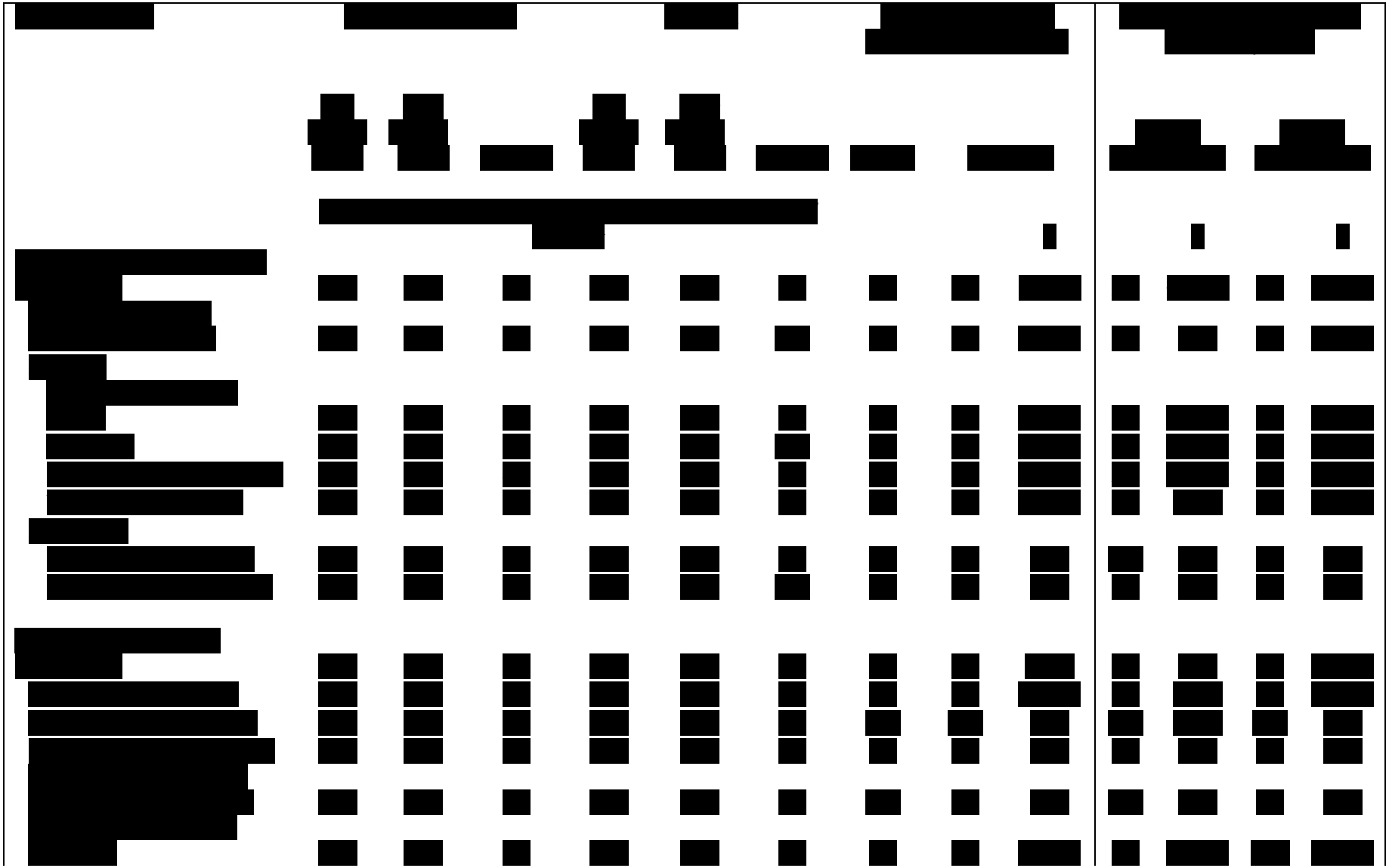
[††] These sensitivity analyses test the robustness of our main results against changes in the variables or sample. Column 1 is again the main coefficient of interest. The remaining columns show the following modifications: (2) risk scores in deciles rather than a continuous variable; (3) excluding cost sharing from spending; (4) including prescription drug spending; (5) prospective risk score lagged by 1 year; (6) restricting to continuous enrollees over 7 years during the study period; (7) quarterly model at the enrollee level. Columns (8-11) test alternative control groups that were possible to construct using the available data. These alternative control groups have drawbacks that we describe here and note in the paper. (8) HMO only controls from the 8 Northeastern states. This group fails to capture all enrollees in plans comparable to the AQC, which require designating a PCP and have incentives for receiving care in network. Also, this group had significant differences in pre-intervention spending trends compared to the AQC. (9) All Massachusetts control group. This group is not ideal because it contains BCBSMA (treatment) enrollees as well; we could not separate BCBSMA enrollees from Harvard Pilgrim, Tufts, or other private payers in MA due to the absence of payer IDs in the Truven data for confidentiality. Moreover, this control group also had significant differences in pre-intervention spending trends relative to the AQC. (10) Non-AQC BCBSMA control group (enrollees whose providers had not joined the AQC by 2012). This is not an ideal control group because the remaining providers in non-incentive contracts were small, rural practices that received lower fee updates from BCBSMA as a consequence of remaining in fee-for-service. Moreover, this control group also had significant differences in pre-intervention spending trends relative to the AQC. The Massachusetts only control groups are also susceptible to spillover effects. (11) National controls comprising a 10% random sample of enrollees in the 49 non-Massachusetts states in the Truven data. As with the main control group, national controls are susceptible to other factors in Massachusetts affecting the results, which we discuss in the paper. However, the national control group does not contaminate controls with treatment subjects and is less susceptible to AQC spillover effects within Massachusetts. Of note, similar to the baseline control group, this national control group demonstrated no significant differences in pre-intervention spending trends relative to the AQC.

