Preventive Medicine 55 (2012) S80–S85

Contents lists available at SciVerse ScienceDirect

Preventive Medicine

journal homepage: www.elsevier.com/locate/ypmed

Impact of a pay for performance program to improve diabetes care in the safety net

Alyna T. Chien [a,*], Diana Eastman [b], Zhonghe Li [b], Meredith B. Rosenthal [b]

[a] Division of General Pediatrics, Boston Children's Hospital and Harvard Medical School, 21 Autumn Street – Room 223, Boston, MA 02215, USA
[b] Department of Health Policy and Management, Harvard School of Public Health, 677 Huntington Avenue – Room 405, Boston MA 02115, USA

ARTICLE INFO

Keywords:
Pay for performance
Payment strategy
Healthcare quality
Diabetes care
Medicaid

ABSTRACT

Objective. To evaluate the impact of a "piece-rate" pay for performance (P4P) program aimed at improving diabetes care processes, outcomes and related healthcare utilization for patients enrolled in a not-for-profit Medicaid-focused managed care plan.

Methods. To evaluate Hudson Health Plan's P4P program in New York (2003–2007), we conducted: (1) a case-comparison difference-in-difference study using plan-level administrative data; (2) a patient-level claims data analysis; and (3) a cross-sectional survey.

Results. The case-comparison study found that diabetes care processes (e.g., HbA1c, lipid, and dilated eye exam rates) and outcomes (e.g., LDL-C<100 mg/dL) did not improve significantly over the study period. Claims analysis showed that younger adults had significantly increased odds (OR 3.50–3.56, p<0.001) of using emergency and hospital-based services and similarly decreased odds of receiving recommended care process (OR 0.22–0.36, p<0.01–0.001). Survey study indicated that practices lack fundamental quality improvement infrastructures and training.

Conclusions. Recent health legislation mandates the use of P4P incentives in government programs that disproportionately care for patients with lower socioeconomic or minority backgrounds (e.g., Medicaid, Veterans Health Administration, and Tricare). More research is needed in order to understand how to tailor P4P programs for vulnerable care settings.

Introduction

Pay-for-performance (P4P) programs—payment initiatives that reward providers for improving or delivering high quality healthcare—can improve a variety of healthcare processes and outcomes (Dudley et al., 2004; Petersen et al., 2006). Less is known, however, about how effectively these strategies work for practices located in lower socioeconomic (SE) settings, which disproportionately care for medically and socially complex patients, while often lacking the necessary quality improvement (QI) infrastructures or support (Chien et al., 2007a; Weinick et al., 2010; Young et al., 2010).

Evidence suggests that providers caring for lower SE populations are disadvantaged in traditional "all-or-nothing" P4P programs or may forego participation altogether because quality targets seem unattainable (Franks et al., 2003; Hong et al., 2010; Mehta et al., 2008; Rosenthal et al., 2005). While two recent studies show that P4P incentives improved childhood immunization rates in Medicaid-insured populations, both note that health plans provided significant administrative support (e.g., reminder calls, patient incentives) as part of their programs (Chien and Rosenthal, 2010; Felt-Lisk et al., 2007). One of these studies also found that rewarding improvement through "piece-rate" bonuses rather than traditional achievement targets is beneficial in the Medicaid setting (Chien and Rosenthal, 2010). A pre/post analysis of physician-level P4P incentives in a federally qualified health center (FQHC) network, found a significant increase in patients receiving the recommended number of Hba1c tests post-implementation without patient health outcomes moving in tandem; authors hypothesized that the limited impact on outcomes could be due to lack of QI support (Coleman et al., 2007).

This paper evaluates a robust P4P program aimed at improving diabetes care quality and outcomes among enrollees within a Medicaid managed care health plan. This study is important because it evaluates a P4P program that not only followed "best practices" based on current evidence, but also addressed several prior concerns noted about the use of P4P incentives, particularly for providers located in lower SE settings.

First, the sponsoring health plan focused on a health condition (diabetes) with clinically valid measures and an evidence base amenable to P4P incentives (Coleman et al., 2007; Levin-Scherz et al., 2006; Rosenthal

* Corresponding author at: Children's Hospital Boston, 21 Autumn Street – Room 223, Boston, MA 02215, USA.
E-mail addresses: Alyna.chien@childrens.harvard.edu (A.T. Chien), Deastman@hsph.harvard.edu (D. Eastman), ZLI@hsph.harvard.edu (Z. Li), mrosenth@hsph.harvard.edu (M.B. Rosenthal).

