SPECIAL ARTICLE

Health Care Spending, Utilization, and Quality 8 Years into Global Payment

Zirui Song, M.D., Ph.D., Yunan Ji, B.A., Dana G. Safran, Sc.D., and Michael E. Chernew, Ph.D.

ABSTRACT

BACKGROUND

Population-based global payment gives health care providers a spending target for the care of a defined group of patients. We examined changes in spending, utilization, and quality through 8 years of the Alternative Quality Contract (AQC) of Blue Cross Blue Shield (BCBS) of Massachusetts, a population-based payment model that includes financial rewards and penalties (two-sided risk).

METHODS

Using a difference-in-differences method to analyze data from 2006 through 2016, we compared spending among enrollees whose physician organizations entered the AQC starting in 2009 with spending among privately insured enrollees in control states. We examined quantities of sentinel services using an analogous approach. We then compared process and outcome quality measures with averages in New England and the United States.

RESULTS

During the 8-year post-intervention period from 2009 to 2016, the increase in the average annual medical spending on claims for the enrollees in organizations that entered the AQC in 2009 was $461 lower per enrollee than spending in the control states (P<0.001), an 11.7% relative savings on claims. Savings on claims were driven in the early years by lower prices and in the later years by lower utilization of services, including use of laboratory testing, certain imaging tests, and emergency department visits. Most quality measures of processes and outcomes improved more in the AQC cohorts than they did in New England and the nation in unadjusted analyses. Savings were generally larger among subpopulations that were enrolled longer. Enrollees of organizations that entered the AQC in 2010, 2011, and 2012 had medical claims savings of 11.9%, 6.9%, and 2.3%, respectively, by 2016. The savings for the 2012 cohort were statistically less precise than those for the other cohorts. In the later years of the initial AQC cohorts and across the years of the later-entry cohorts, the savings on claims exceeded incentive payments, which included quality bonuses and providers' share of the savings below spending targets.

CONCLUSIONS

During the first 8 years after its introduction, the BCBS population-based payment model was associated with slower growth in medical spending on claims, resulting in savings that over time began to exceed incentive payments. Unadjusted measures of quality under this model were higher than or similar to average regional and national quality measures. (Funded by the National Institutes of Health.)

From the Department of Health Care Policy, Harvard Medical School (Z.S., M.E.C.), the Department of Medicine, Massachusetts General Hospital (Z.S.), the Department of Medicine, Tufts University School of Medicine, and Haven (D.G.S.), Boston, and the Graduate School of Arts and Sciences, Harvard University, Cambridge (Y.J.) — all in Massachusetts. Address reprint requests to Dr. Song at the Department of Health Care Policy, Harvard Medical School, 180A Longwood Ave., Boston, MA 02115, or at song@hcp.med.harvard.edu.

This article was updated on July 18, 2019, at NEJM.org.

N Engl J Med 2019;381:252-63.
DOI: 10.1056/NEJMsa1813621

The reform of health care payment systems has centered on moving providers away from fee-for-service payment. The most clinically comprehensive of the alternative payment models — population-based global payment — gives providers a spending target or budget for the entire continuum of care within a defined population. These providers, often working as accountable care organizations (ACOs), assume responsibility for spending and quality, earning shared savings if spending is below the target, and, in some models, sharing financial risk if spending exceeds the target.[1] Bonuses that are awarded for quality care help to mitigate incentives to underuse appropriate care that a budget may introduce.

Public and private payers have both stimulated growth in ACO arrangements. By 2018, a total of 561 provider organizations were participating in the Medicare Shared Savings Program, 41 in the Medicare Next Generation ACO Model, and 9 in the Pioneer ACO Model, accounting for 12.6 million beneficiaries or more than one fifth of the Medicare population.[2,3] State Medicaid programs have gradually begun to follow suit.[4-6] Enrollees in commercial insurance plans — the largest share of insured populations in the United States — make up the largest share of ACO participants, with more than 19 million enrollees in such arrangements as of 2017.[7,8]

Studies regarding the effects of ACO models have focused on the early years of the programs.[9] Medicare ACOs have shown modest savings on claims and improved experience for patients during the first 3 years, with net savings in a subgroup of ACOs after accounting for bonus payments.[10-13] Oregon's Medicaid global budget program reported savings on claims and some improvements in quality during the first 2 years.[14] Previous studies of the Alternative Quality Contract (AQC) of Blue Cross Blue Shield (BCBS) of Massachusetts showed savings on claims and improved quality during the first 4-year period, with net savings emerging in year 4 after accounting for provider incentive payments.[15-17]

We examined data for the AQC population and a control population to assess changes in spending, utilization, and quality under this large-scale global budget model, which includes financial incentives and penalties (two-sided risk), during the 8-year period from 2009 through 2016. Although some details of the AQC have evolved, its main features have remained unchanged. Providers receive shared savings if spending is below a risk-adjusted budget and incur shared losses if spending exceeds the budget.[18] Providers are evaluated on the quality of care through 64 measures (Table S1 in the Supplementary Appendix, available with the full text of this article at NEJM.org) and receive data and reports that help them to identify areas of potential improvement. The AQC was launched in 2009 in provider organizations that collectively cared for approximately 20% of the members of the BCBS health maintenance organization (HMO). These members were prospectively attributed to the organization of their primary care physician. About 85% of the members and providers in the BCBS network had joined the AQC by 2013, a percentage that remained stable through 2016.

