The 'Alternative Quality Contract' in Massachusetts, Based on Global Budgets, Lowered Medical Spending and Improved Quality

Zirui Song, Ph.D.[1,2], **Dana Gelb Safran, Sc.D.**[3,4], **Bruce E. Landon, M.D., M.B.A., M.Sc.**[1,5], **Mary Beth Landrum, Ph.D.**[1], **Yulei He, Ph.D., M.S.**[1], **Robert E. Mechanic, M.B.A.**[6], **Matthew P. Day, F.S.A., M.A.A.A**[3], and **Michael E. Chernew, Ph.D.**[1]

[1]Department of Health Care Policy, Harvard Medical School

[2]National Bureau of Economic Research

[3]Blue Cross Blue Shield of Massachusetts

[4]Department of Medicine, Tufts University School of Medicine

[5]Division of General Medicine and Primary Care, Beth Israel Deaconess Medical Center

[6]Heller School for Social Policy & Management, Brandeis University

Abstract

Seven provider organizations in Massachusetts entered the Blue Cross Blue Shield Alternative Quality Contract in 2009, followed by four more organizations in 2010. This contract, based on a global budget and pay-for-performance for achieving certain quality benchmarks, places providers at risk for excessive spending and rewards them for quality, similar to the new Pioneer Accountable Care Organizations in Medicare. We analyzed changes in spending and quality associated with the Alternative Quality Contract and found that the rate of increase in spending slowed compared to control groups. Overall, participation in the contract over two years led to a savings of 3.3% (1.9% in year-1, 3.3% in year-2) compared to spending in groups not participating in the contract. The savings were even higher for groups whose previous experience had been only in fee-for-service contracting. Such groups' quarterly savings over two years averaged 8.2% (6.3% in year-1, 9.9% in year-2). Quality of care also improved within organizations participating in the Alternative Quality Contract compared to control organizations in both years. Chronic care management, adult preventive care, and pediatric care improved from year 1 to year 2 within the contracting groups. These results suggest that global budgets coupled with pay-for-performance can begin to slow the underlying growth in medical spending while improving quality.

Amid mounting federal debt, slowing the growth of health care spending is a national priority.[1,2] Much policy attention has been focused on using global budgets within accountable care organizations because these kinds of payments have the potential to lower spending while improving the quality of care. A global budget is a prospective reimbursement to a health care provider, such as a physician or hospital, that reflects the total expected spending of its patient population over the continuum of care for a defined period of time.

Address correspondence to Zirui Song, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115. Phone: 617-432-1294, Fax: 617-432-0173, zirui_song@hms.harvard.edu.

Through the Medicare Shared Savings Program established under the Affordable Care Act,[3] provider groups can become accountable care organizations by choosing a "one-sided" model in which a group shares savings with Medicare if the group's spending is below its prespecified target. Provider groups can also choose a "two-sided" model in which they share savings but also assume risk for excess spending over their targets--in which case a target is analogous to a global budget. .[4] Both models reward providers for meeting quality benchmarks.

In January 2012, thirty-two organizations across the country began Pioneer Accountable Care Organization contracts through the Center for Medicare and Medicaid Innovation, which uses the two-sided model with greater risk and greater reward. Outside of Medicare, about 100 provider organizations have worked with private insurers to implement similar contracts.[5] In Massachusetts, lawmakers are proposing legislation to expand global payment throughout the state.[6] Despite this momentum, however, there is limited evidence on the effects of global payment within accountable care organization contracts on health care spending and quality.[5,7]

In 2009, Blue Cross Blue Shield of Massachusetts entered into global budget accountable care organization agreements with seven provider organizations in the state via the Alternative Quality Contract.[8] In the first year of implementation, the Alternative Quality Contract was associated with a 1.9 percent lower increase in total medical spending and modest quality improvements, compared to control groups.[9] These initial savings were largely achieved through shifting referrals to less expensive providers and settings rather than reductions in utilization[10]--a strategy that is not likely to achieve substantial additional savings year after year.

Four additional organizations joined the Alternative Quality Contract in 2010, bringing total participation to more than 1,600 primary care physicians and 3,200 specialists. Organizations ranged from large physician-hospital organizations to groups of small practices (Appendix Exhibit 1).[11] The contract pays providers a global budget that covers the entire continuum of care for a defined population of enrollees each year. Many analysts believe that global payment with financial risk provides more powerful incentives to control spending than the one-sided "shared savings" payment model based on a spending target.[12–14] In addition, the Alternative Quality Contract rewards provider groups up to 10 percent of their global budget for meeting a set of sixty-four quality measures (Appendix Exhibit 2),[11] which broadly overlaps with the thirty-three quality measures used by Medicare.[15]

The Alternative Quality Contract is a five-year contract that currently applies primarily to enrollees in health maintenance organization plans. These members are required to designate a primary care physician prior to each enrollment year. Thus, an enrollee is "in" the Alternative Quality Contract if his or her primary care physician belongs to an organization that has joined the contract. This works in a way similar to patient-centered medical homes, in which patients are attributed to a primary care physician.[16–19] Blue Cross Blue Shield of Massachusetts provides organizations in the Alternative Quality Contract with technical support, such as ongoing quality and spending data, to assist them in managing their global budgets and improving quality.

In this study, we evaluated the effect of the contract on spending and quality after two years. We also analyzed its first-year effect on the four organizations that entered the contract in 2010. We report on the study's results in this article.

