SPECIAL ARTICLE

Changes in Health Care Spending and Quality 4 Years into Global Payment

Zirui Song, M.D., Ph.D., Sherri Rose, Ph.D., Dana G. Safran, Sc.D., Bruce E. Landon, M.D., M.B.A., Matthew P. Day, F.S.A., M.A.A.A., and Michael E. Chernew, Ph.D.

ABSTRACT

BACKGROUND

Spending and quality under global budgets remain unknown beyond 2 years. We evaluated spending and quality measures during the first 4 years of the Blue Cross Blue Shield of Massachusetts Alternative Quality Contract (AQC).

METHODS

We compared spending and quality among enrollees whose physician organizations entered the AQC from 2009 through 2012 with those among persons in control states. We studied spending changes according to year, category of service, site of care, experience managing risk contracts, and price versus utilization. We evaluated process and outcome quality.

RESULTS

In the 2009 AQC cohort, medical spending on claims grew an average of $62.21 per enrollee per quarter less than it did in the control cohort over the 4-year period (P<0.001). This amount is equivalent to a 6.8% savings when calculated as a proportion of the average post-AQC spending level in the 2009 AQC cohort. 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. Claims savings were concentrated in the outpatient-facility setting and in procedures, imaging, and tests, explained by both reduced prices and reduced utilization. Claims savings were exceeded by incentive payments to providers during the period from 2009 through 2011 but exceeded incentive payments in 2012, generating net savings. Improvements in quality among AQC cohorts generally exceeded those seen elsewhere in New England and nationally.

CONCLUSIONS

As compared with similar populations in other states, Massachusetts AQC enrollees had lower spending growth and generally greater quality improvements after 4 years. Although other factors in Massachusetts may have contributed, particularly in the later part of the study period, global budget contracts with quality incentives may encourage changes in practice patterns that help reduce spending and improve quality. (Funded by the Commonwealth Fund and others.)

From the Department of Medicine, Massachusetts General Hospital (Z.S.), Department of Health Care Policy, Harvard Medical School (Z.S., S.R., B.E.L., M.E.C.), Blue Cross Blue Shield of Massachusetts (D.G.S., M.P.D.), the Department of Medicine, Tufts University School of Medicine (D.G.S.), and the Department of Medicine, Beth Israel Deaconess Medical Center (B.E.L.) — all in Boston; and the National Bureau of Economic Research, Cambridge, MA (Z.S., M.E.C.). Address reprint requests to Dr. Song at the Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave., Boston, MA 02115, or at zirui_song@post.harvard.edu.

N Engl J Med 2014;371:1704-14.
DOI: 10.1056/NEJMsa1404026

To slow the growth of health care spending, insurers are moving toward global budgets. Increasingly, physicians are forming or joining accountable care organizations (ACOs) to take on such contracts.[1-3] As of 2014, Medicare has entered into ACO agreements with 360 physician organizations caring for 5.3 million beneficiaries.[4] Combined with a similar growth in the private sector, an estimated 18 million persons in the United States have insurance coverage in which their physicians are in ACO arrangements.[5]

Massachusetts was an early adopter of payment reform.[6] One of the first developments occurred in 2009, when Blue Cross Blue Shield of Massachusetts (BCBS) implemented the Alternative Quality Contract (AQC), which pays providers a risk-adjusted global budget.[7] By 2012, approximately 85% of the physicians in the BCBS network had entered the AQC. Tufts Health Plan, another large insurer in Massachusetts, undertook similar efforts and brought 72% of its commercial managed care enrollees under global budgets by 2012.[8] The Medicare Pioneer ACO program, launched in 2012, includes five organizations from Massachusetts, and other Massachusetts providers have since joined the Medicare Shared Savings Program.

The AQC is a two-sided contract with shared savings if spending is below budget and shared risk if spending exceeds the budget (a so-called risk contract). Organizations receive quality bonuses that are based on 64 measures, including data on processes, outcomes, and patients' experiences in the ambulatory care and hospital settings (Table S1 in the Supplementary Appendix, available with the full text of this article at NEJM.org). Enrollees are prospectively attributed to provider organizations by means of the affiliation of their primary care physician (PCP), whom they designate each year. The organization is then responsible for managing a population budget, similar to the idea of the patient-centered medical home within a "medical neighborhood."[9,10] Organizations also receive periodic reports from BCBS regarding cost and quality performance, including comparisons of their care patterns with those of other organizations, to help providers identify areas of potential overuse and improvement.

