

Original Investigation

Evaluating a Pay-for-Performance Program for Medicaid Children in an Accountable Care Organization

Sean Gleeson, MD; Kelly Kelleher, MD; William Gardner, PhD

IMPORTANCE Pay for performance (P4P) is a mechanism by which purchasers of health care offer greater financial rewards to physicians for improving processes or outcomes of care. To our knowledge, P4P has not been studied within the context of a pediatric accountable care organization (ACO).

OBJECTIVE To determine whether P4P promotes pediatric performance improvement in primary care physicians.

DESIGN, SETTING, AND PARTICIPANTS This retrospective cohort study was conducted from January 1, 2010, to December 31, 2013. A differences-in-differences design was used to test whether P4P improved physician performance in an ACO serving Medicaid children. Data were obtained from 2966 physicians and 323 812 patients. Three groups of physicians were identified: (1) community physicians who received the P4P incentives, (2) nonincentivized community physicians, and (3) nonincentivized physicians employed at a hospital.

INTERVENTION Pay for performance.

MAIN OUTCOMES AND MEASURES Healthcare Effectiveness Data Information Set measure rates for preventive care, chronic care, and acute care primary care services. We examined 21 quality measures, 14 of which were subject to P4P incentives.

RESULTS There were 203 incentivized physicians, 2590 nonincentivized physicians, and 173 nonincentivized hospital physicians. Among them, the incentivized community physicians had greater improvements in performance than the nonincentivized community physicians on 2 of 2 well visits (largest difference was for adolescent well care: odds ratio, 1.05; 99.88% CI, 1.02-1.08), 3 of 10 immunization-incentivized measures (largest difference was for inactivated polio vaccine: odds ratio, 1.14; 99.88% CI, 1.07-1.21), and 2 nonincentivized measures (largest difference was for rotavirus: odds ratio, 1.11; 99.88% CI, 1.04-1.18). The employed physician group at the hospital had greater improvements in performance than the incentivized community physicians on 8 of 14 incentivized measures and 1 of 7 nonincentivized measures (largest difference was for hepatitis A vaccine: odds ratio, 0.34; 99.88% CI, 0.31-0.37).

CONCLUSIONS AND RELEVANCE Pay for performance resulted in modest changes in physician performance in a pediatric ACO, but other interventions at the disposal of the ACO may have been even more effective. Further research is required to find methods to enhance quality improvements across large distributed pediatric health systems.

JAMA Pediatr. doi:10.1001/jamapediatrics.2015.3809
Published online January 25, 2016.

Author Affiliations: Partners for Kids, Columbus, Ohio (Gleeson, Kelleher); Department of Pediatrics, Ohio State University, Columbus (Gleeson, Kelleher); Nationwide Children's Hospital, Columbus, Ohio (Gleeson, Kelleher); Research Institute, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada (Gardner).

Corresponding Author: Sean Gleeson, MD, Partners for Kids, Nationwide Children's Hospital, 700 Children's Dr, Columbus, OH 43205 (sean.gleeson@nationwidechildrens.org).

Pay for performance (P4P) is a mechanism by which purchasers offer financial rewards to physicians on the basis of predetermined criteria to increase quality outcomes. Pay-for-performance programs are decades old, but the use of P4P is growing within state Medicaid programs,¹ private health plans,² and internationally.³

There are conflicting reports about the effectiveness of these incentives,⁴⁻⁶ and some aspects of P4P programs remain understudied.⁶⁻⁸ First, because there have been fewer pediatric studies, it is unclear whether P4P will be as effective for pediatric care as it has been for adult care.⁹ Second, to our knowledge, there are no studies addressing P4P programs within accountable care organizations (ACOs). Accountable care organizations, groups of physicians who join together and accept clinical and financial risk for a population, are proliferating in the marketplace but are relatively unstudied. The Alternative Quality Contract in Massachusetts is a payer-initiated P4P that used a quality bonus combined with risk-sharing contracts¹⁰ to influence population health management. Most P4P programs are initiated by a payer, not a physician-driven entity such as an ACO. A physician-centric, ACO-developed P4P program may win greater physician acceptance because significant clinician design involvement was a success condition identified by previous researchers.^{11,12}

Partners for Kids (PFK) is an ACO accepting financial responsibility for Medicaid managed care children. Partners for Kids teamed with community physicians in an at-risk contracting arrangement and has increased value in the health care system by reducing cost without worsening quality.¹³ In this study, we examined PFK's effort to improve physician performance using P4P incentives. We assessed the effects of a P4P plan on improvement in quality measures in a pediatric ACO serving a Medicaid population. We compared 3 groups of primary care physicians. The first was community physicians who received the P4P incentives. There were also 2 nonincentivized groups: community physicians receiving no incentive and physicians employed at a hospital. We hypothesized that the community physicians receiving P4P incentives would improve their performance more over this period than either the nonincentivized community physicians or the salaried hospital physicians.