Study Data And Methods

Study Population

Our population included Blue Cross Blue Shield of Massachusetts enrollees from January 2006 through December 2010 who were continuously enrolled for at least one calendar year. The 2009 cohort consisted of enrollees whose primary care physician joined the Alternative Quality Contract in 2009. The 2010 cohort consisted of enrollees whose primary care physician joined in 2010. The control group consisted of enrollees whose primary care physician did not enter the contract.

Study Design

We used a difference-in-differences approach to isolate the effect of the Alternative Quality Contract on spending. For the 2009 cohort, the preintervention period was 2006–08 and the postintervention period was 2009–10. Within this cohort, we prespecified two subgroups. The "prior-risk" subgroup consisted of four organizations that had prior experience managing risk-based contracts from Blue Cross Blue Shield of Massachusetts. Enrollees from these organizations constituted 88 percent of the 2009 cohort. The three remaining organizations were placed in the "no-prior-risk" subgroup, which meant they had previously been fee-for-service groups. Enrollees in this subgroup constituted the remaining 12 percent of the 2009 cohort.

We decomposed the average two-year effect on spending into year 1 and year 2 effects. We also studied the year 1 effect on spending in the 2010 cohort. All four organizations in the 2010 cohort entered the contract without risk-contracting experience. Thus, we compared the 2010 cohort's year 1 finding to the analogous year 1 finding in the no-prior-risk subgroup of the 2009 cohort.[9]

We also decomposed the spending effect by clinical category, as defined by the Berenson-Eggers Type of Service classification system;[20] site and type of care; and enrollees' health status using the health risk score.[21] Finally, we decomposed the spending result into a price effect and a utilization (volume) effect. We first repriced claims for each service to their median prices across all providers in 2006–10. Spending results generated using repriced claims reflected only differences in utilization. We further decomposed the price effect into two price-related explanations: differential fee changes between treatment groups (those in the contract) and control groups; and referral pattern changes--enrollees in the contract groups referred to lower-price providers relative to those in the control groups. This was done by repricing claims to the median 2010 price for each service within each practice.

For quality, we compared performance on ambulatory--nonhospital, primary care--process measures using a similar difference-in-differences approach with 2007–10 enrollee-level data. These were primary care-oriented measures under the direct control of providers, based on the Healthcare Effectiveness Data and Information Set (HEDIS) measures used by most health plans. We analyzed aggregate as well as individual measures within chronic care management, adult preventive care, and pediatric care.

Variables

For our spending analysis, the dependent variable was total enrollee medical spending, including enrollees' cost sharing. Spending was computed from claims-level fee-for-service payments made within the global budget. This is an accurate measure of medical spending based on utilization and negotiated fees, but it does not capture the quality bonuses or end-of-year budget reconciliation. For quality, the dependent variable was a dichotomous variable indicating whether the measure was met for an eligible member. *Eligibility* was

defined by diagnoses; for example, only members with diabetes were eligible for diabetes measures. Quality performance was determined annually.

We controlled for age categories, interactions between age and sex, risk score, indicators for benefit design, and secular trends. Risk scores were calculated by Blue Cross Blue Shield of Massachusetts from current-year demographic information and diagnoses grouped by episodes of care--in a manner conceptually similar to the risk-adjustment method used by the Centers for Medicare and Medicaid Services for determining payments to Medicare Advantage plans.[22] Higher scores denote lower health status and higher expected spending.

Statistical Analysis

We analyzed spending (in 2010 dollars) at the enrollee-quarter level using a multivariate linear model with plan fixed effects and propensity weights calculated using age, sex, risk, and cost sharing.[23] Our model was not logarithmic-transformed because the risk score is designed to predict dollar spending and because linear models have been shown to better predict health spending than more complex functional forms.[24–27]

Additional independent variables included indicators for intervention status, quarter, interactions between quarters and intervention, the postintervention period, and the interaction between the postintervention period and the intervention. This final indicator produced our estimate of the policy effect. Standard errors were clustered at the practice level.[28,29]

We conducted a number of sensitivity analyses, including restricting the sample to five-year continuously enrolled subjects, controlling for risk or benefit design differently, and omitting propensity weights or cost sharing. Because 20 percent of enrollees did not have prescription drug coverage through the payer, we excluded drug spending from our base analyses but included it in a sensitivity analysis. We also performed an interaction test of the differential effect of the intervention between prior-risk and no-prior-risk groups.

Because patients with higher risk scores garner larger global payments, we also tested for the possibility that providers may upcode, or charge for a diagnosis or procedure more complex than what actually exists in order to garner higher payment. Upcoding would make intervention subjects seem sicker and thus spending adjusted for health status seem lower.

We used analogous models to evaluate quality at the enrollee-year level, with year indicators in place of quarter indicators. For our aggregate quality analysis, we pooled measures and adjusted for measure-level fixed effects. All analyses used the statistical analysis software STATA, version 11. Results are reported with two-tailed *p* values. The Harvard Medical School Office for Research Subject Protection approved the study.

Limitations

Given that subjects were enrolled in commercial plans in Massachusetts, our results might not generalize to other populations and areas. Commercial enrollees are younger and generally healthier than the Medicare population. Nevertheless, the two-sided contracting models used by the Medicare Pioneer and Shared Savings programs are similar to those used by the Alternative Quality Contract.

Because the intervention was not randomized, self-selection of provider groups into the contract was a concern. Although we did not have data on providers, we tested for differences in preintervention spending trends between intervention and control groups, which may suggest a self-selection bias; we found none.

Moreover, we did not observe details of each contract or provider risk contracting with other payers. The prevalence of global payment has grown in Massachusetts,[6] and contracts with other payers may exert unobserved spillover effects in either our treatment or control groups. Lastly, our process quality measures did not capture all aspects of health care quality.