In 2009 and 2010, annual AQC budget increases were predetermined percentages, and the amount of shared savings or risk was independent of quality. Because environmental factors (e.g., disease outbreaks) could hinder the ability of organizations to meet absolute budget targets, a complex year-end reconciliation process was needed. To address this situation, contracts in 2011 and thereafter had annual budget increases that were tied to regional spending benchmarks. Shared savings and deficits were also tied to quality performance, with higher quality conveying a larger share of savings and a smaller share of deficits to providers. Quality bonuses were defined on a per-member, per-month basis rather than as a percentage of the budget, helping to equalize bonuses across organizations with similar quality.[11]

Early evaluations of the AQC showed improvements in quality and reductions in claims spending, driven by a shifting of care to less expensive providers and by some reduced utilization.[12-15] These initial savings were exceeded by incentive payments to providers.[12,13] Recently, Medicare ACOs also reported early savings and quality improvements.[16,17] However, there is little evidence on spending and quality beyond 1 to 2 years and little evidence regarding whether savings outgrow incentive payouts over time.

Methods

Population

The intervention group consisted of four cohorts of AQC organizations that were defined by their first contract year: 2009, 2010, 2011, or 2012 (Table S2 in the Supplementary Appendix). Within the 2009 AQC cohort, we prespecified a prior-risk subgroup of organizations that had experience managing risk contracts with BCBS and a no-prior-risk subgroup of organizations that did not have such experience. Prior-risk organizations tended to be larger delivery systems, whereas no-prior-risk organizations were smaller groups. The 2010 cohort included only no-prior-risk organizations; the 2011 and 2012 cohorts included mostly no-prior-risk organizations.

The 2011 and 2012 cohorts had made up the majority of the control group in previous AQC evaluations.[12,13] After entering the AQC, they could no longer serve as controls. The same situation applied to the remaining 15% of providers in the BCBS network that were not in a risk contract by 2012; they were small, rural practices that received a different set of payment updates,

rendering them nonrepresentative. Therefore, a different control group was needed.

Our control group comprised commercially insured persons in employer-sponsored plans across all eight other Northeastern states (Connecticut, Maine, New Jersey, New Hampshire, New York, Pennsylvania, Rhode Island, and Vermont) in the Commercial Claims and Encounters database of Truven Health Analytics.[18] All persons were younger than 65 years of age and had been enrolled continuously in a health maintenance organization or point-of-service plan for at least 1 year. Similar to the AQC, these plans all have physician networks and require enrollees to designate a PCP. All the plans were sponsored by employers that reported claims continuously during the study period.

Control states had demographic characteristics and risk scores similar to those in Massachusetts. They also had large and small provider groups and multiple commercial payers. Unlike Massachusetts, these states did not have broad-scale shifts to global payment by large commercial payers during the study period. Medical-home pilot programs in Rhode Island and Pennsylvania were smaller in scale than the Massachusetts AQC and did not involve risk contracts.[19,20] The single-payer system in Vermont, passed in 2011, will not be operational until 2017.[21] In sensitivity analyses, we tested other Massachusetts and nationwide controls.

DATA AND VARIABLES

For spending analyses, the dependent variable was claims payments, which reflected negotiated prices and patient cost sharing. In addition to total medical claims, spending was analyzed according to site of care (inpatient or outpatient), type of claim (facility or professional), and category of care according to the Berenson–Eggers Type of Service classification. Pharmaceutical claims were added in sensitivity analyses; they were excluded from the main analyses because not all the enrollees had drug benefits administered by their primary insurer.

To assess whether incentive payments exceeded claims savings, we computed aggregate incentives by combining shared savings under the budget (specific to each contract), quality bonuses, and infrastructure payments that supported providers' investment in care redesign. Incentive payments are calculated by BCBS with internal auditing and by AQC groups for accuracy before both sides sign a contract amendment accepting the final results. Because of the propriety nature of these contracts and legal restrictions with respect to making incentive payments public, the payments were aggregated according to year and provided as a percentage of claims spending in ranges, which enabled comparisons with estimated claims savings.

A total of 18 measures of ambulatory care process were available for BCBS enrollees from 2007 through 2012 in three categories: chronic disease management, adult preventive care, and pediatric care. Five outcome measures were available: a glycated hemoglobin level of 9% or less, a low-density lipoprotein (LDL) cholesterol level of less than 100 mg per deciliter (2.59 mmol per liter), and a blood pressure of less than 140/80 mm Hg for patients with diabetes; an LDL cholesterol level of less than 100 mg per deciliter for patients with coronary artery disease; and a blood pressure of less than 140/90 mm Hg for patients with hypertension. Because the data provided by Truven Health Analytics did not include quality measures, we compared the AQC data with Healthcare Effectiveness Data and Information Set (HEDIS) scores for the nation and for New England.[22]

To calculate risk scores, we used diagnostic-cost-group (DxCG) software (Verisk Health), which uses data regarding demographic characteristics, enrollment, claims, and diagnoses according to the *International Classification of Diseases, Ninth Revision*.[23] The DxCG method is used by most private payers in the United States as a risk-adjustment tool and is related to the risk scores used by Medicare for its Hierarchical Condition Categories.[24]

STATISTICAL ANALYSIS

We used a difference-in-differences approach to compare changes in spending and quality between the AQC group and the control group.[25] For the 2009 cohort, the preintervention period was from 2006 through 2008, and the postintervention period was from 2009 through 2012. We used these definitions to compare the 2010, 2011, and 2012 cohorts with the control group.