Methods

Setting

Partners for Kids is a physician-hospital organization taking medical and cost risk for 330 000 pediatric Medicaid recipients in Central and Southeast Ohio, a region containing substantial poverty. Partners for Kids contracts on behalf of its member physicians and Nationwide Children's Hospital (NCH), Columbus, Ohio, accepting risk for children through the Medicaid managed care plans. Partners for Kids is financially responsible for all Medicaid health plan members aged 0 to 18 years throughout a 34-county region.

Partners for Kids used 3 payment mechanisms with primary care physician groups from January 1, 2010, to December 31, 2013: (1) fee for service plus an incentive bonus was used

At a Glance

- Community physicians receiving pay-for-performance incentives achieved moderate but significant improvements in quality measures compared with other community physicians.
- Hospital-employed primary care physicians who were targeted with nonincentive-based improvement efforts achieved greater performance improvements than community physicians receiving pay-for-performance incentives.
- Accountable care organizations should use rigorous data analytics to study program implementations for effectiveness.

for the independent physicians contracted to PFK (referred to from here forward as *incentivized*); (2) exclusive fee-for-service payments were provided to the community physicians who were not contracted as members of PFK (*nonmembers*) but instead were directly contracted with the health plans; and (3) full capitation payment for professional services provided was paid to the members of the academic faculty practice plan of NCH. We have less information about the practice characteristics of the nonmembers because they did not directly contract with PFK.

This project used deidentified data, without any interaction with patients, and was therefore identified by the NCH institutional review board as exempt from review.

The Partners for Kids P4P Program

Beginning in 2012, PFK offered a P4P program emphasizing quality. Incentives were paid based on several factors. Practices received \$0.50 per member per quarter if they accepted at least 500 Medicaid members per physician averaged across the practice. They received an additional \$0.50 per member per quarter if they completed a PFK-approved Maintenance of Certification program or were recognized by the National Committee for Quality Assurance as a patient-centered medical home. Finally, the bulk of the incentive funds were dedicated to a select list of Healthcare Effectiveness Data Information Set measures.

Our study compared performance in 2010-2011 (preincentive) with performance in 2012-2013 (postincentive). The quality payments (\$40.18 in 2012 and \$41.39 in 2013) were made per successful patient and were paid to the patient's attributed physician group.

Population

To be included in the PFK population, individuals had to be 18 years old or younger, had to live in 1 of the 34 Central/Southeast Ohio counties, and had to belong to one of Ohio's Medicaid managed care plans. All patients meeting those criteria were patients of the ACO and were included in this study. Patients were attributed to a particular physician as follows. Patients seen by 1 or more primary care physicians were attributed to the physician who had the most visits in the prior 2 years. Those patients who had no primary care visits were attributed to the physician to whom Ohio Medicaid assigned them.

In most cases, all physicians in a practice fell into the same incentive condition. However, some large group practices le-

Table 1. Description of 132 Practices of 2966 Physicians

Variable	Practices Group			P Value
	Incentivized	Nonmembers	NCH	
No. of practices	50	80	2	NA
No. of PCPs	203	2590	173	NA
Unique patients, No. (%)	89 547 (27.7)	184 529 (57.0)	49 736 (15.4)	NA
Patient-years, No. (%)	196 989 (27.4)	407 765 (56.7)	114 520 (15.9)	NA
Period, %				
2010-2011	26.7	57.8	15.5	<.001
2012-2013	28.1	55.6	16.3	
Female, %	48.4	49.5	49.6	<.001
Age, mean (SD), y	8.3 (5.4)	10.1 (5.7)	8.1 (5.4)	<.001

Abbreviations: NA, not applicable; NCH, Nationwide Children's Hospital; PCPs, primary care physicians.

gally shared a tax identification number but consisted of independently run small-group practices using a common business office. Some of these small groups chose to contract with PFK and were incentivized, while others were not. In those situations, we treated the incentivized and nonincentivized groups within the provider organization as different practices.

Four physicians changed from a group in 1 incentive category to 1 in a different category during the study period. These doctors were not included in the study.

Measures

Claims data were used in this analysis. Partners for Kids' data vendor, Valence Health (<http://valencehealth.com/>), combined the claims files provided by the managed care plans into a single database. Valence Health then used the National Committee for Quality Assurance Healthcare Effectiveness Data Information Set 2013 standards¹⁴ to determine each patient's compliance with the quality standards.

Two categories of Healthcare Effectiveness Data Information Set measures were used: incentivized (2 well care, 2 asthma, and 10 immunization) and unincentivized (2 acute illness, 2 attention-deficit/hyperactivity disorder, 2 immunization, and a screening test) measures. The incentivized measures were used by PFK to determine whether incentivized physicians earned a P4P payment. An immunization P4P incentive was paid if the patient had received all its component immunizations. An additional well care measure and 2 components of the childhood immunization (Childhood Immunization Status) composite were part of the P4P but were excluded from the analysis owing to data reliability issues.