Methods

Study Design

In Massachusetts, multiple efforts have been proposed for slowing the growth in health care spending.[19,20] The Centers for Medicare and Medicaid Services (CMS) launched its models for the Medicare Pioneer ACO and Shared Savings ACO in 2012.[21] Other private payers also expanded alternative payment models after the formation of the AQC.[22] Moreover, state regulation has aimed to limit health care spending to a predefined growth rate, and Medicaid ACOs were recently developed.[23] These factors caution against causal interpretation of associations between the AQC and outcomes and have informed our study design, which aims to isolate the effects of the AQC to the extent possible. The study was supported by the National Institutes of Health and was approved by the institutional review board at Harvard Medical School.

Data and Population

We analyzed all claims and enrollment data for the 11-year period from 2006 through 2016. BCBS enrollees were assigned to an AQC cohort if the organization of their primary care physician had joined the AQC, with the cohort defined according to the year of AQC entry (Table S2 in the Supplementary Appendix). Physicians or enrollees may have changed their organizational affiliation during the study period, which could result in withdrawal from the AQC and

potential reentry through a different affiliation. Therefore, we adopted an intention-to-treat framework and attributed all enrollees to an AQC cohort according to the initial year of entry, regardless of subsequent exit or reentry. We excluded enrollee-year observations in which an enrollee switched insurance plans or primary care physician midyear, since such a change could introduce other incentives that might affect health care use.

The control group for analyses of medical claims included enrollees in employer-sponsored commercial plans across the eight other northeastern states (Connecticut, Maine, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont) in the MarketScan Commercial Claims and Encounters Database, owned by Truven Health Analytics.[24] Control participants had enrolled in an HMO or a point-of-service plan, which required designating a primary care physician, similar to the plans in the AQC.[15,25] Moreover, control employers all continuously reported claims to Truven from 2008 through 2016.

In our main analyses, all the participants in the AQC cohorts and the control group had been enrolled for at least 1 calendar year. However, because of the broad national trend in employer-sponsored insurance populations moving from an HMO to a preferred provider organization (PPO) or other type of plan (e.g., a high-deductible health plan), we performed sensitivity analyses to compare two subgroups of participants in AQC cohorts and the control group who were continuously enrolled in our HMO sample for at least 5 years and for all 11 years, respectively.

Although cost-control efforts may have existed in control states, broad shifts to global payment by commercial payers were generally absent. Pennsylvania experimented with medical-home models, although the scale was limited.[26] Rhode Island implemented state-based affordability standards but contributed a small sample to the control group.[27] Nevertheless, the control group may not have been devoid of alternative payment models, so to the extent that other cost-control initiatives were present, our findings may be conservative relative to a hypothetical no-cost-control comparison group.

VARIABLES

In our spending analysis, the dependent variable was claims spending at the enrollee-year level, which reflects negotiated prices. We evaluated spending according to the site of care (inpatient or outpatient) and type of claim (facility or professional). For analyses of utilization (volume), the dependent variable was the number of services delivered.

In addition to age and sex, we derived individual risk scores using the Diagnostic Cost Groups (DxCG) model from Verisk Health, which predicts spending on the basis of demographic characteristics and diagnoses (analogous to the CMS Hierarchical Condition Category risk-adjustment model).[28,29]

For our quality analyses, we compared AQC data regarding process and outcome measures of ambulatory care with New England and national average quality performance from the Healthcare Effectiveness Data and Information Set (HEDIS) of the National Committee for Quality Assurance for 2007 through 2016.[30] We studied three domains of process measures (chronic disease management, adult preventive care, and pediatric care) and a set of outcome measures. The definition of each measure was binary in that it either met or did not meet a performance threshold for enrollees eligible for the measure (e.g., glycated hemoglobin testing for those with diabetes). An enrollee could be eligible for multiple measures. Organizational performance on a measure was the percentage of eligible patients in whom the measure of care had met the threshold. Each AQC organization had an aggregate quality score that was calculated on the basis of performance across all measures (with outcome measures triple weighted) in each year; this score determined the quality bonus and the size of shared savings and shared risk under the budget (Table S1 in the Supplementary Appendix). We averaged HEDIS quality measures into the same domains for comparison with the AQC cohorts. Because HEDIS quality data were not available at the individual level, quality analyses were unadjusted.

STATISTICAL ANALYSIS

We compared spending and utilization in the AQC cohorts with those in the control group using a difference-in-differences approach within an ordinary least-squares regression model at the individual-year level (see the Methods section in the Supplementary Appendix).[31] To estimate changes in spending in large samples, we used

a linear model, which is often preferred in estimating averages despite less precision at the tails of the distribution.[32]

For the 2009 cohort, pre-intervention was defined as the period from 2006 through 2008; post-intervention was defined as the period from 2009 through 2016. Independent variables included age categories, interactions between age categories and sex, DxCG risk score, indicator variable for the AQC, year indicator variables, and interactions between the AQC and year — which produced our coefficients of interest. The model also included fixed effects for each individual insurance plan and the enrollee's state of residence to account for benefit design and time-invariant factors. Standard errors were clustered according to the individual plan.[33,34] The analyses of spending contained 1 outcome: total medical spending. For utilization, we analyzed 10 sentinel outcomes and adjusted for the family-wise error rate using the Bonferroni correction. We tested for differences in pre-intervention trends between the AQC and the control group.

We defined savings on claims in percentage terms as a decrease in spending for medical claims associated with the AQC divided by post-intervention spending in the AQC. To evaluate net savings, we compared savings on claims with incentive payments that providers received, including shared savings, quality bonuses, and infrastructure support (e.g., for electronic medical records). Incentive payments, which were audited by BCBS and providers, were proprietary and not observed at the contract level. However, we report these numbers as percentages of claims spending in ranges aggregated across cohorts and time, which allowed for the determination of rough comparisons with savings on claims.