We used enrollee-level linear models with adjustment for age, sex, risk score, indicator for AQC, year indicators, and interactions between the AQC indicator and year. These interactions estimated changes attributable to the AQC but

should be interpreted with the recognition that other factors in Massachusetts may have influenced spending or quality. We scaled dollar results for each year into percentages by dividing them by the claims costs for the current year. We included state and plan fixed effects to account for such time-invariant characteristics among individual enrollees. We used a linear model because we were most interested in estimating average effects. (Given 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.[26-29])

We tested for differences in preintervention trends between the AQC group and the control group and estimated the extent to which changes in spending were explained by changes in prices or utilization (volume). By substituting median prices across all the providers for each service, changes in spending reflected changes in utilization.

We performed a number of sensitivity analyses. We also tested for differential changes in risk scores between the AQC group and the control group, which may reflect changes in coding intensity. Standard errors were clustered according to plan and are reported with two-tailed P values.[30,31]

Analyses of quality used an unadjusted difference-in-differences approach, given that HEDIS quality scores were aggregated. Performance was considered to be the percentage of enrollees who were eligible for a measure whose care met qual-

Table 1. Characteristics of Alternative Quality Contract (AQC) Cohorts and Control Group.*

Characteristic	2009 Cohort	2010 Cohort	2011 Cohort	2012 Cohort	Control Group
Enrollees (no.)†	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 (no.)	7	4	1	5	NA
Primary care physicians	1151	469	420	2115	NA
Specialists	2197	1010	1319	7260	NA
Affiliated hospitals	15	13	2	10	NA

* Plus–minus values are means ±SD.

† Data are shown for persons who were enrolled for at least 1 year in the study period. Persons in the ACQ cohorts selected primary care physicians working in organizations that joined the AQC. Persons in the control group were enrolled in employer-sponsored plans in eight other northeastern states (Connecticut, Maine, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont). No data on provider organizations were available for controls (NA). Age, sex, health risk score, and cost sharing (i.e., the portion of spending paid by the enrollee, calculated as a percentage annually) were pooled across all the enrollees during the entire study period. The characteristics of the enrollees in the four AQC cohorts did not differ substantially from those in the control group. Given the large sample sizes, however, there were significant differences in the characteristics of the enrollees in the AQC cohorts, as compared with those in the control group. P values for the differences are shown in Table S3 in the Supplementary Appendix. The characteristics of the enrollees were adjusted for in the base statistical model and were excluded in the sensitivity analyses (Table S6 in the Supplementary Appendix) to test the robustness of the main results.

‡ The diagnostic-cost-group (DxCG) risk score is a measure of enrollee health status, calculated with the use of coefficients from a statistical model run on a national claims database that relates spending to diagnoses according to the *International Classification of Diseases, Ninth Revision*, and demographic characteristics. The DxCG method is similar to the Medicare Hierarchical Condition Category risk scores and is commonly used for risk adjustment. The average risk score across all persons was 1.03 (interquartile range, 0.18 to 1.07); higher values indicate higher expected spending.

ity criteria in a given year. All the analyses were performed with the use of Stata software, version 13 (StataCorp).

RESULTS

POPULATION

The characteristics of the four AQC cohorts were similar to those in the control group (Table 1). The 2009 AQC cohort comprised seven organizations with approximately 1100 PCPs and 2000 specialists. Comparisons between each cohort and the control group, both before and after the intervention, are provided in Table S3 in the Supplementary Appendix.

SPENDING ON CLAIMS

Unadjusted claims spending in the 2009 AQC cohort grew slower after entering the contract, as compared with the control group (Fig. 1). In the adjusted analysis, medical spending grew an average of $62.21 per enrollee per quarter less in the AQC cohort than in the control group during the 4-year postintervention period (P<0.001), which was an average savings of 6.8% as compared with the average postintervention claims spending level in the AQC cohort (Table 2). Preintervention trends did not differ significantly between the AQC group and the control group (difference, −$4.57; P=0.86), suggesting that differences in postintervention spending were not driven by inherently different trajectories of spending. No differential changes in the DxCG risk scores were found between the AQC group and the control group (difference, −0.0015; P=0.57), suggesting that coding behavior did not meaningfully affect the results.

Savings were most pronounced in the outpatient setting over the 4-year postintervention period: 4.0% for professional spending (P=0.004) and 19.3% for facility spending (P<0.001) (Table S4 in the Supplementary Appendix). Among the categories of services, savings were largest with respect to procedures (8.7%, P<0.001), imaging (10.9%, P<0.001), and tests (9.7%, P<0.001). The prior-risk and no-prior-risk subgroups had similar changes in spending, as compared with the control group (−6.3% and −7.7%, respectively; P<0.001) (Table S5 in the Supplementary Appendix). Sensitivity analyses showed that the main estimates were robust to alterations in the model, variables, or sample (Table S6 in the Supplementary Appendix).