Nonincentivized measures had no bearing on the financial payments to physicians. We examined the nonincentivized measures to determine whether the P4P incentive affected only the incentivized measures or whether P4P's effects generalized to other clinician behaviors.

Design

This study used a differences-in-differences design, a method commonly used to distinguish changes from background trends.¹⁵ That is, we examined the differences in quality of care when comparing the years 2010-2011 (before the P4P incentive program) and 2012-2013 (during the incentive program). We compared those changes in the group of physicians who

received the incentives with each of the 2 groups of physicians that did not receive the incentives. If the P4P incentives work, there should be a greater improvement in quality for the incentivized physicians than in either of the other physician groups. Similarly, the P4P should improve the incentivized measures more than nonincentivized measures.

Physicians were not randomized into the incentivized, NCH, and nonmember groups. To the contrary, physician assignment to a group reflected both self-selection and selection by PFK, which sought to include the best community physicians in its incentive contract.

Statistical Analysis

The analysis focused on patient-level binary outcomes of whether the quality criterion was met for a patient in a given year. For each physician group and each criterion, we calculated an odds ratio (OR) that measured whether performance improved from 2010-2011 to 2012-2013.

In addition, for each quality criterion, a mixed-effects logistic regression model was used to compare the average odds that the criterion was met between the preincentive (2010-2011) and postincentive (2012-2013) periods among the incentivized, NCH, and nonmember physicians. Estimation of these differences-in-differences estimates was achieved by including group-by-period interaction terms in the regression model. All models adjusted for the patients' ages. All models included physician- and practice-level random intercepts to account for the correlation due to clustering of patient quality outcomes within physicians and practices. Results were expressed as ORs and, where necessary, inverted such that if the incentivized group's performance improved more from the preincentive to postincentive periods, then the OR was greater than 1.0.

These regression analyses focused on the 42 ORs that reflected the differences between physician incentive groups in the change in performance from 2010-2011 to 2012-2013 (ie, the differences in differences). To correct for our reporting of multiple comparisons, we reported these ORs with 99.88% CIs (because this level of confidence in 42 independent test leads to a 95% CI that fewer than 1 null test would be reported positive as a result of sampling error).

Analyses were performed using R 3.02 (R Foundation for Statistical Computing), using the `glmer` function in the `lme4` package.

Results

Sample Characteristics

We studied 2966 unique physicians in at least 132 practices (Table 1). The number of practice groups underreports the true number of practices because 57.9% (n = 1500) of the 2590 physicians in the nonmember category were listed under a single group name termed *other group* because we could not access the specific group practice affiliation status for these physicians. These physicians or groups did not contract with PFK. The proportions of patient counts and the patient-years over the study period were nearly identical, with trend toward smaller patient volumes at the older age groups. However, the relative proportion of patients attributed to the nonmember physicians declined slightly over the study period relative to the incentivized physicians and the NCH physicians, falling from 57.8% to 55.6% of total patient-years. There were slight differences in the proportions of female patients (48.4% vs 49.5% vs 49.6%). Patients cared for by nonmember physicians were almost 2 years older than patients of incentivized or NCH physicians (10.1 years vs 8.3 years, respectively).

Quality Scores

Table 2 presents all physicians' performances on the metrics calculated across the entire 2010-2013 period. Table 3 presents performances for each physician group in the 2010-2011 and 2012-2013 periods. Almost all measures improved for all physician groups over the years. The incentivized physicians performed better than nonmembers across the entire 2010-2013 period. The incentivized and NCH physicians had generally similar overall performance on measures across the study period. This overall trend of general improvement and the aforementioned group comparisons applied to both the incentivized and nonincentivized measures.

Figure 1 presents the differences between incentivized and nonmember physicians in the change in performance between 2010-2011 and 2012-2013. The incentivized physicians improved faster than the nonmember physicians on most of the incentivized measures. However, the ORs for these differences in differences were greater than 1 with 99.88% confidence for only 5 of 14 incentivized measures. The largest difference among these significant ORs (1.14; 99.88% CI, 1.07-1.21) was for inactivated polio vaccine and the smallest difference (1.04; 99.88% CI, 1.01-1.07) was for well child visits for those aged 3 to 6 years. These 5 differences included well child care and immunizations but not asthma.

The NCH physicians improved more rapidly than the incentivized physicians between 2010-2011 and 2012-2013 on most measures (OR < 1.0 in Figure 2). The ORs favoring the NCH physicians differed from 1.0 with 99.88% confidence for 8 of 14 incentivized measures primarily in the immunization measures. The largest difference among these significant ORs (0.34; 99.88% CI, 0.31-0.37) was for hepatitis A vaccine and the smallest difference (0.86; 99.88% CI, 0.78-0.95) was for meningococcal vaccine. Incentivized physicians showed more rapid improvement in well child care measures, and progress between the 2 groups for the asthma measures was the same.