To break down the changes in medical spending into changes in prices and in utilization, we first applied median prices at the claims level to estimate changes in spending that were due to utilization rather than price. Because this approach is fairly crude, we directly examined quantities of key services using an analogous difference-in-differences model.

We examined whether enrollment in the AQC was associated with changes in risk scores. We also performed sensitivity analyses. To separate the AQC effects from the Massachusetts secular trend, we compared AQC spending with the Massachusetts MarketScan sample, even though the latter contained BCBS enrollees whom we could not identify (thus producing a conservative estimate). We examined spending in prespecified analyses that included pharmaceutical claims (which were excluded from the main analyses, since not all enrollees had drug benefits) and tested other changes to the model. P values were calculated only for the primary analysis because there was no adjustment for multiple comparisons; confidence intervals alone are reported for other key comparisons.

RESULTS

SPENDING ON CLAIMS

Table 1 shows the characteristics of the AQC and control populations, with further details provided in Table S2 in the Supplementary Appendix. In the 2009 cohort, unadjusted spending grew more slowly after entry in the AQC than in the control group and in the overall population of commercially insured enrollees in Massachusetts (Fig. 1). The largest gap in spending between the AQC and the control group was in outpatient facilities (Fig. S1 in the Supplementary Appendix). For the cohorts that entered the AQC in 2010 through 2012, analogous plots are shown in Figures S2, S3, and S4 in the Supplementary Appendix. (Plots with the subsample of persons who were continuously enrolled for all 11 years are shown in Figures S5 through S8 in the Supplementary Appendix.)

In adjusted analysis, during the 8-year period from 2009 to 2016, the increase in the average annual medical spending per enrollee on claims in the 2009 AQC cohort was lower than the increase in the average medical spending in the control states by $461 (95% confidence interval [CI], −576 to −346; P<0.001), an 11.7% relative savings on claims (Table 2, and Table S3 in the Supplementary Appendix). The difference in pre-intervention trends between the two groups was not significant (P=0.55). The savings on claims in the comparison between the AQC and the Massachusetts control group (−$434; 95% CI, −689 to −178) was similar to the overall comparison, as were other sensitivity analyses (Table S4 in the Supplementary Appendix). The savings were larger in the more stable samples of participants who had been continuously enrolled for a minimum of 5 years and for all 11 years, in which the differences in pre-intervention trends

Table 1. Characteristics of the AQC and Control Populations.*

Characteristic	2009 Cohort	2010 Cohort	2011 Cohort	2012 Cohort	Control Group
No. of enrollees	613,054	239,544	133,063	699,878	1,039,469
Age (yr)	35.5±18.7	37.9±18.1	42.3±15.1	32.8±19.7	33.7±18.4
Female sex (%)	52.0	51.4	52.4	51.7	50.0
DxCG risk score†					
Mean	1.10	1.16	1.31	1.08	0.94
Median (IQR)	0.50 (0.20–1.11)	0.53 (0.21–1.20)	0.62 (0.24–1.38)	0.47 (0.19–1.06)	0.37 (0.13–0.91)
Enrollee percentage of cost sharing					
Mean	12.1	12.1	12.8	10.6	18.7
Median (IQR)	8.3 (3.7–15.8)	8.3 (4.0–15.5)	8.2 (3.6–16.7)	7.1 (3.3–13.4)	14.3 (7.4–24.9)
No. of provider organizations‡	7	4	1	5	NA
Type of provider (no.)					
Primary care physician	1151	469	420	2115	NA
Specialist	2197	1010	1319	7260	NA
Affiliated hospital	15	13	2	10	NA

* Plus–minus values are means ±SD. Beneficiaries were enrolled for at least 1 year during the study period. Enrollees in the Alternative Quality Contract (AQC) of Blue Cross Blue Shield of Massachusetts were required to designate a primary care physician. The control group consisted of enrollees in similar employer-sponsored insurance plans in eight other northeastern states (Connecticut, Maine, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont), in which the AQC was not offered. No data on provider organizations were available for controls. Data regarding age, sex, risk score, and cost sharing were pooled across all enrollees in the entire study period. Cost sharing is the portion of spending paid by the enrollee (the sum of deductibles, copayments, and coinsurance premiums) and is calculated as an annual percentage. IQR denotes interquartile range, and NA not available.

† The Diagnostic Cost Groups (DxCG) risk score is a measure of enrollee health status that is calculated with the use of coefficients from a statistical model that relates spending to diagnoses and demographic characteristics. The DxCG risk score is similar to the Medicare Hierarchical Condition Category risk score and is commonly used for risk adjustment. The average risk score across all plan participants is approximately 1. Higher values denote higher expected spending.

‡ The numbers of provider organizations and providers were reported at the beginning of the contract for each cohort. During the contract, enrollees may have entered or left the cohort. We used an intention-to-treat framework in which all the physicians who were initially included in the contract continued to be designated as a part of the treatment cohort throughout the duration of the study period.

were also not significant (Tables S5 and S6 in the Supplementary Appendix).

Analogously, in the 2010 cohort, the increase in the average annual medical spending per enrollee in the AQC was lower than control spending by $477 (95% CI, −608 to −347), an 11.9% relative savings; lower by $312 (95% CI, −483 to −141), a 6.9% relative savings, in the 2011 cohort; and lower by $102 (95% CI, −225 to 22), a 2.3% relative savings, in the 2012 cohort (Table 2). The between-group difference in pre-intervention trends was not significant in the 2010 cohort (P=0.23) or in the 2012 cohort (P=0.52), but the increase in spending was significantly slower in the 2011 cohort (P=0.04). Estimates relative to the Massachusetts comparison group are provided in Tables S7 and S8 in the Supplementary Appendix. Spending reductions were generally larger in the more stable subgroups of patients who were enrolled for at least 5 years or for all 11 years, in which pre-intervention trends were generally not significantly different from control trends (Tables S5 and S6 in the Supplementary Appendix).