Study Results

2009 Intervention Cohort: Year 1 And Year 2 Spending Effects

There were 428,892 subjects with at least one year of continuous enrollment during 2006–10 in the 2009 cohort and 1,339,798 control subjects. Characteristics of the population are shown in Exhibit 1.

After implementation of the Alternative Quality Contract, average health care spending increased for both intervention and control enrollees, but the increase was smaller for intervention enrollees. Overall in 2009–10, statistical estimates indicated that the intervention was associated with a $22.58 ($p$=0.04) decrease in average spending per enrollee per quarter relative to what spending would have been without the intervention (Exhibit 2). This amounted to a 2.8 percent average savings over two years. Our prior analyses showed that savings in year 1 (2009) were 1.9 percent--a reduction of $15.51 in quarterly spending per enrollee (p=0.009).[9] By comparison, in year 2 (2010) the intervention was associated with a $26.72 ($p$=0.04) reduction in average quarterly spending--a 3.3 percent savings. The increased savings in year 2 were unrelated to the removal of the 2010 cohort from the control group in year 2 (the 2010 cohort belonged to the control group in year 1). The interaction of the secular trend with the intervention indicator produced a small and statistically insignificant coefficient ($0.62, p=0.65), which meant there were no significant differences in spending trends between intervention and control groups prior to the intervention.

Savings were substantially larger in the no-prior-risk subgroup, whose providers entered the contract from fee-for-service. Over the first two years, the no-prior-risk subgroup demonstrated an average reduction of $60.75 ($p$=0.02) in quarterly spending--an 8.2 percent savings. Broken down by year, its year 1 savings were 6.3 percent (p=0.006) and year 2 savings were 9.9 percent (p=0.003) (Exhibit 3). In contrast, the prior-risk subgroup saw smaller and statistically insignificant reductions of $13.42 ($p$=0.19) per member per quarter over the postintervention period. By year, reductions were 1.1 percent in year 1 (p=0.13) and 1.8 percent in year 2 (p=0.23).

Throughout 2009–10, savings were concentrated in procedures, imaging, and tests. Moreover, reductions in outpatient facility spending accounted for 75 percent of the savings. The largest reductions in spending were found in enrollees in the highest risk quartile (Appendix Exhibit 3).[11] Analogous decompositions of savings in both subgroups were consistent (Appendix Exhibit 4).[11]

All sensitivity analyses supported our main results (Appendix Exhibit 5).[11] An interaction test of the differential treatment effect between the prior-risk and no-prior-risk subgroups yielded −$22.58 ($p$=0.04). Furthermore, the Alternative Quality Contract was associated with a 0.02 (p=0.002) annual rise in enrollee risk scores. This effect can explain about 5 percent of the intervention's effect on spending. However, we cannot distinguish between a true increase in risk and an upcoding effect due to incentives, to accurately record all diagnoses.

2010 Intervention Cohort: Year 1 Effect

The 2010 cohort consisted of 183,655 enrollees whose characteristics were similar to those of the 2009 cohort (Exhibit 1). In its first year, the 2010 cohort--all no-prior-risk organizations--experienced a decrease of $42.82 ($p$=0.001), or 5.5 percent, in average quarterly spending. This is similar to the 6.3 percent year 1 savings of the 2009 cohort's no-prior-risk subgroup (Exhibit 3). Decomposition of the 2010 cohort's year 1 savings produced similar findings as in the 2009 cohort's no-prior-risk subgroup (Appendix Exhibit 6).[11]

Price Versus Utilization

Estimates from the 2010 cohort revealed that reductions in utilization relative to the control group accounted for about 50 percent of its savings in year 1. This result was consistent with findings from the 2009 cohort's no-prior-risk subgroup. However, we found no statistically significant changes in utilization in the prior-risk subgroup, which constituted the bulk of the 2009 cohort. This was consistent with the insignificant spending effect in this subgroup.

Direct analyses of utilization by clinical category showed heterogeneous effects on utilization. Some suggestive evidence of substitution from higher- to lower-cost services in areas such as imaging and tests was evident. However, we were not able, using claims data, to evaluate implications for the value or appropriateness of services due to these changes.

We found no differences in price trends between intervention and control providers. Models with prices standardized by practice revealed that the price effect was largely due to referrals of intervention subjects to facilities and providers charging lower fees (Appendix Exhibit 7).[11]

Quality

Improvements in ambulatory care quality measures associated with the intervention were larger in year 2 than in year 1. After two years, the contract was associated with a 3.7-percentage-point increase per member per year among eligible members meeting quality performance targets for chronic care management (p<0.001) (Exhibit 4). Broken down by year, we estimated a 4.7-percentage-point increase in year 2 (p<0.001) and a 2.6-percentage-point increase in year 1 (p<0.001).

For adult preventive care, statistical estimates showed a 0.3-percentage-point (p=0.008) yearly improvement. Year 2 improvement was 0.7 percentage points (p<0.001), while that of year 1 was an insignificant 0.1 percentage points (p=0.67).

Ambulatory measures for pediatric care improved 1.3 percentage points (p<0.001) per year over the two years, with similarly larger effects in year 2 (1.9 percentage points, p<0.001) than in year 1 (0.7 percentage points, p=0.001). Analogous estimates among the two subgroups and the 2010 cohort yielded similar results (not shown).

Formal evaluation of outcome quality measures could not be conducted because of the lack of preintervention enrollee-level outcomes data. However, an unadjusted analysis of weighted averages for five outcome metrics across provider organizations suggests that intervention groups achieved better or comparable outcomes in 2009–10 relative to recent Blue Cross Blue Shield of Massachusetts network averages (Exhibit 5).