Models that used standardized prices indicated that approximately 40% of the claims savings (P<0.001) were explained by decreases in volume, with the remainder due to lower prices. Over the 4-year period, the proportion of savings associated with volume reductions was approximately 60% for procedures, 25% for imaging, and 60% for tests (P<0.001 for all comparisons).

Unadjusted spending in the 2010, 2011, and 2012 cohorts is shown in Figures S1, S2, and S3, respectively, in the Supplementary Appendix. Over their respective 3-year, 2-year, and 1-year contract periods, as compared with the control group, medical spending per enrollee per quarter grew an average of $81.92 less (−8.8%, P<0.001), $97.10 less (−9.1%, P<0.001), and $59.39 less (−5.8%, P=0.04) (Table 2). Savings were similarly concentrated in the outpatient facility setting and among procedures, imaging, and tests (Table S4 in the Supplementary Appendix). The proportion of savings attributable to volume in the 2010, 2011, and 2012 cohorts was approximately 25% (P=0.001), 50% (P<0.001), and 25% (P=0.08), respectively.

TOTAL PAYMENTS

As noted in earlier work, incentive payments (shared savings, quality bonuses, and infrastructure support) exceeded savings on claims in the period from 2009 through 2010, reflecting investment in the early years of payment reform.[12,13] This pattern continued into 2011, with a smaller gap, but reversed in 2012, when claims savings exceeded incentive payments to generate a net savings (Table 2). By 2012, the 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 that was based on controls.

QUALITY

The average performance of the 2009 AQC cohort on measures of chronic disease management increased from 79.6% in the period from 2007 through 2008 to 84.5% in the period from 2009 through 2012, as compared with 79.8% and 80.8% in the respective periods for the HEDIS national average, rendering an unadjusted increase of 3.9 percentage points over the control group during the 4-year period (Table 3). Analogously, unadjusted performance in adult preventive care and pediatric care increased by 2.7 percentage points and 2.4 percentage points, respectively. The 2010 and 2011 cohorts also saw increases in