Unadjusted risk scores are shown in Figures S9 through S12 in the Supplementary Appendix. Between-group differences in risk scores were unchanged in the 2009 AQC cohort relative to the control group (−0.02; 95% CI, −0.04 to 0.00); those in the 2010–2012 cohorts changed by −0.05 (95% CI, −0.07 to −0.02) to −0.09 (95% CI, −0.14 to −0.04) relative to the control group. Between-group differences in risk scores were generally smaller in magnitude or were not significant in the subgroups of patients who were enrolled for at least 5 years or for all 11 years (Tables S9, S10, and S11 in the Supplementary Appendix; sample sizes are shown in Table S12). Although there was a secular trend in employer-sponsored insurance plans that were moving away from HMO

plans, less attrition occurred in the AQC than in the control group.

PRICE VERSUS UTILIZATION

A general breakdown of the change in medical spending on claims in the 2009 cohort relative to the control group on the basis of median prices showed that 71% of the relative decrease in spending was attributable to lower provision of services during the 8-year period. The differences during the early years of the contract were explained by lower prices achieved through referrals to lower-priced providers,[15-17] whereas in later years the difference was more often explained by lower utilization (Table S13 in the Supplementary Appendix).

Supporting these results are direct analyses of the level of utilization in the AQC and the control group (Table 3, and Table S14 in the Supplementary Appendix). Across all AQC cohorts after 2009, a lower frequency of emergency department visits, radiography and echocardiography, and laboratory testing was observed than in the control group. The use of computed tomography was lower in the 2009–2011 AQC cohorts but not in the 2012 cohort, whereas changes in magnetic resonance imaging, positron-emission tomography, and nuclear imaging were more mixed. In some AQC cohorts, the number of prescriptions for specialty drugs was lower than that in the control groups. For preventive care, there were mixed results with respect to the use of colonoscopy among enrollees between the ages of 50 and 85 years and the use of mammography among women 40 years of age or older. No consistent between-group differences in changes were observed for inpatient admissions or outpatient visits or consultations. Results for the comparisons between the AQC and the Massachusetts comparison group are shown in Table S15 in the Supplementary Appendix.

NET FISCAL PERFORMANCE

Weighted average savings on claims (unadjusted and adjusted) were compared with unadjusted BCBS incentive payments. In the 2009–2010 cohorts, claims savings were exceeded by incentive payments in the early years — a period of initial investments. In later years, claims savings generally exceeded incentive payments to produce net savings, especially in more stably enrolled samples (Table 2, and Tables S16 and S17 in the Supplementary Appendix). In the 2011–2012 cohorts, savings were generally larger than incentive payments. Claims savings during the period from 2009 through 2012 differed from those in previous evaluations because of differences in the control group, which in this study has been further restricted to employers that continuously reported claims through 2016.[15-17]

Missing from this comparison were any incentive payments to providers in control states, which were not captured in the claims. Any quality bonuses, shared savings under alternative payment models, or other incentive payments would render claims spending in control states a conservative estimate of total system spending in those states.

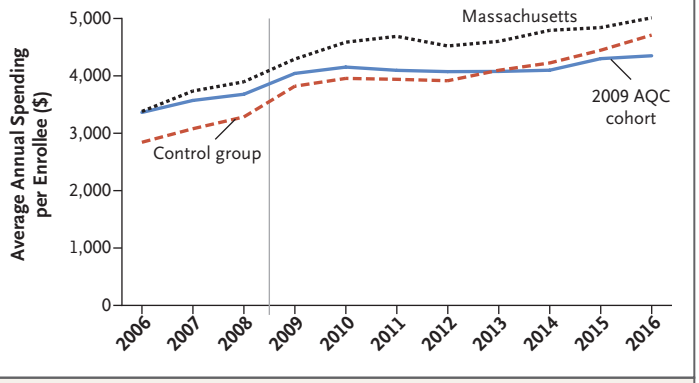

Figure 1. Medical Spending on Claims in the 2009 AQC and Control Populations.

Shown is the unadjusted medical spending on claims for the 2009 Alternative Quality Contract (AQC) cohort of Blue Cross Blue Shield of Massachusetts, the control group consisting of enrollees in similar employer-sponsored plans across eight northeastern states (Connecticut, Maine, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont), and enrollees in similar employer-sponsored plans in Massachusetts, as defined in the MarketScan Commercial Claims and Encounters Database. This Massachusetts comparison group includes Blue Cross Blue Shield of Massachusetts enrollees, who could not be separated from enrollees of other private insurers in the state. The gray vertical line indicates the initiation of the AQC.

QUALITY

Unadjusted quality measures for the 2009 cohort and New England and national averages are shown in Figure 2. Within process measures, the percentage of eligible enrollees who met the criteria for quality care with respect to chronic disease management (e.g., diabetes care) improved from an average of 81% before the initiation of the AQC to 88% after the initiation,

Table 2. Changes in Medical Spending and Net Fiscal Performance in the AQC and Control Populations.*