Blue Cross Blue Shield Payments

Our findings do not imply that overall spending fell for Blue Cross Blue Shield of Massachusetts in 2009–10. Ten of the eleven organizations in the contract spent below their

2010 budget targets, earning a budget surplus payment. In general, surplus payments earned by each cohort were no more than half of the medical savings generated by the cohort. An exception to this was the 2009 "prior risk" cohort, for which total payments exceeded savings generated. All organizations earned a 2010 quality bonus, and most received infrastructure support, which makes it likely that total Blue Cross Blue Shield payments to groups in 2010 exceeded medical savings achieved by the group that year.

Discussion

In year 2, the Alternative Quality Contract was associated with increased slowing of medical spending and larger improvements in ambulatory care quality compared to year 1. Outpatient facility spending on procedures, imaging, and tests accounted for the bulk of savings. In both years, savings were largest among enrollees with the highest expected spending and were concentrated in provider organizations without risk-contracting experience. Results were robust to sensitivity analyses and did not appear to be attributable to upcoding. Lower prices from referral-pattern changes drove the spending slowdown overall. However, lower utilization explained about half of the savings among enrollees of providers without risk-contracting experience. These results contrast with findings from Medicare's Physician Group Practice Demonstration, in which provider organizations in one-sided shared savings contracts improved quality but generally did not lower spending.[7,30]

In year 1, total Blue Cross Blue Shield payouts to groups in the contract likely exceeded savings under the global budget.[9] In year 2, savings achieved by the intervention group were generally larger than the surplus payments received. However, total payments to groups, including surplus sharing, quality bonuses, and infrastructure support, probably exceeded the savings achieved by most groups that year. This reflects the design of the contract, which set targets based on actuarial projections to save money over its five-year duration, accounting for anticipated quality bonuses and other payments.

In addition, health care spending growth in Massachusetts slowed over this time period as a result of general economic factors. These were probably not anticipated by Blue Cross Blue Shield when the contracts were signed and budget targets for the 2009–10 period were agreed to. Initial investments by Blue Cross Blue Shield probably helped motivate participation and support the changes required for providers to succeed in managing spending and improving quality. The long-term success of the model will depend both on how well budgets and bonuses are set and how well groups are able to allocate resources and improve quality within budgets that grow more slowly each year.

The increased slowing of spending from year 1 to year 2 suggests that global budgets may be an effective tool to use in helping control health care spending, but also that organizations need time to implement changes. At the same time, improvements in quality supported by sizable pay-for-performance incentives may effectively buffer against stinting on care. Collectively, global payment and pay-for-performance may provide a palatable set of incentives for providers groups to participate in delivery system reforms that encourage accountability and reduce waste. Successful transitions to practice models that support global budgets will require providers to protect quality and patient satisfaction.

This model is informative for the broader movement towards accountable care organizations.[3,31–33] Of the thirty-two Medicare Pioneer Accountable Care Organizations that began three-year risk contracts in 2012, five also participated in the Alternative Quality Contract. Our findings suggest that changes in utilization are possible in the early years of a global payment contract for participants entering from fee-for-service, whereas savings

among providers with risk-contracting experience derived mostly from referrals to less costly providers. Medicare, which regulates prices, may be able to achieve savings only through changes in utilization. Thus, incentives for Medicare beneficiaries and their providers to lower volume without sacrificing quality are key. As global payment contracts expand across the country,[34,35] supportive partnerships between payers and provider groups may help providers take accountability for spending while improving the quality of care for patients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Biographies

Zirui Song is a medical degree candidate at Harvard Medical School. His research focuses on the effect of payment reform on health care spending and quality in Massachusetts, Medicare fee policy, and the competitive bidding system in Medicare Advantage. Song is a member of the Task Force on Health Care Reform for the Massachusetts Medical Society and the Boston Advisory Board of the Albert Schweitzer Fellowship. He earned his doctorate in health policy (economics track) from Harvard University and was a predoctoral fellow in Aging and Health Economics at the National Bureau of Economic Research.

Dana Gelb Safran is the senior vice president for performance measurement and improvement at Blue Cross Blue Shield of Massachusetts (BCBSMA) and an associate professor of medicine at the Tufts University School of Medicine. She was among the lead developers of the BCBSMA Alternative Quality Contract, a provider contract model launched in 2009 with the twin goals of improving quality and outcomes, while significantly slowing spending growth. Safran earned her doctoral and master's degrees in health policy from the Harvard University School of Public Health.

Bruce Landon is a professor of health care policy and medicine at Harvard Medical School. He also practices medicine at Beth Israel Deaconess Medical Center. Landon's primary research interest has been assessing the impact of different characteristics of physicians and health care organizations on the provision of health care services. He earned his medical degree from the University of Pennsylvania School of Medicine; a master's degree in health policy from the Harvard School of Public Health; and an MBA, with a concentration in health care management, from the University of Pennsylvania Wharton School.

Mary Beth Landrum is an associate professor of health care policy, with a specialty in biostatistics, in the Harvard Medical School Department of Health Care Policy. Her primary research focus is on the development and application of statistical methodology for human

services research. This research has several related themes, including the development of medical guidelines. Landrum has developed and applied methodology for profiling providers—in particular, methods for assessing providers on more than one dimension of care. She earned her doctoral and master's degrees in biostatistics from the University of Michigan.