Table 2. HEDIS 2013 Quality Measures

Measure	Performance, No./Total No. (%) ^a
Incentivized measures	
Adolescent well care visits	106 123/294 165 (36.1)
Well child visits at 3-6 y	148 819/247 289 (60.2)
Asthma, y	
12-18	6181/7850 (78.7)
5-11	11 050/13 422 (82.3)
Immunizations: adolescents	
Meningococcal	20 620/46 351 (44.5)
Td/Tdap	24 997/46 351 (53.9)
Immunizations: children	
DTP	22 101/61 656 (35.8)
Hepatitis A	38 436/61 656 (62.3)
IPV	29 574/61 656 (48)
MMR	47 845/61 656 (77.6)
Pneumococcal conjugate	23 740/61 656 (38.5)
Varicella	47 780/61 656 (77.5)
Nonincentivized measures	
Pharyngitis	24 459/36 828 (66.4)
Upper respiratory infection	74 642/91 263 (81.8)
ADHD	
Maintenance	1841/4641 (39.7)
Initiation	4980/14 315 (34.8)
Lead screening	38 574/61 656 (62.6)
Influenza	19 767/61 656 (32.1)
Rotavirus	21 031/61 656 (34.1)

Abbreviations: ADHD, attention-deficit/hyperactivity disorder; DTP, diphtheria, tetanus toxoids, and pertussis vaccine; HEDIS, Healthcare Effectiveness Data Information Set; IPV, inactivated polio vaccine; MMR, measles, mumps, and rubella; Td, tetanus-diphtheria toxoids; Tdap, tetanus, diphtheria, and pertussis.

^a The denominator of the performance measure is the count of patient-years for which we have data. The numerator is the count of patient-years in which the quality criterion was met.

There was no discernible pattern to the performance improvement in either physician group comparison for the nonincentivized measures, a marked difference from the incentivized measures.

Discussion

Comparing incentivized physicians with other community physicians within a pediatric Medicaid ACO, we found moderate but statistically significant differences in performance improvement on several quality outcomes. The differences-in-differences analysis distinguished the impact of the P4P program from the overall trends of improvement in the measured indices. This finding of an effect of P4P on pediatric well care and immunization measures is consistent with the literature showing that incentive programs can influence behavior sufficiently to change the results on quality measures.^{6,16-18} The modest nature of the improvements in metrics with P4P places our results squarely in the middle of the adult studies on P4P and is roughly consistent with the limited pediatric

Table 3. Change in Quality Measures by Years and Incentive Status

Measure	Group, %						Odds Ratio (99.88% CI) ^a		
	Incentivized		Nonmembers		NCH		Incentivized	Nonmembers	NCH
	2010-2011	2012-2013	2010-2011	2012-2013	2010-2011	2012-2013			
Incentivized measures									
Adolescent well care visits	41.0	43.3	32.1	32.7	42.7	44.0	1.10 (1.04-1.15)	1.03 (1.00-1.06)	1.05 (0.99-1.13)
Well child visits, 3-6 y	62.6	64.4	56.0	57.1	63.6	64.8	1.08 (1.03-1.13)	1.05 (1.00-1.08)	1.05 (1.00-1.11)
Asthma, y									
12-18	79.3	79.7	78.7	76.6	81.8	81.6	1.02 (0.71-1.43)	0.89 (0.70-1.09)	0.98 (0.61-1.49)
5-11	83.0	81.1	84.5	78.9	84.1	84.6	0.88 (0.68-1.11)	0.68 (0.56-0.84)	1.04 (0.74-1.40)
Immunizations: adolescents									
Meningococcal	39.9	52.8	27.3	40.2	49.9	70.3	1.68 (1.51-1.88)	1.79 (1.65-1.93)	2.38 (2.04-2.77)
Td/Tdap	45.0	55.8	30.6	42.7	54.4	72.3	1.54 (1.38-1.73)	1.69 (1.55-1.81)	2.19 (1.86-2.55)
Td/Tdap	49.7	65.2	40.8	56.2	55.4	77.5	1.89 (1.69-2.12)	1.86 (1.72-2.00)	2.77 (2.35-3.24)
Immunizations: children									
DTP	21.1	40.4	15.9	31.0	9.2	45.4	2.54 (2.29-2.79)	2.37 (2.17-2.59)	8.22 (6.99-9.56)
Hepatitis A	28.2	51.1	24.6	43.0	17.9	50.5	2.66 (2.43-2.91)	2.31 (2.14-2.50)	4.70 (4.12-5.32)
IPV	64.3	75.1	52.8	63.4	32.8	88.0	1.67 (1.53-1.84)	1.55 (1.44-1.66)	15.01 (12.92-17.32)
MMR	37.6	66.2	35.5	58.6	24.7	62.3	3.25 (2.97-3.56)	2.57 (2.38-2.76)	5.04 (4.46-5.68)
Pneumococcal conjugate	79.9	79.3	72.8	72.7	85.3	86.9	0.96 (0.87-1.07)	0.99 (0.91-1.07)	1.14 (0.97-1.33)
Varicella	32.1	51.9	26.9	43.0	31.8	50.6	2.29 (2.09-2.49)	2.04 (1.88-2.20)	2.19 (1.96-2.46)
Nonincentivized measures									
Pharyngitis	80.1	80.2	72.2	72.7	83.4	87.4	1.00 (0.91-1.12)	1.02 (0.94-1.11)	1.37 (1.17-1.60)
Upper respiratory infection	83.0	81.2	58.9	59.2	86.6	84.1	0.89 (0.75-1.03)	1.01 (0.93-1.09)	0.81 (0.51-1.20)
ADHD									
Maintenance	88.8	90.0	75.3	77.2	93.6	94.1	1.13 (1.01-1.29)	1.11 (1.05-1.19)	1.10 (0.83-1.41)
Initiation	40.0	43.8	35.2	39.6	40.9	48.3	1.17 (0.81-1.62)	1.21 (0.94-1.52)	1.35 (0.68-2.33)
Lead screening	34.8	37.1	31.7	36.5	31.2	35.9	1.10 (0.90-1.33)	1.24 (1.07-1.43)	1.24 (0.91-1.63)
Influenza	68.9	70.8	48.9	49.0	82.8	84.4	1.10 (0.99-1.19)	1.00 (0.94-1.08)	1.13 (0.97-1.30)
Rotavirus	30.7	42.9	21.0	31.3	32.2	44.9	1.70 (1.54-1.86)	1.72 (1.58-1.86)	1.72 (1.53-1.92)
	26.5	50.5	21.4	39.1	19.8	51.9	2.84 (2.57-3.11)	2.36 (2.17-2.55)	4.37 (3.84-4.93)