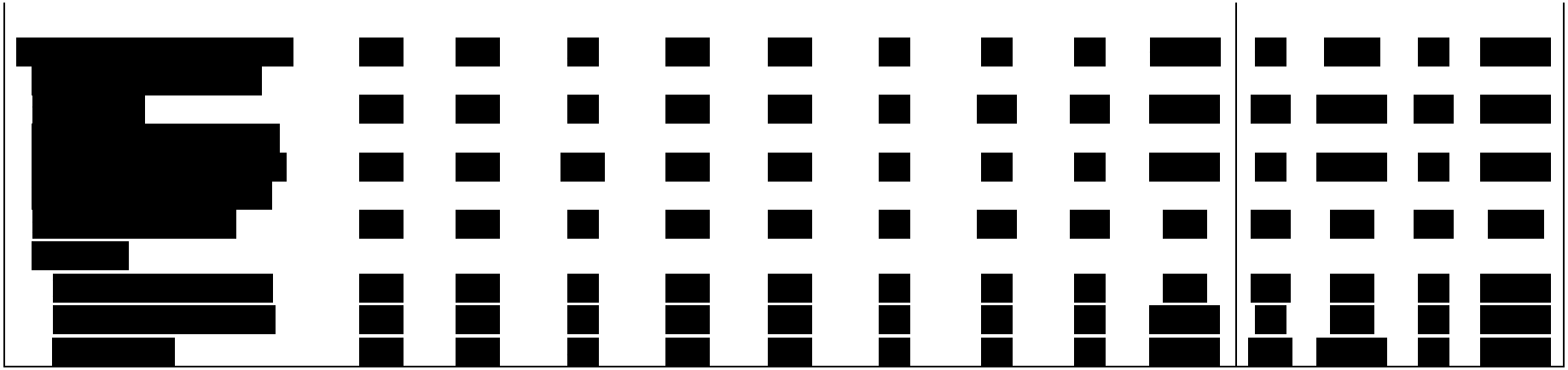
Table 9. Changes in Utilization in Treatment and Control Groups (volume per 1000 enrollees per quarter)