Yulei He is an assistant professor of health care policy, with a specialty in biostatistics, in the Harvard Medical School Department of Health Care Policy. His research focuses on the development and application of statistical methods for health services and policy research. He is a member of the Statistical Coordinating Center of CanCORS (Cancer Care Outcomes, Research, and Surveillance), a multisite study sponsored by the National Cancer Institute to study the quality and patterns of cancer care and their subsequent impact on patient outcomes. He earned a doctorate in biostatistics from the University of Michigan and a master's degree in public health from the University of Illinois at Chicago.

Robert Mechanic is a senior fellow at the Heller School for Social Policy and Management at Brandeis University and executive director of the Health Industry Forum, a national organization devoted to developing strategies to improve the quality and effectiveness of US health care. Mechanic previously ran a strategic health care consultancy; worked with the Massachusetts Hospital Association; and was a vice president of the Lewin Group, where his practice focused on hospital finance, state health policy, and health care reform. Mechanic holds an MBA in finance from the University of Pennsylvania Wharton School.

Matthew Day is the vice president and senior actuary for provider financial management at BCBSMA. He is responsible for the overall financial management and direction of its physician, hospital, ancillary, and global payment provider contracts. Day holds a bachelor's degree in mathematics and economics from Fairfield University.

Michael Chernew is a professor of health care policy in the Department of Health Care Policy at Harvard Medical School and serves as the director of Harvard's program for value-based insurance design. His research examines several areas related to controlling health care spending growth, while maintaining or improving the quality of care. He is the coeditor of the *American Journal of Managed Care* and senior associate editor of *Health Services Research*. Chernew has served on the editorial boards of *Health Affairs* and *Medical Care Research and Review*. Chernew earned a doctorate in economics from Stanford University.

Notes

1. Aaron HJ. The central question for health policy in deficit reduction. N Engl J Med. 2011; 365(18): 1655–1657. [PubMed: 21995336]

2. 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; 362(13):1166–1168. [PubMed: 20237338]

3. Centers for Medicare and Medicaid Services. Medicare Shared Savings Program: accountable care organizations. Fed Regist. 2011 Apr 7; 76(67):19528–19654.

4. Berwick DM. Making good on ACOs' promise--the final rule for the Medicare Shared Savings Program. N Engl J Med. 2011; 365(19):1753–1756. [PubMed: 22013899]

5. Fisher ES, McClellan MB, Safran DG. Building the path to accountable care. N Engl J Med. 2011; 365(26):2445–2447. [PubMed: 22204720]

6. Song Z, Landon BE. Controlling health care spending--the Massachusetts experiment. N Engl J Med. 2012; 366(17):1560–1561. [PubMed: 22494111]

7. Wilensky GR. Lessons from the Physician Group Practice Demonstration--a sobering reflection. N Engl J Med. 2011; 365(18):1659–1661. [PubMed: 21916636]

8. Chernew ME, Mechanic RE, Landon BE, Safran DG. Private-payer innovation in Massachusetts: the "Alternative Quality Contract". Health Aff (Millwood). 2011; 30(1):51–61. [PubMed: 21209437]

9. Song Z, Safran DG, Landon BE, He Y, Ellis RP, Mechanic RE, et al. Health care spending and quality in year 1 of the alternative quality contract. N Engl J Med. 2011; 365(10):909–918. [PubMed: 21751900]

10. 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):1734–1742. [PubMed: 21900665]

11. To access the Appendix, click on the Appendix link in the box to the right of the article online.

12. Rosenthal MB, Cutler DM, Feder J. The ACO rules--striking the balance between participation and transformative potential. N Engl J Med. 2011; 365(4):e6. [PubMed: 21751898]

13. Engelberg Center for Health Care Reform. Washington (DC): Brookings Institution; 2009 Aug. Bending the curve: effective steps to address long-term health care spending growth [Internet]. Available from: http://www.brookings.edu/research/reports/2009/09/01-btc

14. Miller HD. From volume to value: better ways to pay for health care. Health Aff (Millwood). 2009; 28(5):1418–1428. [PubMed: 19738259]

15. Centers for Medicare and Medicaid Services. [cited 2012 June 1] 2012 ACO Narrative Quality Measures Specifications Manual [Internet]. 2011 Dec 12. Available from: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Quality_Measures_Standards.html

16. Bodenheimer T, Grumbach K, Berenson RA. A lifeline for primary care. N Engl J Med. 2009; 360(26):2693–2696. [PubMed: 19553643]

17. Rittenhouse DR, Shortell SM. The patient-centered medical home: will it stand the test of health reform? JAMA. 2009; 301:2038–2040. [PubMed: 19454643]

18. Merrell K, Berenson RA. Structuring payment for medical homes. Health Aff (Millwood). 2010; 29(5):852–858. [PubMed: 20439871]

19. Kilo CM, Wasson JH. Practice redesign and the patient-centered medical home: history, promises, and challenges. Health Aff (Millwood). 2010; 29(5):773–778. [PubMed: 20439860]

20. Centers for Medicare and Medicaid Services. Berenson-Eggers Type of Service (BETOS) [Internet]. Baltimore (MD): CMS; The codes for 2010 are available for download from: https://www.cms.gov/Medicare/Coding/HCPCSReleaseCodeSets/BETOS.html

21. The health risk score is calculated using a risk-adjustment method similar to the diagnostic-cost-group (DxCG) scoring system (Verisk Health), which uses statistical analyses based on a national claims database to relate current-year spending to current-year diagnoses and demographic information. The DxCG method is a commonly used proprietary method similar to Medicare's Hierarchical Condition Category (HCC) system, which is used for risk adjustment of prospective payments to Medicare Advantage plans (and was developed by the same organization). DxCGs are designed for people younger than age sixty-five and are more detailed than the HCC system. Among all of the subjects, the mean risk score ± SD was 1.11 ± 1.93, and the median was 0.49.