Abbreviations: ADHD, attention-deficit/hyperactivity disorder; DTP, diphtheria, tetanus toxoids, and pertussis vaccine; IPV, inactivated polio vaccine; MMR, measles, mumps, and rubella; NCH, Nationwide Children's Hospital; Td, tetanus-diphtheria toxoids; Tdap, tetanus, diphtheria, and pertussis.

^a Odds ratios measure the change in performance from 2010-2011 to 2012-2013 within each group of physicians.

literature.^{5,8,19-21} It is not clear how much the physician committee charged with design and implementation of the P4P program influenced program effectiveness, although it is our impression the effect was real.

Unexpectedly, in addition to the modest improvement noted from incentivized relative to nonmember practices, for immunizations, we found an even greater improvement among NCH physicians compared with incentivized physicians. This latter finding may reflect the different starting points of the measures. Nationwide Children's Hospital immunization measures (childhood immunization; diphtheria, tetanus toxoids, and pertussis vaccine; hepatitis A; and inactivated polio vaccine) with low baseline performance had the largest performance gains relative to the incentivized physicians. This finding is consistent with other studies that have found that those measures on which performance is weakest often have the most rapid improvement.⁶ The large performance improvements in the NCH group may also

have resulted from significant quality-improvement efforts at the hospital, including quality-improvement support, and changes in the electronic medical record providing point-of-care clinical decision support (S.G., oral communication, August 2015).

One of the Institute of Medicine's core recommendations for learning health care systems was that health care systems should evaluate their services through the use of sophisticated methods of observational data analysis, such as the differences-in-differences analysis used here.^{22,23} Large systems are often challenged by the expense and complexity of conducting randomized trials for reimbursement and contracting, so observational methods with simulated trials and other tools can provide valuable information. Going forward, learning health care systems should use these tools. A lesson from this study is that it is essential to make every effort to capture attributes of patients, physicians, practices, and communities that might influence outcomes.

Figure 1. Odds Ratios Representing the Interactions Between the Years' Effects and the Incentivized vs Nonmembers' Effects