Year†	AQC Cohort		Control Group		Difference in Differences		Average Savings on Medical Claims‡		BCBS Incentive Payments to Providers§	
	Unadjusted Pre-AQC	Unadjusted Post-AQC	Unadjusted Pre-AQC	Unadjusted Post-AQC	Adjusted (95% CI)	Relative Change	First Half of Contract¶	Second Half of Contract¶	First Half of Contract¶	Second Half of Contract¶
	$ (U.S.)					%	*unadjusted (adjusted)* %		%	
2009 and 2010 cohorts							8.3 (9.0)	18.2 (14.2)	16–17	13–14
2009 cohort‖	3,409	3,946	3,098	4,066	−461 (−576 to −346)	−11.7				
2010 cohort	3,824	4,022	3,282	4,121	−477 (−608 to −347)	−11.9				
2011 and 2012 cohorts							8.0 (4.7)	16.7 (2.0)	2–3	1–2
2011 cohort	4,531	4,514	3,398	4,157	−312 (−483 to −141)	−6.9				
2012 cohort	4,172	4,444	3,484	4,226	−102 (−225 to 22)	−2.3				

* All values for medical spending (inflation-adjusted to 2016 dollars) are for each enrollee per year. Adjusted results were obtained from a difference-in-differences regression analysis that evaluated changes in spending in the Blue Cross Blue Shield (BCBS) AQC cohorts minus those in the control group (consisting of eight other northeastern states) after adjustment for covariates. Adjusted results for each cohort were scaled to percentages relative to the cohort's average spending level after joining the AQC. Comparisons of pre-intervention trends between each of the four AQC cohorts and the control group produced P values of 0.55, 0.23, 0.04, and 0.52 with the use of joint F-tests of the hypothesis that pre-intervention interactions between the AQC indicator and year indicators were zero.

† Values for net fiscal performance for average savings on claims and incentive payments are presented collectively for the 2009–2010 and 2011–2012 cohorts over the first and second halves of their contract periods owing to the confidentiality of contracts between BCBS and provider organizations that precluded reporting of incentive payments according to the cohort or year level. Of note, any incentive payments that were made outside of claims from insurers to providers in the control group were not captured in the claims data for the control group. Thus, comparisons of net spending between the AQC cohorts and the control group are limited by the absence of incentive payments by control-group payers.

‡ Average savings on claims were weighted across cohorts and years; they were scaled into percentages by dividing average savings on claims by average fee-for-service claims spending weighted across cohorts and years. Unadjusted savings were calculated on the basis of changes in raw spending on claims observed in the AQC cohorts and the control group. Adjusted savings were calculated on the basis of results from the difference-in-differences analyses.

§ BCBS incentive payments included shared savings under the budget, quality bonuses, and infrastructure bonuses for all enrollees in the program (with no exclusion criteria for continuous enrollment of any duration), scaled into percentages by dividing the incentive payments by average fee-for-service claims spending weighted across cohorts and years for all enrollees in the program. Beginning in 2011, the AQC model moved from a fixed-budget trend to a trend based on the average across the BCBS network. Moreover, quality bonuses evolved from a percentage of the budget to a per-member-per-month payment (not based on the budget level). On average, these bonuses resulted in lower average incentive payments over time. These payments were reported in ranges owing to the confidentiality of contracts between the BCBS and provider organizations. All the listed values in this category are unadjusted.

¶ For the 2009–2010 cohorts, the first half of contracts spanned the year of entry through 2012, and the second half spanned the 2013–2016 period. For the 2011–2012 cohorts, the first half of contracts spanned the year of entry through 2013, and the second half spanned the 2014–2016 period.

‖ P<0.001 for this primary comparison.

Table 3. Changes in Utilization in the 2009 AQC and Control Populations.*

Category of Service†	2009 AQC Cohort			Control Group			Difference in Differences‡		
	Pre-AQC (2006–2008)	Post-AQC (2009–2016)	Difference	Pre-AQC (2006–2008)	Post-AQC (2009–2016)	Difference	Unadjusted	Adjusted (95% CI)	Relative Change
	number of services/1000 enrollees/yr								%
Preventive care									
Colonoscopy	178.3	181.9	3.6	141.1	133.5	−7.6	11.2	18.3 (5.8 to 30.8)	10.1
Mammography	1333.7	1565.3	231.7	943.6	1116.1	172.5	59.1	60.2 (−15.7 to 136.2)	3.8
Imaging									
Radiography or echocardiography	874.3	840.0	−34.3	754.8	801.1	46.2	−80.6	−40.0 (−72.5 to −7.6)	−4.8
CT	128.4	99.3	−29.1	99.1	89.3	−9.9	−19.2	−13.5 (−21.0 to −6.0)	−13.6
MRI/PET/nuclear imaging	143.0	108.0	−35.0	126.1	109.4	−16.7	−18.3	−4.8 (−16.1 to 6.6)	−4.4
Specialty-drug prescription	54.4	62.5	8.1	50.6	60.6	10.0	−1.9	−13.1 (−24.7 to −1.5)	−21.0
Laboratory test	7929.5	8232.4	302.9	5766.1	6365.4	599.3	−296.4	−1365.9 (−1728.3 to −1003.4)	−16.6
Office visit or consultation	4122.6	4352.0	229.4	3967.0	4149.6	182.6	46.8	−74.4 (−183.6 to 34.8)	−1.7
Emergency department visit	279.3	273.6	−5.7	170.4	183.7	13.2	−18.9	−34.8 (−57.1 to −12.5)	−12.7
Inpatient admission	54.9	53.3	−1.7	52.5	50.5	−2.0	0.4	0.9 (−2.1 to 3.8)	1.6

* All participants were enrolled for at least 1 calendar year. Adjusted results for each cohort were scaled to percentages relative to the average utilization in the cohort after joining the AQC. CT denotes computed tomography, MRI magnetic resonance imaging, and PET positron-emission tomography.

† Utilization is expressed as the number of services per 1000 enrollees per year. Consistent with screening guidelines, the rate of colonoscopy was reported in enrollees who were between the ages of 50 and 85 years, and the rate of mammography was studied in female enrollees who were 40 years of age or older.