22. Pope GC, Kautter J, Ellis RP, Ash AS, Ayanian JZ, Iezzoni LI, et al. Risk adjustment of Medicare capitation payments using the CMS-HCC model. Health Care Financ Rev. 2004; 25(4):119–141. [PubMed: 15493448]

23. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983; 70(1):41–55.

24. Zaslavsky AM, Buntin MB. Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures. J Health Econ. 2004; 23:525–542. [PubMed: 15120469]

25. Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. J Health Econ. 2005; 24(3):465–488. [PubMed: 15811539]

26. Ellis RP, McGuire TG. Predictability and predictiveness in health care spending. J Health Econ. 2007; 26(1):25–48. [PubMed: 16908083]

27. Ai C, Norton EC. Interaction terms in logit and probit models. Economics Letters. 2003; 80:123–129.

28. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica. 1980; 48:817–830.

29. Huber, PJ. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley (CA): University of California Press; 1967. The behavior of maximum likelihood estimates under non-standard conditions; p. 221-233.

30. Medicare Payment Advisory Commission. Washington (DC): MedPAC; 2009 Jun. Report to the Congress: improving incentives in the Medicare program.

31. Fisher ES, Staiger DO, Bynum JP, Gottlieb DJ. Creating accountable care organizations: the extended hospital medical staff. Health Aff (Millwood). 2007; 26(1):w44–w57. [PubMed: 17148490]

32. 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):982–990. [PubMed: 20439895]

33. Shortell SM, Casalino LP. Implementing qualifications criteria and technical assistance for accountable care organizations. JAMA. 2010; 303(17):1747–1748. [PubMed: 20442391]

34. Shortell SM, Casalino LP, Fisher ES. How the Center for Medicare and Medicaid Innovation should test accountable care organizations. Health Aff (Millwood). 2010; 29(7):1293–1298. [PubMed: 20606176]

35. Berenson, RA.; Burton, RA. Accountable care organizations in Medicare and the private sector: a status update. Washington (DC): Urban Institute; 2011 Nov 3. date.
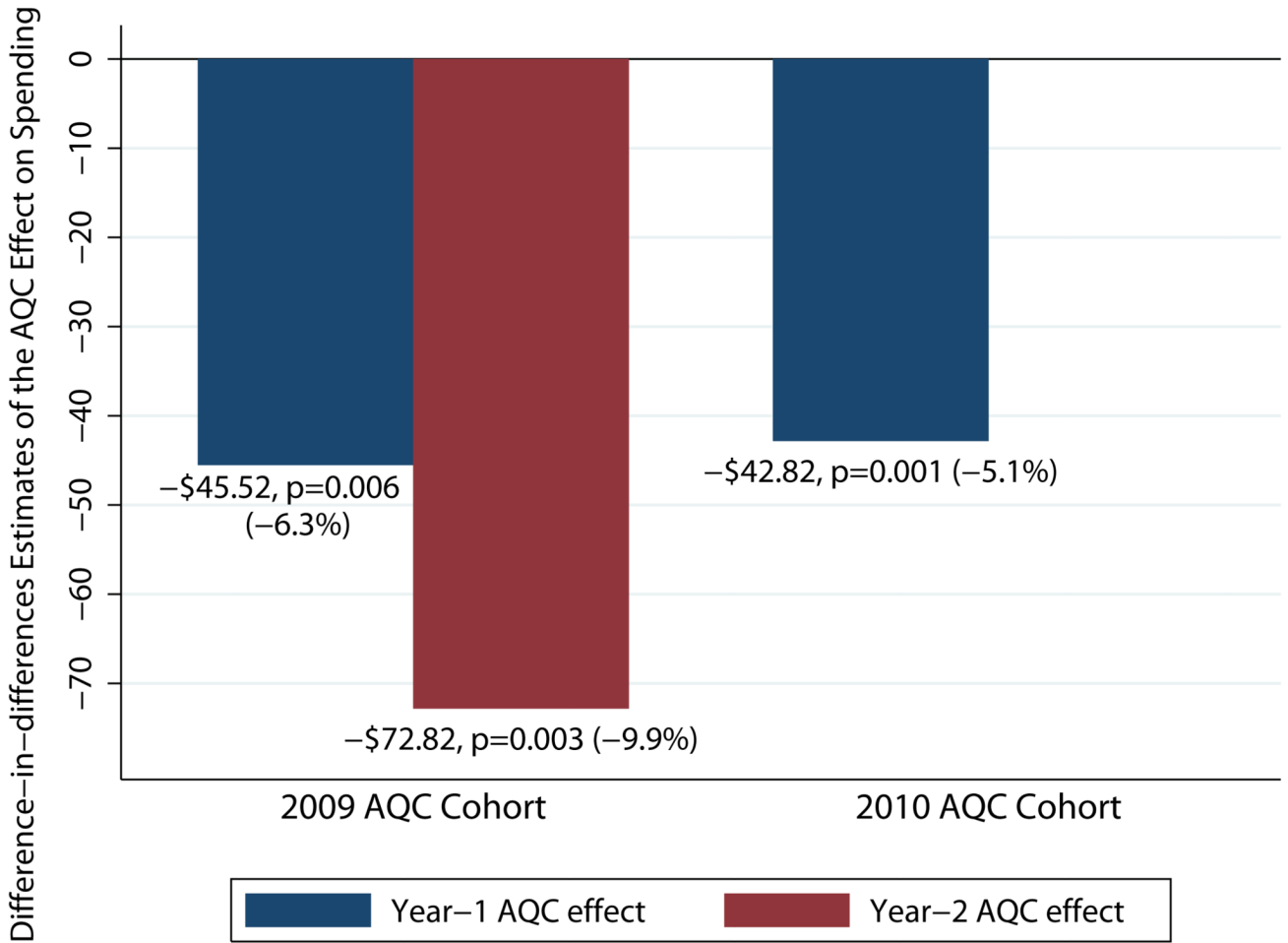
Exhibit 3.
Estimated Year 1 And Year 2 Effects Of The Alternative Quality Contract (AQC) On Spending In The 2009 And 2010 Cohorts' No-Prior-Risk Groups, Blue Cross Blue Shield Of Massachusetts
Source/Notes: SOURCE Authors' analysis of 2006–10 claims data from Blue Cross Blue Shield of Massachusetts. NOTES Exhibit 3 shows difference-in-differences estimates of the separate year 1 and year 2 effects of the AQC on health care spending per member per quarter. For descriptions of the 2009 and 2010 cohorts, see the text.