This table was previously published in Song Z, Fendrick AM, Safran DG, Landon B, Chernew ME. Global Budgets and Technology-Intensive Medical Services. Healthcare (Amst). 2013 Jun;1(1-2):15-21.[42]

Table 10. Change in Performance on Measures of Quality of Ambulatory Care in the 2009 AQC Cohort and Control Groups.





* Adjusted results are from a difference-in-differences multivariate model at the enrollee-year level. The intervention group was the 2009 AQC cohort. The control group comprised Blue Cross Blue Shield of Massachusetts enrollees whose primary care physicians belonged to organizations that did not enter the AQC. Pooled observations were used for the aggregate analyses of chronic care management, adult preventive care, and pediatric care. Analyses were further adjusted for measure-level fixed effects.

This table was previously published in Song Z, Safran DG, Landon BE, Landrum MB, He Y, Mechanic RE, Day MP, Chernew ME. The 'Alternative Quality Contract,' based on a global budget, lowered medical spending and improved quality. Health Aff (Millwood). 2012 Aug;31(8):1885-94.[22]

Table 11. Ambulatory Process Quality: AQC Cohorts vs. HEDIS National Average*

	AQC Cohorts			HEDIS National Average			Unadjusted Difference-in-differences
	Pre	**Post**	**Change**	**Pre**	**Post**	**Change**	
2009 AQC Cohort	**2006-08**	**2009-12**		**2006-08**	**2009-12**		*Over 4 Years*
Chronic Care Management	79.6	84.5	5.0	79.8	80.8	1.1	3.9
Adult Preventive Care	75.9	80.7	4.8	57.5	59.6	2.1	2.7
Pediatric Care	79.5	84.0	4.5	68.8	70.9	2.1	2.4
2010 AQC Cohort	**2006-09**	**2010-12**		**2006-09**	**2010-12**		*Over 3 Years*
Chronic Care Management	80.30	82.59	2.3	79.97	80.97	1.0	1.29
Adult Preventive Care	74.77	79.93	5.2	58.03	59.73	1.7	3.45
Pediatric Care	75.94	80.75	4.8	69.10	71.33	2.2	2.58
2011 AQC Cohort	**2006-10**	**2011-12**		**2006-10**	**2011-12**		*Over 2 Years*
Chronic Care Management	79.39	81.37	2.0	80.25	80.90	0.7	1.33
Adult Preventive Care	72.77	79.11	6.3	58.38	59.90	1.5	4.81
Pediatric Care	75.26	79.89	4.6	69.50	71.65	2.2	2.48
2012 AQC Cohort	**2006-11**	**2012**		**2006-11**	**2012**		*Over 1 Year*
Chronic Care Management	82.08	80.54	-1.5	80.36	81.00	0.6	-2.18
Adult Preventive Care	77.26	78.04	0.8	58.68	59.90	1.2	-0.44
Pediatric Care	80.14	81.86	1.7	69.92	71.70	1.8	-0.06

* Values designate the percent of eligible enrollees for a measure whose care achieved threshold performance for the measure. These 3 aggregate ambulatory process measures are weighted averages of individual measures in each category. Chronic Care Management measures are: cardiovascular LDL screening; hemoglobin A1c testing, eye exam, LDL screening, and nephrology screening for patients with diabetes; and short-term and maintenance prescription for patients with depression. Adult Preventive Care measures are: breast cancer, cervical cancer, and colorectal cancer screening; chlamydia screening for enrollees aged 21-24 years; and no antibiotics for acute bronchitis. Pediatric measures are: appropriate testing for pharyngitis; chlamydia screening for enrollees aged 16-20 years; no antibiotics for upper respiratory infections; and well care for babies (<15 months), children (3-6 years), and adolescents (12-21 years). All analyses are unadjusted. In other words, they are calculations based on raw weighted averages in the groups before and after their respective intervention dates.