‡ Unadjusted between-group differences were calculated as the difference in the changes between the AQC cohort and the control group (i.e., enrollees in eight other northeastern states). Adjusted between-group differences are estimates from the statistical model, which controls for covariates with adjustment for the family-wise error rate with the use of the Bonferroni correction.

The New England Journal of Medicine

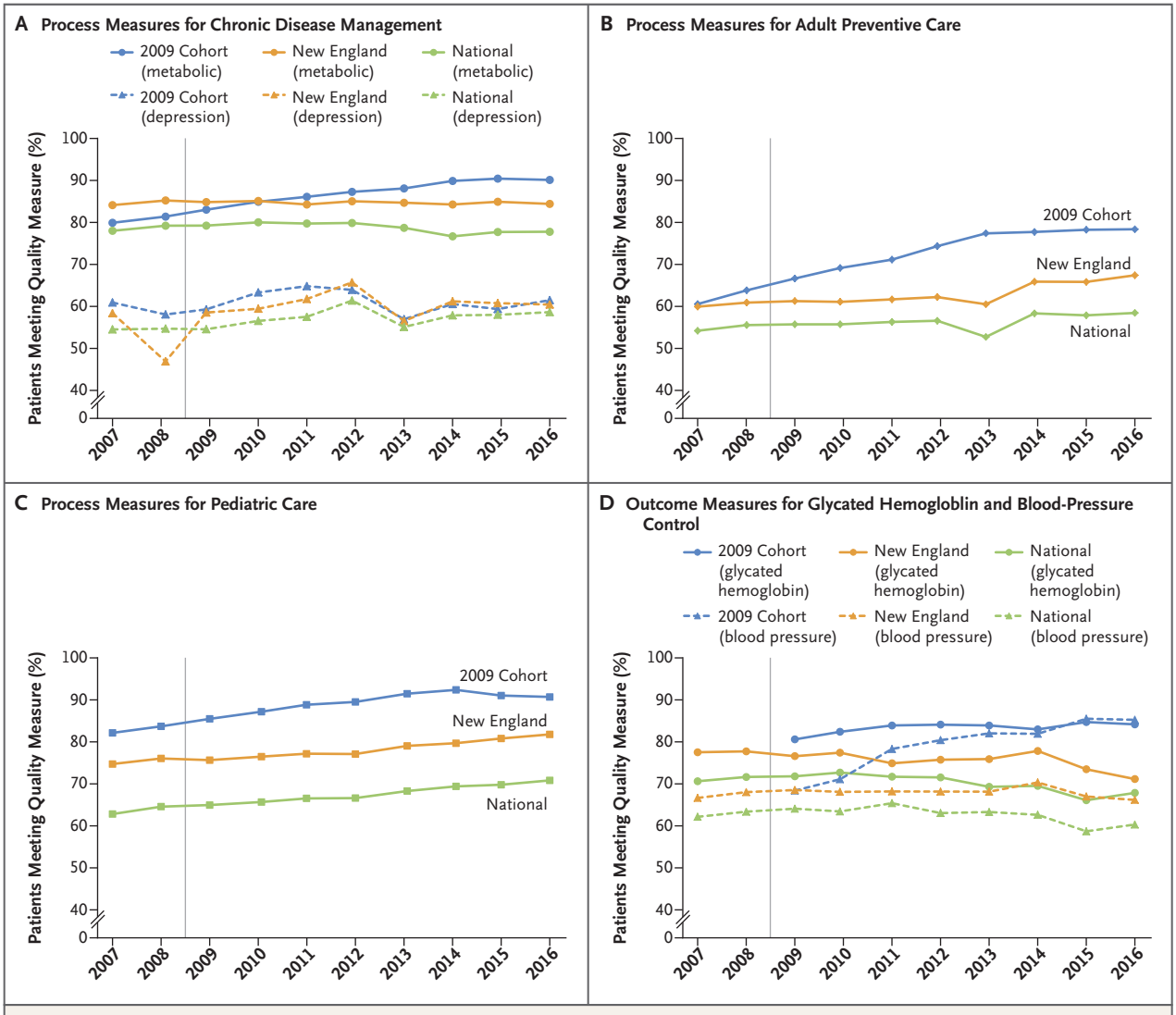

Figure 2. Quality Measures of Process and Outcome in the 2009 AQC Cohort, as Compared with New England and National Averages.

Shown are process quality measures, which were divided into three domains: chronic disease management (Panel A), adult preventive care (Panel B), and pediatric care (Panel C). All the process measure plots are averages of individual measures in each domain, as outlined below. Also shown are outcome quality measures, including blood-pressure control for enrollees with hypertension (target level, <140/90 mm Hg) and glycated hemoglobin control for enrollees with diabetes (target level, <9%) (Panel D). The gray vertical line indicates the initiation of the AQC. The domain of chronic disease management included six measures: cardiovascular testing for screening of low-density lipoprotein (LDL) cholesterol; glycated hemoglobin testing, eye examination, and nephropathy screening for enrollees with diabetes (metabolic subcategory); and short-term and maintenance pharmacologic treatment for enrollees with depression (depression subcategory). The domain of adult preventive care included five measures: breast, cervical, and colorectal cancer screening; chlamydia screening for enrollees between the ages of 21 and 24 years; and no prescription of antibiotics for acute bronchitis. The domain of pediatric care included six measures: appropriate testing for pharyngitis; chlamydia screening for adolescents between the ages of 16 and 20 years; no prescription of antibiotics for upper respiratory infection; and well care for babies under the age of 15 months, children between the ages of 3 and 6 years, and adolescents between the ages of 12 and 21 years. No pre-AQC data at the enrollee level were available for outcome measures in the AQC. There were changes in definitions for three other outcome measures — blood-pressure control in enrollees with diabetes and control of LDL cholesterol in enrollees with diabetes or cardiovascular disease — or the measures were discontinued by the National Committee for Quality Assurance. The results of those measures are provided in Figure S13 in the Supplementary Appendix.

whereas New England and national averages were unchanged at 85% and 79%, respectively. Measures for the treatment of depression trended similarly to the New England and national averages, with values generally ranging from approximately 55 to 65%. The percentage of enrollees who met the criteria for quality care with respect to adult preventive care improved from 62% before the initiation of the AQC to 74% after the initiation. New England and national averages improved from 60% to 63% and 55% to 57%, respectively. Measures of pediatric care in the AQC improved from 83% to 90%, as compared with improved values of 75% to 79% for New England and 64% to 68% nationally.