Exhibit 1

Characteristics Of The Study Population: Effects Of Alternative Quality Contract (AQC) In Blue Cross Blue Shield Of Massachusetts, 2006–10

| Characteristic | AQC groups (intervention) | | | | Non-AQC groups (control) (N=1,339,798) | |
| | 2009 cohort (N=428,892) | | 2010 cohort (N=183,655) | | | |
	Pre AQC (2006–08)	Post AQC (2009–10)	Pre AQC (2006–09)	Post AQC (2010)	Pre AQC (2006–08)	Post AQC (2009–10)
Age (years)	34.4 ± 18.6	35.5 ± 18.5	36.2 ± 18.2	37.9 ± 18.2	35.2 ± 18.7	35.3 ± 19.0
Percent female	52.6	52.1	51.7	52.0	51.8	51.1
Health risk score[a]						
Mean	1.08	1.16	1.18	1.25	1.11	1.15
Interquartile range	0.12–1.29	0.13–1.39	0.13–1.43	0.14–1.53	0.10–1.31	0.11–1.37
Cost sharing (%)						
Mean	13.9	16.1	14.4	16.7	13.8	15.8
Interquartile range	11.3–16.3	11.0–18.6	11.2–17.3	12.7–18.5	11.2–16.3	11.0–18.6

Source/Notes: SOURCE Authors' analysis of 2006–2010 Blue Cross Blue Shield of Massachusetts enrollment data. NOTES Plus–minus values are means ± SD (standard deviation). The total number of enrollees in the intervention and control groups exceeds 1,655,745 because there were enrollees who had one primary care physician in the intervention group and another in the control group for at least one year in each case. Cost sharing represents the average percentage of spending paid for out of pocket by the enrollee on common services, defined by using the most frequent Current Procedural Terminology (CPT) codes.

[a]The health risk score takes into account the health status of the enrollee and expected spending. See Note 21 in text.

Exhibit 2

Change In Average Health Care Spending Per Member Per Quarter In The 2009 Intervention Cohort And Control Groups, Blue Cross Blue Shield Of Michigan Alternative Quality Contract (AQC)

Category/site and type of care	2009 AQC cohort (intervention)		Non-AQC groups (control)		Between-group difference		Between-group difference by year			
	Pre AQC (2006–08)	Post AQC (2009–10)	Pre AQC (2006–08)	Post AQC (2009–10)	Average 2-year effect	p value	Year 1 (2009) effect	p value	Year 2 (2010) effect	p value
Total quarterly spending ($)	803.98	863.26	842.63	924.49	−22.58	0.04	−15.51	0.009	−26.72	0.04
By BETOS category ($)										
E&M	182.48	217.66	183.40	222.63	−4.06	0.28	−2.22	0.002	−5.32	0.22
Procedures	168.74	188.91	171.26	198.43	−7.00	0.02	−5.96	0.001	−7.62	0.05
Imaging	95.66	103.23	103.16	116.41	−5.67	0.001	−3.47	<0.001	−6.86	0.001
Test	68.35	78.72	76.15	90.76	−4.24	0.007	−3.72	<0.001	−4.28	0.03
DME	9.96	12.39	11.15	14.01	−0.44	0.19	−0.14	0.68	−0.72	0.14
Other	49.24	53.04	55.80	57.41	2.19	0.39	0.80	0.72	3.41	0.25
Unclassified	195.66	201.86	206.06	216.73	−4.48	0.22	−0.80	0.84	−6.97	0.16
By site and type of care ($)										
Inpatient Professional	35.55	38.64	35.39	39.22	−0.73	0.38	−0.72	0.38	−0.51	0.67
Facility	157.17	165.02	163.54	174.91	−3.53	0.32	0.23	0.95	−6.23	0.19
Outpatient Professional	319.28	367.61	302.66	352.55	−1.57	0.79	−0.28	0.8	−2.31	0.72
Facility	218.06	243.02	263.69	306.79	−18.14	0.004	−14.50	<0.001	−20.00	0.01
Ancillary	40.03	41.50	41.71	42.92	0.26	0.85	−0.24	0.86	0.68	0.73

SOURCE Authors' analysis of 2006–2010 Blue Cross Blue Shield of Massachusetts claims data. NOTES Sample sizes are presented in Exhibit 1. BETOS categories are Berenson–Eggers Type of Service (BETOS) classification, 2010 version. All spending figures are in 2010 US dollars.