Table 12. Outcome Quality: 2009 AQC Cohort vs. HEDIS National Average*

	2009 AQC Cohort				HEDIS National Average	
	2009	**2010**	**2011**	**2012**	**2011**	**2012**
Outcome Measures	*Percent of population achieving performance (%)*				*Percent of population achieving performance (%)*	
Diabetic HbA1c Control (≤9 percent)	80.6	82.4	83.9	84.1	71.7	71.5
Diabetic LDL Cholesterol Control (<100mg)	57.7	61.0	64.1	65.2	48.1	48.4
Diabetic Blood Pressure Control (<140/80)	51.6	54.8	60.6	65.4	44.2	44.3
LDL Cholesterol Control in Cardiovascular Patients (<100mg)	69.9	72.3	74.0	74.8	59.8	59.9
Blood Pressure Control in Cardiovascular Patients (<140/90)	68.4	71.1	78.3	80.4	65.4	63.0
Average	65.6	68.3	72.2	74.0	57.8	57.4

* Values designate the percent of eligible enrollees for a measure whose care achieved a defined threshold of quality performance for the measure. “HEDIS” is the Healthcare Effectiveness Data and Information Set. “HbA1c” is hemoglobin A1c. “LDL” is low-density lipoprotein cholesterol.

References

[1] Aaron HJ. The Central Question for Health Policy in Deficit Reduction. *N Engl J Med* 2011; 365:1655-1657.

[2] Orszag PR, Ellis P. The challenge of rising health care costs — a view from the Congressional Budget Office. N Engl J Med 2007;357:1793-5.

[3] Chernew ME, Baicker K, Hsu J. The specter of financial armageddon—health care and federal debt in the United States. *N Engl J Med*. 2010 Apr 1;362(13):1166-8.

[4] Chernew ME, Hirth RA, Cutler DM. Increased spending on health care: long- term implications for the nation. Health Aff (Millwood) 2009;28:1253-5.

[5] Fisher ES, McClellan MB, Safran DG. Building the path to accountable care. N Engl J Med. 2011;365(26): 2445–7.

[6] Berwick DM. Making good on ACOs' promise--the final rule for the Medicare shared savings program. N Engl J Med. 2011 Nov 10;365(19):1753-6.

[7] Shortell SM, Casalino LP, Fisher ES. How the center for Medicare and Medicaid innovation should test accountable care organizations. Health Aff (Millwood). 2010 Jul;29(7):1293-8.

[8] Fisher ES, Staiger DO, Bynum JP, Gottlieb DJ. Creating accountable care organizations: the extended hospital medical staff. Health Aff (Millwood). 2007;26(1):w44–57.

[9] McClellan M, McKethan AN, Lewis JL, Roski J, Fisher ES. A national strategy to put accountable care into practice. Health Aff (Millwood). 2010;29(5):982–90.

[10] Fisher ES, Shortell SM. Accountable care organizations: accountable for what, to whom, and how. JAMA. 2010 Oct 20;304(15):1715-6.

[11] Centers for Medicare and Medicaid Services. Medicare Shared Savings Program: Accountable Care Organizations. *Federal Register* 2011 Apr 7;76(67):19528-654.

[12] Muhlestein, D. Accountable Care Growth In 2014: A Look Ahead. Health Affairs Blog. 2014 Jan 29.

[13] Song Z, Landon BE. Controlling health care spending--the Massachusetts experiment. N Engl J Med. 2012 Apr 26;366(17):1560-1.

[14] Bebinger M. Mission not yet accom- plished? Massachusetts contemplates ma- jor moves on cost containment. Health Aff (Millwood) 2009;28:1373-81.