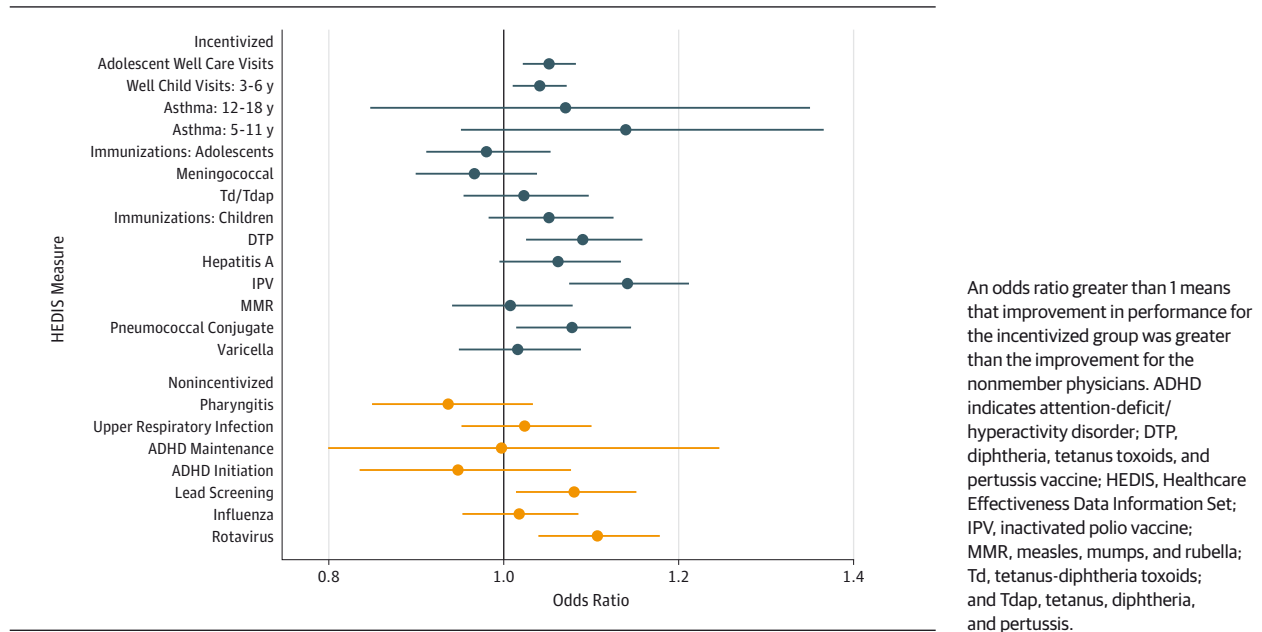
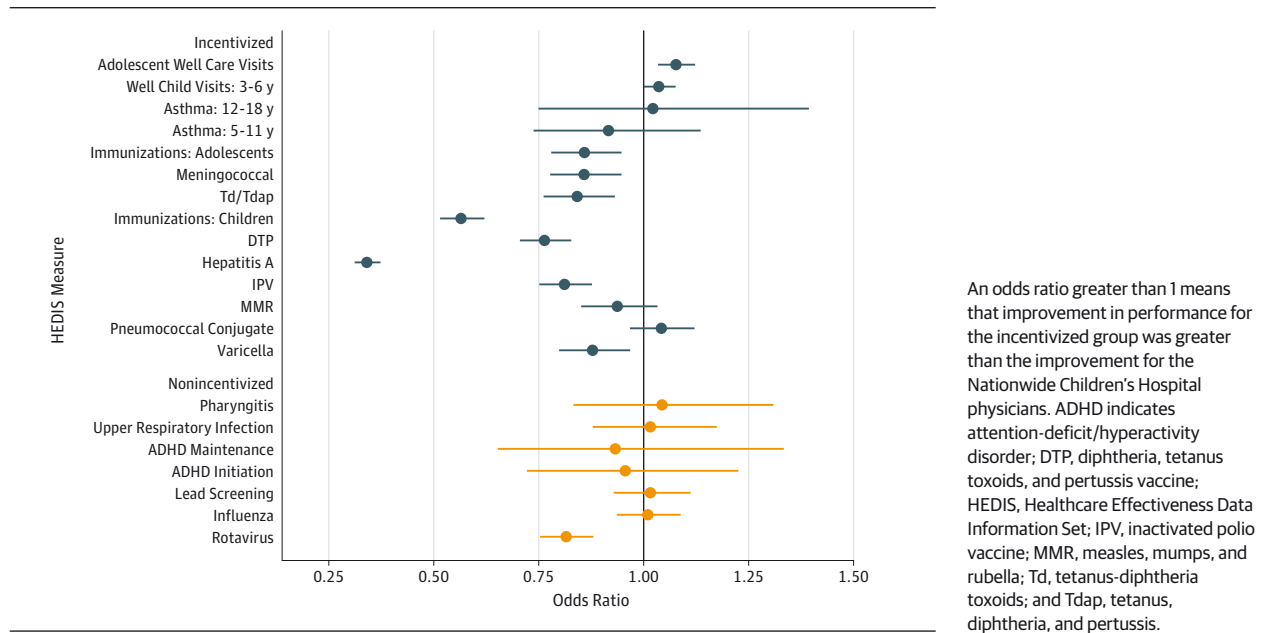


Figure 2. Odds Ratios Representing the Interactions Between the Years' Effects and the Incentivized vs Nationwide Children's Hospital Effects



Our results suggest that P4P can work in a pediatric ACO, but that it may not deliver the magnitude of improvement in quality that organizations seek. Future research should investigate whether P4P's effectiveness can be increased and explore other ways to motivate physician behavior change. Other investigators have speculated that the size of the incentives,^{16,19} the penetration of the specific insured population in the practice,²⁴ practice skill at transformation,^{25,26} and the duration of the incentive²⁷ are all factors that influence the likelihood that incentives can be effective.