Outcome measures for hypertension and control of glycated hemoglobin among enrollees with diabetes — the only measures with complete post-AQC data and no changes in the measure definition — improved from 75% in 2009 to 85% in 2016. Meanwhile, New England and national averages declined slightly (Fig. 2D). In the Supplementary Appendix, outcome measures with incomplete data are shown in Figure S13, and quality measures in the 2010–2012 cohorts are shown in Figures S14, S15, and S16.

DISCUSSION

Medical spending on claims in the AQC grew slower than spending in the control group during the 8-year period from 2009 through 2016. In the early years after the AQC initiation, these savings on claims — which reflect changes in provider behavior — were largely generated through referring patients to lower-priced providers or places of service. In later years, the savings on claims were generated increasingly through lower utilization, including the use of laboratory testing, certain imaging tests, and emergency department visits. The use of some services declined among all the AQC cohorts, whereas changes in the use of other services varied.

Although it was challenging to measure the net fiscal performance of the AQC against the control group, savings on claims exceeded incentive payments in the later years of the initial AQC cohorts and across the years of the later-entry cohorts.[15] Most quality measures for enrollees in the AQC were better than the New England and national averages of the National Committee for Quality Assurance, although a lack of enrollee-level comparison data precluded the use of statistical analyses. It is likely that the changes in provider behavior in the AQC cohorts were aided by the contract's built-in incentives, by data and reports from BCBS, and by peer support among the providers. Savings on claims in the AQC, which budgeted the entire continuum of care, appeared to be larger than in other models that budgeted a segment of care, such as inpatient spending[35-37] and patient-centered medical-home models.[26,38,39]

We did not find a greater intensity of diagnostic disease coding on claims (which would denote a larger disease burden and garner larger global budgets, a concern for risk-adjusted payment models) among providers in the AQC than in the control group. On the contrary, the differential decreases in AQC risk scores may be explained by an increase in coding intensity that took place in the control populations or by changes in health status, which are difficult to separate. The latter could reflect changes in health (possibly attributable to the AQC) or in the case mix — perhaps enrollees with lower risk scores joined the AQC and higher-risk ones withdrew (e.g., to join PPOs or other insurers). Although members who were enrolled longer in the AQC had lower risk scores than those with a shorter duration of enrollment, the rate of attrition in the AQC was substantially lower than that in the control group, which suggests that the introduction of the AQC probably did not induce substantial withdrawal from the HMO population.

This study has several limitations. First, conditions in Massachusetts (e.g., the presence of Medicare ACOs, payment reform among other commercial payers, and state policies) may have contributed to the findings, especially in recent years.[40] Such factors are difficult to disentangle from the AQC effects. However, conservative estimates that are based on data from the Massachusetts control group (which also contained BCBS enrollees who would have cancelled out any AQC effects) still suggested savings on claims in the AQC, although some estimates were, as expected, less statistically precise (Tables S7 and S8 in the Supplementary Appendix).

Second, control states may have pursued cost-

control methods, such as affordability standards in Rhode Island. If such efforts slowed spending, our estimated savings may be conservative. Third, voluntary participation in the AQC invokes concern about selection bias, although providers faced disincentives for nonparticipation and the vast majority of providers in the BCBS network had entered the AQC by 2012.

Fourth, our results may not have generalizability for other ACO arrangements (e.g., one-sided models in which providers receive potential financial rewards but not risks), other payers (e.g., Medicare, which has largely uniform prices), or other states. Finally, the association between the AQC and quality is limited by unadjusted analysis owing to the lack of enrollee-level comparison data. However, previous adjusted analyses with the use of BCBS enrollees who were not in the AQC as controls showed better quality in most measures than that in the control group.[16,17]

In conclusion, during the 8-year period after the initiation of the AQC, the growth of spending on medical claims was lower in the AQC than in a control population. Changes in referral patterns during the early years of the contract were followed by reductions in utilization of certain services. These findings suggest that an ACO model with both financial rewards and penalties, including quality incentives, may offer a framework for slowing the growth in medical spending without sacrificing the quality of care for patients.

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Supported by a grant (DP5-OD024564, to Dr. Song) from the National Institutes of Health.

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

We thank Sarah Chiodi, Matthew Day, Gabriella Diamandis, Christian Lassonde, Angela Li, Yiwen Yang, and Wei Ying for their assistance with data and Andrew Hicks for the derivation of risk scores.