[15] Tufts Health Plan. Coordinated Care Model. 2011. Available at: http://www.tuftshealthplan.com/employers/pdfs/coordinated_care_brochure.pdf

[16] Merrell K, Berenson RA. Structuring payment for medical homes. Health Aff (Millwood) 2010;29:852-8.

[17] Kilo CM, Wasson JH. Practice rede- sign and the patient-centered medical home: history, promises, and challenges. Health Aff (Millwood) 2010;29:773-8.

[18] Rittenhouse DR, Shortell SM. The pa- tient-centered medical home: will it stand the test of health reform? JAMA 2009;301: 2038-40.

[19] Bitton A, Martin C, Landon BE. A nationwide survey of patient centered medical home demonstration projects. J Gen Intern Med. 2010 Jun;25(6):584-92.

[20] Song Z, Safran DG, Landon BE, He Y, Ellis RP, Mechanic RE, Day MP, Chernew ME. Health care spending and quality in year 1 of the alternative quality contract. N Engl J Med. 2011 Sep 8;365(10):909-18.

[21] Mechanic RE, Santos P, Landon BE, Chernew ME. Medical group responses to global payment: early lessons from the "Alternative Quality Contract" in Massachusetts. Health Aff (Millwood). 2011;30(9):1734–42.

[22] Song Z, Safran DG, Landon BE, Landrum MB, He Y, Mechanic RE, Day MP, Chernew ME. The 'Alternative Quality Contract,' based on a global budget, lowered medical spending and improved quality. Health Aff (Millwood). 2012 Aug;31(8):1885-94.

[23] McWilliams JM, Landon BE, Chernew ME. Changes in health care spending and quality for Medicare beneficiaries associated with a commercial ACO contract. JAMA. 2013 Aug 28;310(8):829-36.

[24] Chien AT, Song Z, Chernew ME, Landon BE, McNeil BJ, Safran DG, Schuster MA. Two-year impact of the alternative quality contract on pediatric health care quality and spending. Pediatrics. 2014 Jan;133(1):96-104.

[25] Center for Medicare and Medicaid Services. Pioneer Accountable Care Organizations succeed in improving care, lowering costs. 2013 Jul 16. (http://www.cms.gov/Newsroom/MediaReleaseDatabase/Press-Releases/2013-Press-Releases-Items/2013-07-16.html)

[26] Song Z. Pioneer Accountable Care Organizations: Lessons from Year 1. In Gregory Curfman, Stephen Morrissey, Gardiner Morse, and Steve Prokesch (eds.), Insight Center: Leading Health Care Innovation. New England Journal of Medicine/Harvard Business Review. 2013 Oct 8. (http://images.nejm.org/editorial/supplementary/2013/hbr06-song.pdf)

[27] Center for Medicare and Medicaid Services. Medicare's delivery system reform initiatives achieve significant savings and quality improvements - off to a strong start. 2014 Jan 30. (http://www.hhs.gov/news/press/2014pres/01/20140130a.html)

[28] Epstein AM, Jha AK, Orav EJ, Liebman DL, Audet AM, Zezza MA, Guterman S. Analysis of early accountable care organizations defines patient, structural, cost, and quality-of-care characteristics. Health Aff (Millwood). 2014 Jan;33(1):95-102.

[29] McWilliams JM, Chernew ME, Zaslavsky AM, Hamed P, Landon BE. Delivery system integration and health care spending and quality for Medicare beneficiaries. JAMA Intern Med. 2013 Aug 12;173(15):1447-56.

[30] Hansen LG, Chang S. Health Research Data for the Read World: the Thomson Reuters MarketScan Databases. Thomson Reuters White Paper, 2010 Jan. (http://MarketScan.thomsonreuters.com/MarketScanportal/)

[31] Verisk Health. DxCG Risk Solutions: DxCG Science Guide. Verisk Health, Inc. Waltham, MA, 2012.