Exhibit 4

Change In Performance On Ambulatory Care Quality Measures In The Intervention And Control Groups, Alternative Quality Contract (AQC), Blue Cross Blue Shield Of Massachusetts

Quality metric	Percent of eligible enrollees for whom performance threshold was met				Between-group difference (intervention-control)			Between-group difference by year			
	2009 AQC cohort (intervention)		Non-AQC groups (control)								
	Pre AQC	Post AQC	Pre AQC	Post AQC	Raw	Adj.[a]	p value	Year 1 (2009) effect	p value	Year 2 (2010) effect	p value
Chronic care management (aggregate)	79.1	83.3	79.7	80.0	3.9	3.7	<0.001	2.6	<0.001	4.7	<0.001
Cardiovascular LDL cholesterol screening	88.6	91.1	90.2	89.8	2.9	3.0	<0.001	1.8	0.04	4.5	<0.001
Diabetes											
HbA1c testing	89.3	92.4	89.3	90.3	2.1	2.1	<0.001	1.7	<0.001	2.5	<0.001
Eye exam	58.5	65.2	61.7	61.2	7.2	7.2	<0.001	5.5	<0.001	8.8	<0.001
LDL cholesterol screening	86.6	90.6	86.2	86.9	3.3	3.3	<0.001	2.8	<0.001	3.8	<0.001
Nephrology screening	85.1	88.3	83.6	83.7	3.1	2.9	<0.001	1.6	0.001	4.2	<0.001
Depression											
Short-term Rx	67.2	68.0	66.9	66.9	0.8	0.5	0.78	−1.1	0.59	1.6	0.44
Maintenance Rx	51.2	53.8	51.1	50.5	3.2	2.9	0.09	1.1	0.59	3.9	0.07
Adult preventive care (aggregate)	75.7	80.0	72.8	76.5	0.6	0.4	0.004	0.1	0.67	0.7	<0.001
Breast cancer screening	80.2	83.7	79.6	81.1	2.0	1.2	<0.001	0.6	0.006	1.9	<0.001
Cervical cancer screening	87.3	87.7	84.3	85.1	−0.4	−0.4	0.01	−0.5	0.002	−0.3	0.14
Colorectal cancer screening	64.2	71.7	59.7	67.1	0.1	0.0	0.92	0.0	0.97	0.3	0.26
Chlamydia screening for enrollees ages 21–24	58.6	65.8	53.9	61.2	−0.1	0.0	0.99	−0.8	0.41	0.7	0.51
No antibiotics for acute bronchitis	18.7	28.1	19.9	21.1	8.2	9.4	<0.001	5.5	<0.001	13.1	<0.001
Pediatric care (aggregate)	79.5	82.8	74.7	77.1	0.9	1.3	<0.001	0.7	0.001	1.9	<0.001
Appropriate testing for pharyngitis	93.9	96.1	81.8	90.5	−6.5	−6.1	<0.001	−3.9	<0.001	−7.5	<0.001
Chlamydia screening for enrollees ages 16–20	54.8	66.0	51.3	55.9	6.6	6.8	<0.001	5.4	<0.001	8.2	<0.001
No antibiotics for upper respiratory infection	94.9	95.5	92.1	93.7	−1.0	−1.0	0.04	−0.4	0.52	−1.8	0.006
Well care											
Babies age <15 months age	93.0	94.0	92.5	93.4	0.1	0.2	0.77	−0.1	0.91	0.6	<0.001

Quality metric	Percent of eligible enrollees for whom performance threshold was met												
	2009 AQC cohort (intervention)		Non-AQC groups (control)		Between-group difference (intervention-control)				Between-group difference by year				
	Pre AQC	Post AQC	Pre AQC	Post AQC	Raw	Adj.[a]	p value		Year 1 (2009) effect	p value	Year 2 (2010) effect	p value	
Children ages 3–6 years	92.3	94.8	90.0	91.3	1.2	1.1	<0.001		0.6	0.09	1.6	<0.001	
Adolescents ages 12–21 years	73.8	77.9	69.1	71.9	1.3	1.7	<0.001		0.09	<0.001	2.5	<0.001	

SOURCE Authors' analysis of 2007–10 quality data, Blue Cross Blue Shield of Massachusetts. NOTES The pre AQC period was 2007–08, and the post AQC period was 2009–10. For descriptions of intervention and control groups, see the text. Adjusted results are from a propensity-weighted difference-in-differences model controlling for all covariates. Pooled observations were used for the aggregate analyses of chronic care management, adult preventive care, and pediatric care, and the analyses were further adjusted for measure-level fixed effects. LDL is low-density lipoprotein.

Exhibit 5

Outcome Quality For Alternative Quality Contract (AQC) Groups And The Blue Cross Blue Shield Of Massachusetts Network Average, 2006–10

Condition/outcome measure	BCBS network average (%)					AQC weighted average (%)		
	2007	2008	2009	2010		2009 cohort (2009–10)	2010 cohort (2010)	All AQC 2010
Diabetes								
HbA1c control (<9 percent)	83.7	79.8	82.0	80.7		80.7	79.2	81.2
LDL cholesterol control (<100 mg/dL)	45.7	51.3	51.3	54.7		57.7	54.3	58.0
Blood pressure control (130/80)	30.9	36.7	38.0	35.8		44.3	38.9	46.0
Hypertension								
Blood pressure control (140/90)	68.4	70.3	69.5	67.5		68.4	71.1	73.0
Cardiovascular disease								
LDL cholesterol control (<100 mg/dL)	64.2	69.5	69.5	69.5		69.9	63.9	69.0

SOURCE Authors' analysis of 2006–10 claims data, Blue Cross Blue Shield of Massachusetts. NOTES Scores denote the percentage of eligible enrollees who met the quality criteria as defined. Scores are weighted by eligible members for each measure and are unadjusted averages.