The involvement of physician leadership in helping determine the incentives may have been useful in engaging physicians and increasing acceptance. The physician-led nature of ACOs should position these organizations to create programs of greater impact. At present, there is insufficient data about the structure of medical leadership in ACOs to draw conclusions about best practices.

This study had a few limitations. The institutional context of this study—a pediatric ACO contracting within a multipayer environment—limits the generalizability of the con-

clusions. As with any large geographically diverse organization, effective communication is a constant challenge, particularly with independent physicians having only a contractual relationship to the ACO. Some incentivized physicians were likely not conscious of the incentive owing to incomplete communication, limiting the program's measured impact. Effective communication strategies have been identified as key components of a successful program.¹⁹ In addition, the amount of the incentive payment (approximately \$40 per patient per year) was small. Other authors who studied larger incentives found larger effects.^{16,17} This program was implemented exclusively within a Medicaid population. Therefore, the conclusions may not apply to a commercial population.

Because the number of clinicians and patients in the nonmember incentive condition dwarfed those in the other conditions, it may have been easier to detect a difference in performance improvement when incentivized and nonmember physicians were compared vs when incentivized and NCH physicians were compared.

The most important limitation in our study was likely the preexisting differences among the different physician groups and, relatedly, our limited data on these differences. Unstudied factors that may have differed between the physician groups included the degree of electronic medical record use, the availability of support for quality-improvement efforts, and the organizational structure of their group practice. Although we did not have data on the training of physicians, the patients of nonmembers were older, suggesting that the proportion of family physicians was higher among the nonmembers. If so, this probably reflects a selection bias because PFK preferentially recruited pediatricians when the network was being formed. There was also an important difference in quality at baseline, with incentivized and NCH practices frequently performing at a higher level than the nonmembers. Relatedly, we also did not have in-

formation about practice memberships for a large number of nonmembers; therefore, we were unable to include that information in the statistical modeling. That likely led to an underestimation of the standard errors of the ORs and hence their confidence intervals as well.

Concern about preexisting differences between the physician groups was mitigated in part by the differences-in-differences design, in which each group's performance is compared with its own baseline. However, all physicians in this study—like other American physicians during this period—were exposed to messages promoting quality improvement. Therefore, it is possible that the P4P program (or in the case of the NCH physicians, the quality-improvement interventions at the hospital) was not the cause of the differences in performance improvement that we observed. We cannot rule out the possibility that preexisting differences between physicians made some more responsive to quality-improvement messaging. However, our view is that messaging alone would be unlikely to produce significant differences in differences in performance as it has been shown that P4P combined with public reporting is more effective than public reporting alone.²⁸

Conclusions

In pediatrics, as in other populations, P4P used alone had modest effects on improving quality measures in a pediatric Medicaid ACO. Combined with other quality-improvement activities and selection of high-quality physicians for networks, P4P may have an important role in quality improvement. Our study demonstrates that networks can use observational data analyses of their growing volume of claims, electronic health record, and data monitoring to gain insight on methods to improve the quality of care.

ARTICLE INFORMATION

Accepted for Publication: October 15, 2015.

Published Online: January 25, 2016.
doi:10.1001/jamapediatrics.2015.3809.

Author Contributions: Drs Gleeson and Gardner had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: All authors.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: All authors.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Gleeson, Gardner.

Administrative, technical, or material support: Gleeson, Kelleher.

Study supervision: Kelleher.

Conflict of Interest Disclosures: Dr Gleeson was a paid medical director (now president) of Partners for Kids, a paid vice president for community health at Nationwide Children's Hospital, and a practicing pediatrician in the Nationwide Children's Hospital physicians group studied. Dr Kelleher is an unpaid board member of Partners for Kids and vice president for health services research at the Nationwide Children's Research Institute, which is

owned by Nationwide Children's Hospital, which also controls Partners for Kids, the accountable care organization studied in this article. Dr Gardner was a professor at the Nationwide Children's Research Institute during the period these data were collected.

Additional Contributions: We gratefully acknowledge the following individuals for their unpaid contribution: Elaine Damo and Andy Wilhelm (Nationwide Children's Hospital) for their assistance in obtaining and structuring the data used for the analysis; Alyna Chien, MD (Boston Children's Hospital) for her early assistance in designing the study; and Dawn Lewis, David Whittridge, Eric Boamah, and John DeVore (Nationwide Children's Hospital) for their assistance in facilitating the necessary data infrastructure to allow for data analysis.