[32] Pope GC, Kautter J, Ellis RP, Ash AS, Ayanian JZ, Iezzoni LI, Ingber MJ, Levy JM, and Robst J. Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC model. Health Care Financing Review 2004;25(4):119-41.

[33] Wooldridge J. Econometric Analysis of Cross Section and Panel Data. MIT Press: Cambridge, 2001.

[34] Centers for Medicare and Medicaid Services. Berenson-Eggers Type of Service, 2010. Available at: https://www.cms.gov/HCPCSReleaseCodeSets/20_BETOS.asp.

[35] Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. J Health Econ 2005;24(3):465-88.

[36] Zaslavsky AM, Buntin MB. Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures. J Health Econ 2004;23:525-42.

[37] Ai C, Norton EC. Interaction terms in logit and probit models. Economic Letters 2003;80:123-9.

[38] Ellis RP, McGuire TG. Predictability and predictiveness in health care spending. J Health Econ 2007;26(1):25-48.

[39] White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 1980;48(4):817-30.

[40] Huber PJ. The behavior of maximum likelihood estimates under non-standard conditions. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1967:221-33.

[41] Medicare Payment Advisory Commission. Report to the Congress: improving incentives in the Medicare program. June 2009. (http://www.medpac.gov/documents/ Jun09_EntireReport.pdf.)

[42] Song Z, Fendrick AM, Safran DG, Landon BE, Chernew BE. Global Budgets and Technology-Intensive Medical Services. Health Care: The Journal of Delivery Science and Innovation. 2013;1(1-2):15-21.

[43] McWilliams JM, Landon BE, Chernew ME. Changes in health care spending and quality for Medicare beneficiaries associated with a commercial ACO contract. JAMA. 2013 Aug 28;310(8):829-36.

[44] 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 Nov 11;173(20):1907-13.

[45] 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 Feb 26;311(8):815-25.

[46] Rajkumar R, Patel A, Murphy K, Colmers JM, Blum JD, Conway PH, Sharfstein JM. Maryland's all-payer approach to delivery-system reform. N Engl J Med. 2014 Feb 6;370(6):493-5.

[47] Ayanian JZ. Michigan's approach to Medicaid expansion and reform. N Engl J Med. 2013 Nov 7;369(19):1773-5.

[48] Wroten D. Arkansas payment reform... what happens now? J Ark Med Soc. 2012 Aug;109(2):28.

[49] Stecker EC. The Oregon ACO experiment--bold design, challenging execution. N Engl J Med. 2013 Mar 14;368(11):982-5.

[50] Silow-Carroll S, Edwards JN, Rodin D. How Colorado, Minnesota, and Vermont are reforming care delivery and payment to improve health and lower costs. Issue Brief (Commonw Fund). 2013 Mar;10:1-9.

[51] Song Z, Lee TH. The era of delivery system reform begins. JAMA. 2013 Jan 2;309(1):35-6.

[52] Lee TH, Bothe A, Steele GD. How Geisinger structures its physicians' compen- sation to support improvements in quality, efficiency, and volume. Health Aff (Millwood). 2012;31(9):2068-2073.

[53] Lee TH. Care redesign: a path forward for providers. N Engl J Med. 2012; 367(5):466-472.

[54] Landon BE, Keating NL, Barnett ML, et al. Variation in patient-sharing net- works of physicians across the United States. JAMA. 2012;308(3):265-273.

[55] Kullgren JT, Troxel AB, Loewenstein G, Asch DA, Norton LA, Wesby L, Tao Y, Zhu J, Volpp KG. Individual- versus group-based financial incentives for weight loss: a randomized, controlled trial. Ann Intern Med. 2013 Apr 2;158(7):505-14.

[56] Volpp KG, Asch DA, Galvin R, Loewenstein G. Redesigning employee health incentives--lessons from behavioral economics. N Engl J Med. 2011 Aug 4;365(5):388-90.