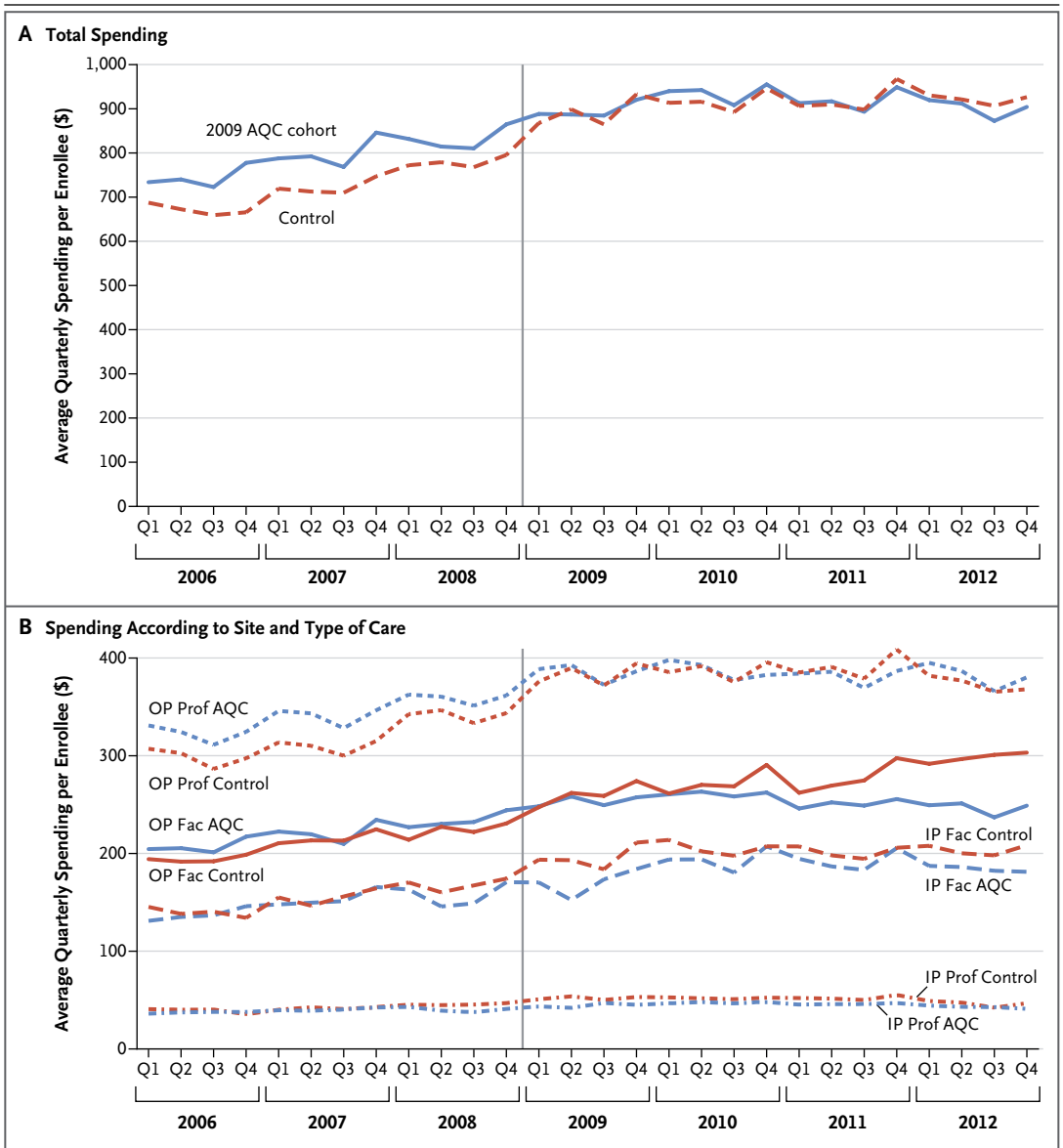

Figure 1. Unadjusted Spending in the 2009 Alternative Quality Contract (AQC) Cohort versus the Control Group, 2006–2012.

Panel A shows the total unadjusted spending. Panel B shows the results according to site of care (inpatient [IP] or outpatient [OP]) and type of claim (facility [Fac] or professional [Prof]). The control group comprised commercially insured enrollees in employer-sponsored plans across eight Northeastern states: Connecticut, Maine, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. The vertical line at the start of 2009 indicates the start of the AQC period.

unadjusted performance as compared with HEDIS, whereas the 2012 cohort, which had a higher level of performance to begin with, did not improve as compared with HEDIS during the first year. Comparisons with HEDIS New England scores were qualitatively similar (Table S7 in the Supplementary Appendix).

Aggregate outcome quality in the 2009 AQC cohort increased each year in the postintervention period, as compared with both the HEDIS national average and the HEDIS New England average (Fig. 2). On average, achievement of control of the glycated hemoglobin level, the LDL cholesterol level, and blood pressure grew by 2.1 percentage points per year after entry into the AQC, whereas the HEDIS data remained

Table 2. Changes in Medical Spending and Total Payments Associated with the AQC, According to Cohort and Year.*

Variable	2009		2010		2011		2012		Cohort Average		
	Change	P Value	Change	P Value	Change	P Value	Change	P Value	Change	P Value	Change as Share of Postintervention Cohort Spending
											%
Change in medical spending ($)†											
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 (% of current-yr FFS claims)‡	2.4		3.1		8.4		10.0			NA	
Incentive payments to providers (% of current-yr FFS claims)§	6–9		9–12		10–13		6–9			NA	
Implication	BCBS payments to providers, including shared savings and bonuses for quality and infrastructure, exceeded savings on claims		Payments to providers exceeded savings on claims		Payments exceeded savings on claims, but by a smaller amount than in earlier years		Savings on claims exceeded payments, rendering net savings			NA	
Scope of adoption in Massachusetts (% of BCBS providers in AQC)	Approximately 20		Approximately 25		Approximately 33		Approximately 75			NA	

* BCBS denotes Blue Cross Blue Shield of Massachusetts, and NA not applicable.

† All values are per enrollee per quarter. Changes in spending on claims are from a difference-in-differences regression analysis with adjustment for covariates. Negative values represent savings. Cohort averages were scaled into a percentage by dividing the average savings of the cohort in the AQC by its average spending levels after participation in the AQC. Dollars are inflation-adjusted to 2012 dollars.

‡ Average savings on claims were weighted across cohorts in each year, scaled into percentages by dividing into the weighted average of current-year fee-for-service (FFS) claims spending measured 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 owing to the confidentiality of contracts between BCBS and provider organizations.

The New England Journal of Medicine

Table 3. Ambulatory Process Quality in the AQC Cohorts versus the Healthcare Effectiveness Data and Information Set (HEDIS) National Average.*

AQC Cohort and Aggregate Process	AQC Group			HEDIS National Average			Unadjusted Difference-in-Differences Analysis†
	Preintervention Period‡	Postintervention Period‡	Change	Preintervention Period‡	Postintervention Period‡	Change	
	percent		*percentage points*	*percent*			*percentage points*
2009 cohort							
Chronic disease 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 cohort							
Chronic disease management	80.3	82.6	2.3	80.0	81.0	1.0	1.3
Adult preventive care	74.8	79.9	5.2	58.0	59.7	1.7	3.4
Pediatric care	75.9	80.8	4.8	69.1	71.3	2.2	2.6
2011 cohort							
Chronic disease management	79.4	81.4	2.0	80.2	80.9	0.7	1.3
Adult preventive care	72.8	79.1	6.3	58.4	59.9	1.5	4.8
Pediatric care	75.3	79.9	4.6	69.5	71.6	2.2	2.5
2012 cohort							
Chronic disease management	82.1	80.5	−1.5	80.4	81.0	0.6	−2.2
Adult preventive care	77.3	78.0	0.8	58.7	59.9	1.2	−0.4
Pediatric care	80.1	81.9	1.7	69.9	71.7	1.8	−0.1