REFERENCES

1. Kuhmerker K, Hartman T. *Pay-for-Performance in State Medicaid Programs: A Survey of State Medicaid Directors and Programs*. New York, NY: The Commonwealth Fund; 2007:55.

2. Rosenthal MB, Landon BE, Normand SL, Frank RG, Epstein AM. Pay for performance in commercial HMOs. *N Engl J Med*. 2006;355(18):1895-1902.

3. Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of pay for performance on the quality of primary care in England. *N Engl J Med*. 2009;361(4):368-378.

4. Chien AT, Song Z, Chernew ME, et al. Two-year impact of the alternative quality contract on pediatric health care quality and spending. *Pediatrics*. 2014;133(1):96-104.

5. Eijkenaar F, Emmert M, Scheppach M, Schöffski O. Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy*. 2013;110(2-3):115-130.

6. Van Herck P, De Smedt D, Annemans L, Remmen R, Rosenthal MB, Sermeus W. Systematic review: effects, design choices, and context of pay-for-performance in health care. *BMC Health Serv Res*. 2010;10:247.

7. Eijkenaar F. Pay for performance in health care: an international overview of initiatives. *Med Care Res Rev*. 2012;69(3):251-276.

8. de Bruin SR, Baan CA, Struijs JN. Pay-for-performance in disease management:

a systematic review of the literature. *BMC Health Serv Res.* 2011;11:272.

9. Chien AT, Conti RM, Pollack HA.

A pediatric-focused review of the performance incentive literature. *Curr Opin Pediatr.* 2007;19(6):719-725.

10. Song Z, Rose S, Safran DG, Landon BE, Day MP, Cherner ME. Changes in health care spending and quality 4 years into global payment. *N Engl J Med.* 2014;371(18):1704-1714.

11. Chien AT, Colman MW, Ross LF. Qualitative insights into how pediatric pay-for-performance programs are being designed. *Acad Pediatr.* 2009;9(3):185-191.

12. Lemak CH, Cohen GR, Erb N. Engaging primary care physicians in quality improvement: lessons from a payer-provider partnership. *J Healthc Manag.* 2013;58(6):429-443; discussion 444-445.

13. Kelleher KJ, Cooper J, Deans K, et al. Cost saving and quality of care in a pediatric accountable care organization. *Pediatrics.* 2015;135(3):e582-e589.

14. National Committee for Quality Assurance. *Healthcare Effectiveness Data and Information Set: Technical Specifications for Health Plans.* Vol 2. Washington, DC: National Committee for Quality Assurance; 2013.

15. Long SK, Stockley K. The impacts of state health reform initiatives on adults in New York and Massachusetts. *Health Serv Res.* 2011;46(1, pt 2):365-387.

16. Chien AT, Li Z, Rosenthal MB. Improving timely childhood immunizations through pay for performance in Medicaid-managed care. *Health Serv Res.* 2010;45(6, pt 2):1934-1947.

17. Rosenthal MB, Li Z, Robertson AD, Milstein A. Impact of financial incentives for prenatal care on birth outcomes and spending. *Health Serv Res.* 2009;44(5, pt 1):1465-1479.

18. Bardach NS, Wang JJ, De Leon SF, et al. Effect of pay-for-performance incentives on quality of care in small practices with electronic health records: a randomized trial. *JAMA.* 2013;310(10):1051-1059.

19. Felt-Lisk S, Gimm G, Peterson S. Making pay-for-performance work in Medicaid. *Health Aff (Millwood).* 2007;26(4):w516-w527.

20. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med.* 2006;145(4):265-272.

21. 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(10):909-918.

22. Institute of Medicine. *Integrating Research and Practice: Health System Leaders Working Toward High-Value Care: Workshop Summary.* Washington DC: Institute of Medicine; 2015.

23. Gardner W, Kelleher K. A learning health care system for pediatrics. *JAMA Pediatr.* 2014;168(4):303-304.

24. Rosenthal MB, Frank RG. What is the empirical basis for paying for quality in health care? *Med Care Res Rev.* 2006;63(2):135-157.

25. Payne TH, Galvin M, Taplin SH, Austin B, Savarino J, Wagner EH. Practicing population-based care in an HMO: evaluation after 18 months. *HMO Pract.* 1995;9(3):101-106.

26. Mandel KE, Kotagal UR. Pay for performance alone cannot drive quality. *Arch Pediatr Adolesc Med.* 2007;161(7):650-655.

27. Langdown C, Peckham S. The use of financial incentives to help improve health outcomes: is the quality and outcomes framework fit for purpose? a systematic review. *J Public Health (Oxf).* 2014;36(2):251-258.

28. Mehrotra A, Damberg CL, Sorbero ME, Teleki SS. Pay for performance in the hospital setting: what is the state of the evidence? *Am J Med Qual.* 2009;24(1):19-28.