REFERENCES

1. Ginsburg PB, Patel KK. Physician payment reform — progress to date. N Engl J Med 2017;377:285-92.
2. Centers for Medicare & Medicaid Services. Medicare Shared Savings Program: fast facts. January 2018 (https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/SSP-2018-Fast-Facts.pdf).
3. Centers for Medicare & Medicaid Services. Next Generation Accountable Care Organization (ACO) model fact sheet. 2018 (https://innovation.cms.gov/Files/fact-sheet/nextgenaco-fs.pdf).
4. McConnell KJ. Oregon's Medicaid coordinated care organizations. JAMA 2016;315:869-70.
5. Blewett LA, Spencer D, Huckfeldt P. Minnesota Integrated Health Partnership demonstration: implementation of a Medicaid ACO model. J Health Polit Policy Law 2017;42:1127-42.
6. MassHealth. MassHealth accountable care organizations (ACOs). March 6, 2018 (https://www.mass.gov/files/documents/2018/03/06/MassHealth%20ACO%20One-Pager_0.pdf).
7. Muhlestein D, Saunders R, McClellan M. Growth of ACOs and alternative payment models in 2017. Health Affairs Blog. June 28, 2017 (https://www.healthaffairs.org/do/10.1377/hblog20170628.060719/full/).
8. Catalyst for Payment Reform. National scorecard on payment reform. 2014 (https://www.catalyze.org/product/2014-national-scorecard/).
9. Song Z, Fisher ES. The ACO experiment in infancy — looking back and looking forward. JAMA 2016;316:705-6.
10. McWilliams JM, Hatfield LA, Landon BE, Hamed P, Chernew ME. Medicare spending after 3 years of the Medicare Shared Savings Program. N Engl J Med 2018;379:1139-49.
11. McWilliams JM, Chernew ME, Landon BE, Schwartz AL. Performance differences in year 1 of Pioneer accountable care organizations. N Engl J Med 2015;372:1927-36.
12. Nyweide DJ, Lee W, Cuerdon TT, et al. Association of Pioneer Accountable Care Organizations vs traditional Medicare fee for service with spending, utilization, and patient experience. JAMA 2015;313:2152-61.
13. McWilliams JM, Landon BE, Chernew ME, Zaslavsky AM. Changes in patients' experiences in Medicare accountable care organizations. N Engl J Med 2014;371:1715-24.
14. McConnell KJ, Renfro S, Lindrooth RC, Cohen DJ, Wallace NT, Chernew ME. Oregon's Medicaid reform and transition to global budgets were associated with reductions in expenditures. Health Aff (Millwood) 2017;36:451-9.
15. Song Z, Rose S, Safran DG, Landon BE, Day MP, Chernew ME. Changes in health care spending and quality 4 years into global payment. N Engl J Med 2014;371:1704-14.
16. 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.
17. 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.
18. 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.
19. Ayanian JZ, Van der Wees PJ. Tackling rising health care costs in Massachusetts. N Engl J Med 2012;367:790-3.
20. Song Z, Landon BE. Controlling health care spending — the Massachusetts experiment. N Engl J Med 2012;366:1560-1.
21. Centers for Medicare & Medicaid Services. Pioneer ACO model fact sheet. July 30, 2018 (https://innovation.cms.gov/initiatives/Pioneer-ACO-Model/PioneerACO-FactSheet.html).
22. Tufts Health Plan. Coordinated care model. 2011 (http://www.tuftshealthplan.com/employers/pdfs/coordinated_care_brochure.pdf).
23. Massachusetts Health Policy Commission. 2017 Annual health care cost trends report. March 2018 (https://www.mass.gov/files/documents/2018/03/28/Cost%20Trends%20Report%202017.pdf).
24. Hansen L. The Truven Health MarketScan Databases for life sciences researchers. Truven Health Analytics, IBM Watson Health white paper. March 2017 (https://

truvenhealth.com/Portals/0/Assets/2017-MarketScan-Databases-Life-Sciences-Researchers-WP.pdf).
25. Song Z, Rose S, Chernew ME, Safran DG. Lower- versus higher-income populations in the Alternative Quality Contract: improved quality and similar spending. Health Aff (Millwood) 2017;36:74-82.
26. 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.
27. Koller CF, Brennan TA, Bailit MH. Rhode Island's novel experiment to rebuild primary care from the insurance side. Health Aff (Millwood) 2010;29:941-7.
28. DxCG risk solutions: DxCG science guide. Waltham, MA: Verisk Health, 2012.
29. 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.
30. National Committee for Quality Assurance. Quality Compass: benchmark and compare quality data (http://www.ncqa.org/HEDISQualityMeasurement/QualityMeasurementProducts/QualityCompass.aspx).
31. Wooldridge J. Econometric analysis of cross section and panel data. Cambridge, MA: MIT Press, 2001.
32. 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.
33. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 1980;48:817-30.
34. 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, CA, December 27, 1965–January 7, 1966. Berkeley: University of California Press, 1967:221-33.
35. Roberts ET, Hatfield LA, McWilliams JM, et al. Changes in hospital utilization three years into Maryland's global budget program for rural hospitals. Health Aff (Millwood) 2018;37:644-53.
36. Roberts ET, McWilliams JM, Hatfield LA, et al. Changes in health care use associated with the introduction of hospital global budgets in Maryland. JAMA Intern Med 2018;178:260-8.
37. McConnell KJ, Renfro S, Chan BK, et al. Early performance in Medicaid accountable care organizations: a comparison of Oregon and Colorado. JAMA Intern Med 2017;177:538-45.
38. Dale SB, Ghosh A, Peikes DN, et al. Two-year costs and quality in the Comprehensive Primary Care Initiative. N Engl J Med 2016;374:2345-56.
39. Friedberg MW, Rosenthal MB, Werner RM, Volpp KG, Schneider EC. Effects of a medical home and shared savings intervention on quality and utilization of care. JAMA Intern Med 2015;175:1362-8.
40. Massachusetts Health Policy Commission. 2017 Health care cost trends report. March 28, 2018 (https://www.mass.gov/doc/2017-health-care-cost-trends-report).