* Values indicate the percentage of eligible enrollees whose care achieved a threshold performance for the measure. These three aggregate ambulatory process measures are weighted averages of individual measures in each category. The measures for chronic disease management were LDL cholesterol screening for patients with cardiovascular disease; testing of the glycated hemoglobin level, eye examination, LDL cholesterol screening, and nephrology screening for patients with diabetes; and short-term and maintenance prescriptions for patients with depression. The adult preventive care measures were screening for breast cancer, cervical cancer, and colorectal cancer; chlamydia screening for enrollees 21 to 24 years of age; and no prescribing of antibiotic agents for acute bronchitis. The pediatric care measures were appropriate testing for pharyngitis; chlamydia screening for enrollees 16 to 20 years of age; no prescribing of antibiotics for upper respiratory infection; and well-child visits in the first 15 months of life, well-child visits in the 3rd to 6th years of life, and adolescent well care visits in the 12th to 21st years of life. Well care visits for children between but not including 15 months of life and the 3rd year of life, as well as between but not including the 6th year of life and the 12th year of life, were not part of the HEDIS set of quality measures. All analyses were unadjusted; the results were calculated on the basis of raw weighted averages in the groups before and after their respective intervention dates.

† The unadjusted analysis used a difference-in-differences approach, so the average varied according to the cohort year. The postintervention data for the 2009 cohort were averaged over a 4-year period, the postintervention data for the 2010 cohort over a 3-year period, the postintervention data for the 2011 cohort over a 2-year period, and the postintervention data for the 2012 cohort over a 1-year period.

‡ For each cohort, the preintervention period was from 2006 to the year preceding the cohort year, and the postintervention period was from the cohort year through 2012.

The New England Journal of Medicine

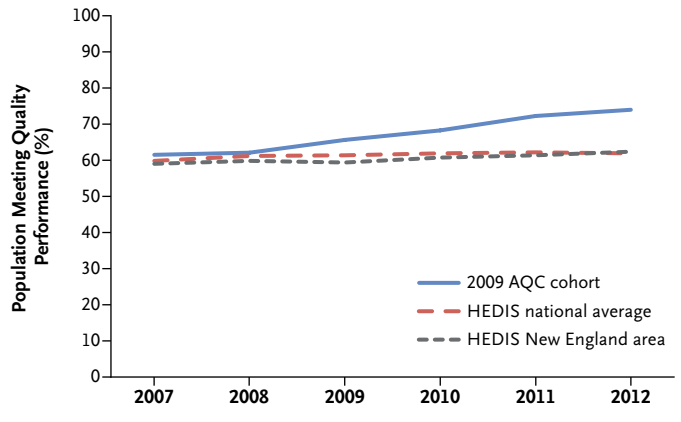

Figure 2. Outcome Quality in the 2009 AQC Cohort versus the Healthcare Effectiveness Data and Information Set (HEDIS), 2007–2012.

Outcome quality consisted of the following five measures: control of the glycated hemoglobin level (≤9%), control of the low-density lipoprotein (LDL) cholesterol level (<100 mg per deciliter [2.6 mmol per liter]), and blood-pressure control (<140/80 mm Hg) in patients with diabetes; the same level of control of LDL cholesterol in patients with coronary artery disease; and a blood-pressure control level of 140/90 mm Hg in patients with hypertension.

largely unchanged. General improvements in outcome measures were also seen in the 2010, 2011, and 2012 cohorts, as compared with the HEDIS data (Table S8 in the Supplementary Appendix).

DISCUSSION

As compared with similar populations in other states, Massachusetts enrollees in the AQC had slower spending growth in the period from 2009 through 2012. Savings were mostly concentrated in the outpatient facility setting and explained by both reduced prices and reduced utilization after 4 years. Improvements in process and outcome quality in the measured domains were generally larger than those seen outside Massachusetts.

Other factors in Massachusetts could have influenced spending and quality. The 2012 Massachusetts payment reform legislation created the state Health Policy Commission and encouraged ACO adoption. Also, global budget contracts with other payers may have spillover effects on the BCBS population.[32] However, reforms in Massachusetts mostly postdate our study period. Moreover, the Medicare Pioneer ACO program was launched in 2012; Tufts Health Plan and Harvard Pilgrim Health Care began large-scale global payment contracts around 2012. Therefore, although our findings for 2012 may be susceptible to spillover effects, and anticipatory effects from other contracts may also play a role, our prior analyses that used internal controls, the consistency of the sensitivity analyses, and qualitative findings from interviews with providers suggest that the AQC played a meaningful role.[12-15]

Our study has several limitations. First, selection bias is a concern because participation in the AQC is voluntary. The lack of differences in preintervention trends between the AQC group and the control group attenuates this concern. The fact that most provider organizations in Massachusetts entered the AQC by year 4 further alleviates this concern. Nevertheless, a potential selection bias cannot be eliminated, because there remain unobserved factors that may have influenced participation as well as spending.

Second, internal validity is also threatened if control states underwent payment reform. However, we know of no broad-scale reforms among large private insurers in these states during the study period. Small-scale medical-home pilot programs in some states have thus far not been shown to affect spending substantially.[19,20] Nevertheless, payment reform was an active issue in many states. We could not identify specific providers or insurers in the data from Truven Health Analytics, which prevented us from rigorously testing these concerns. However, to the extent that any payment reforms occurred in a subset of control states, their effects would be minimized by pooling across states. Moreover, to the extent that payment reforms might have slowed spending in control states, our estimated savings would be conservative.

The key question is whether our control group serves as a good counterfactual scenario. We believe that the lack of differences between the AQC group and the control group in preintervention trends and the pooling of control states boost the fidelity of the control group. This control group, which differed from that of prior AQC evaluations involving BCBS enrollees whose providers were not participating in the AQC, generated the same savings over the first 2 years in the 2009 cohort (calculated as 2.8% with the use of both our control group and the prior control group) and similar savings in the first year of the 2010 cohort (calculated as 2.9% and 4.9%, respectively).[12,13] Moreover, this control group appears to be more appropriate than alternative Massachusetts control groups that

have data limitations, such as concurrent inclusion of intervention participants in the control population, differences in payment updates, susceptibility to spillover effects, and divergent trends in preintervention spending, although alternative controls nevertheless produced findings in the same general direction (Table S6 in the Supplementary Appendix). Results of analyses that used nationwide controls were similar to our baseline estimates (Table S6 in the Supplementary Appendix).

Third, our results may not be generalizable to ACOs in Medicare. Most Medicare ACO contracts are one-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, our results may not be generalizable to other states, which face different constraints and challenges.[33-37]

Fourth, our quality analyses were based on aggregate data, rather than derived from a statistical model. Earlier work with the use of models analogous to our spending analysis showed significant improvements in the three dimensions of process quality, consistent with our descriptive results.[15] Our measures also do not capture all dimensions of quality. Process measures are primary care–centered, and the five outcome measures leave numerous important outcomes unmeasured.

The AQC experience may be useful to policymakers, insurers, and providers embarking on payment reform.[38] Although it is still early, these results suggest that a two-sided global budget model may serve as a foundation for slowing spending and improving quality. Shared savings coupled with quality bonuses can exceed savings on claims. Over time, however, savings on claims may outgrow incentive payments, as was the case with the AQC. Incentive payments themselves may serve meaningful purposes, such as the adoption of quality measures that protect against stinting or the sharing of potential losses that ease providers into risk contracts. Changes in spending owing to lower prices or utilization suggest that new payment models may help modify underlying care patterns, a likely prerequisite for sustainable reform. Going forward, the relationship between payers and providers will be crucial for the success of reforms in payment and delivery systems.[39,40]

The views expressed in this article are those of the authors and do not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health.

Supported by a grant from the Commonwealth Fund (to Dr. Chernew), a predoctoral M.D./Ph.D. National Research Service Award from the National Institute on Aging (F30-AG039175, to Dr. Song), a grant from the National Bureau of Economic Research Fellowship in Aging and Health Economics (T32-AG000186, to Dr. Song), and a grant from the Charles H. Hood Foundation (to Dr. Chernew).

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Angela Li, Katherine Barrett, Rosemary Brown, Kay Miller, and Michael McKellar for assistance with data; and Yanmei Liu for programming assistance.

REFERENCES

1. Epstein AM, Jha AK, Orav EJ, et al. Analysis of early accountable care organizations defines patient, structural, cost, and quality-of-care characteristics. Health Aff (Millwood) 2014;33:95-102.
2. Lewis VA, Colla CH, Carluzzo KL, Kler SE, Fisher ES. Accountable care organizations in the United States: market and demographic factors associated with formation. Health Serv Res 2013;48:1840-58.
3. Shortell SM, McClellan SR, Ramsay PP, Casalino LP, Ryan AM, Copeland KR. Physician practice participation in accountable care organizations: the emergence of the unicorn. Health Serv Res 2014 March 15 (Epub ahead of print).
4. More partnerships between doctors and hospitals strengthen coordinated care for Medicare beneficiaries. Press release of the Centers for Medicare and Medicaid Services, December 23, 2013 (http://www.cms.gov/Newsroom/MediaReleaseDatabase/Press-Releases/2013-Press-Releases-Items/2013-12-23.html).
5. Muhlestein D. Accountable care growth in 2014: a look ahead. Health Affairs Blog. January 29, 2014 (http://healthaffairs.org/blog/2014/01/29/accountable-care-growth-in-2014-a-look-ahead/).
6. Song Z, Landon BE. Controlling health care spending — the Massachusetts experiment. N Engl J Med 2012;366:1560-1.
7. Chernew ME, Mechanic RE, Landon BE, Safran DG. Private-payer innovation in Massachusetts: the 'alternative quality contract.' Health Aff (Millwood) 2011;30:51-61.
8. 2011 Leading through change, guided by mission: 2011 annual report. Watertown, MA: Tufts Health Plan. 2011.
9. Merrell K, Berenson RA. Structuring payment for medical homes. Health Aff (Millwood) 2010;29:852-8.
10. Greenberg JO, Barnett ML, Spinks MA, Dudley JC, Frolkis JP. The "medical neighborhood": integrating primary and specialty care for ambulatory patients. JAMA Intern Med 2014;174:454-7.
11. Blue Cross Blue Shield of Massachusetts. Massachusetts payment reform model: results and lessons. 2012 (https://www.bluecrossma.com/visitor/pdf/aqc-results-white-paper.pdf).
12. Song Z, Safran DG, Landon BE, et al. Health care spending and quality in year 1 of the alternative quality contract. N Engl J Med 2011;365:909-18.
13. Song Z, Safran DG, Landon BE, et al. The 'Alternative Quality Contract,' based on a global budget, lowered medical spending and improved quality. Health Aff (Millwood) 2012;31:1885-94.
14. Song Z, Fendrick AM, Safran DG, Landon B, Chernew ME. Global budgets and technology-intensive medical services. Healthcare (Amst) 2013;1:15-21.
15. 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;30:1734-42.
16. Pioneer Accountable Care Organizations succeed in improving care, lowering

costs. Press release of the Centers for Medicare and Medicaid Services, July 16, 2013 (http://www.cms.gov/Newsroom/MediaReleaseDatabase/Press-Releases/2013-Press-Releases-Items/2013-07-16.html).
17. Medicare ACOs continue to succeed in improving care, lowering cost growth. Press release of the Centers for Medicare and Medicaid Services, September 16, 2014 (http://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2014-Fact-sheets-items/2014-09-16.html).
18. Adamson DM, Chang S, Hansen LG. Health research data for the real world: the MarketScan Databases. Ann Arbor, MI: Thomson Reuters, January 2010 (http://depts.washington.edu/chaseall/pdfs/white%20paper_Thomson%20Reuters%20WP%20MarketScan%20Databases%200708.pdf).
19. Rosenthal MB, Friedberg MW, Singer SJ, Eastman D, Li Z, Schneider EC. Effect of a multipayer patient-centered medical home on health care utilization and quality: the Rhode Island Chronic Care Sustainability Initiative pilot program. JAMA Intern Med 2013;173:1907-13.
20. Friedberg MW, Schneider EC, Rosenthal MB, Volpp KG, Werner RM. Association between participation in a multipayer medical home intervention and changes in quality, utilization, and costs of care. JAMA 2014;311:815-25.
21. Wallack AR. Single payer ahead — cost control and the evolving Vermont model. N Engl J Med 2011;365:584-5.
22. National Committee for Quality Assurance. Quality Compass home page (http://www.ncqa.org/HEDISQualityMeasurement/QualityMeasurementProducts/QualityCompass.aspx).
23. DxCG risk solutions: DxCG science guide. Waltham, MA: Verisk Health, 2012.
24. Pope GC, Kautter J, Ellis RP, et al. Risk adjustment of Medicare capitation payments using the CMS-HCC model. Health Care Financ Rev 2004;25:119-41.
25. Wooldridge J. Econometric analysis of cross section and panel data. Cambridge, MA: MIT Press, 2001.
26. Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. J Health Econ 2005;24:465-88.
27. Buntin MB, Zaslavsky AM. Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures. J Health Econ 2004;23:525-42.
28. Ai C, Norton EC. Interaction terms in logit and probit models. Econ Lett 2003;80:123-9.
29. Ellis RP, McGuire TG. Predictability and predictiveness in health care spending. J Health Econ 2007;26:25-48.
30. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 1980;48:817-30.
31. 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.
32. McWilliams JM, Landon BE, Chernew ME. Changes in health care spending and quality for Medicare beneficiaries associated with a commercial ACO contract. JAMA 2013;310:829-36.
33. Rajkumar R, Patel A, Murphy K, et al. Maryland's all-payer approach to delivery-system reform. N Engl J Med 2014;370:493-5.
34. Ayanian JZ. Michigan's approach to Medicaid expansion and reform. N Engl J Med 2013;369:1773-5.
35. Wroten D. Arkansas payment reform . . . what happens now? J Ark Med Soc 2012;109:28.
36. Stecker EC. The Oregon ACO experiment — bold design, challenging execution. N Engl J Med 2013;368:982-5.
37. 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;10:1-9.
38. Fisher ES, McClellan MB, Safran DG. Building the path to accountable care. N Engl J Med 2011;365:2445-7.
39. Fisher ES, Shortell SM. Accountable care organizations: accountable for what, to whom, and how. JAMA 2010;304:1715-6.
40. Song Z, Lee TH. The era of delivery system reform begins. JAMA 2013;309:35-6.