0091-7435/$ – see front matter © 2012 Published by Elsevier Inc.
doi:10.1016/j.ypmed.2012.05.004

et al., 2005). The program also rewarded both process and outcome measures, addressing the criticism that P4P initiatives inadequately link payment with desired health outcomes (Averill et al., 2011). Second, the bonus amount was set well above typical levels and was substantial compared to office visit fees for a Medicaid population (Baker and Carter, 2005) and comparable to the size of the bonuses used in other P4P programs that were found to be effective in increasing childhood immunization rates the Medicaid setting (Chien, Felt-Lisk). Third, the health plan delivered P4P rewards in a "piece-rate" (as opposed to a "tournament" or all-or-nothing achievement) basis, which encourages continuous improvement independent of baseline performance and minimizes provider incentives to avoid high-risk or non-adherent patients (Hofer et al., 1999). Fourth, recognizing that smaller or less resourced practices may not have the infrastructure or personnel to create registries, identify patients in need of services, or send patient reminders, the health plan provided substantial administrative support as part of the program (Chien and Rosenthal, 2010; Felt-Lisk et al., 2007). Fifth, to improve upon claims-based measurement, the health plan allowed physicians to supplement billing information with chart-based data.

Thus, the goal of this paper is to evaluate the effectiveness of a Medicaid-sponsored P4P program designed with practices serving lower SE populations in mind. A subsidiary aim is to evaluate how the impact of P4P incentives may vary by gender, racial/ethnic group, medical complexity (e.g., number of comorbidities), and age because there is concern that P4P incentives may cause providers to improve care disproportionately for those groups whose behavior makes reaping financial rewards easier (Chien et al., 2007a; Millett et al., 2009) and little is known about care for younger adults. A third aim is to use survey data to provide insight into the strategies that participating practices used in response to these P4P incentives.

Methods

Study setting

Hudson Health Plan (Hudson) is a not-for-profit Medicaid-focused managed care health plan serving the Hudson Valley region of New York (Westchester, Rockland, Orange, Dutchess, Sullivan and Ulster counties). At the time of this study, Hudson contracted with a provider network that included 118 primary care practices (620 physicians) to care for 100,000 enrollees—approximately 7% of whom are adults with diabetes. Of the eligible practices, about 65% were solo or small practices and 35% were medium (4–9 full-time physicians) to large-sized (10 or more full-time physicians) practices. Smaller practices tended to be privately owned whereas larger ones tended to be hospital-based or FQHCs. More than half of Hudson's members identify as Hispanic, with the remaining population divided relatively evenly between Blacks, Whites and those of other race/ethnicity.

P4P program features

Late in 2003 Hudson piloted a diabetes improvement initiative (Operation Diabetes) in 6 of 118 participating practices. This program targeted members who were missing one or more of the following clinical tests: HbA1c, LDL cholesterol, dilated retinal exam, and microalbuminuria. At that time, providers were offered $100 for each patient completing all the missing care processes. Hudson's evaluation of the program led to a revised program (heightened communication supports, larger P4P bonuses) and launched that program in August 2004. In the beginning of 2005, the program was revised a second time such that incentive amounts in 2005 P4P incentive were 3 times that offered in 2003 and more than twice the 2004 bonus. Hudson also added diabetes quality measures that reflected desired health outcomes (e.g., not just whether annual HbA1c and LDL tests were delivered, but whether HbA1c and LDL levels were lowered or at goal levels). The diabetes quality measures used in the P4P program were also aligned with state and national quality efforts and based on those used in New York State's Quality Assessment and Reporting Requirements (QARR) program and in the Health Plan Employer Data and Information Set (HEDIS) (NYS DOH, 2005). The measures and dollar amount tied to each measure are provided in Appendix A.

Hudson designed the P4P program's administrative aspects in conjunction with participating practices. Each March, Hudson generated patient reports identifying adult enrollees with diabetes and any care elements that were missing or below national goals. At the end of the calendar year, the plan sent practices an updated snapshot and gave providers the opportunity to rectify report errors by supplying supplemental chart data within 90 days. Hudson representatives hand-delivered final reports and payments to physician practices and were available to discuss results and identify opportunities for improvement; additional follow-up and coaching occurred at 2, 4, and 6 weeks later.

Study designs, data, and analytic approaches

We conducted two quasi-experimental studies and one cross-sectional study to evaluate the effectiveness of Hudson's P4P program. First, we conducted a case-comparison difference-in-difference analysis to compare Hudson's diabetes care processes and outcomes to those of non-Hudson Medicaid-focused health plans within New York. For this aspect of the evaluation, we used health plan-level QARR data, assembled by New York State's Department of Health (NYS DOH, 2004–2008a) from all certified managed care health plans operating within the state. Quality Assurance Reporting Requirement (QARR) measures are rotated so none of the diabetes process measures were reported in 2005, the HbA1c<9 measure was not reported in 2006, and the BP<130/80 measure was not reporting 2003 or 2004. These data are subject to a full audit of administrative claims records and patient charts by a National Committee for Quality Assurance (NCQA)-certified auditor. We compared Hudson to other NY Medicaid plans that operate outside of the counties in which Hudson enrolls its members (N = 11) to minimize the likelihood that our comparison group would be contaminated by spillover effects from Hudson's P4P program. We also systematically examined the two largest publicly available repositories of health plan QI efforts and determined that none of the plans in our comparison group reported using P4P strategies to incentivize adult diabetes care measures (Leapfrog Group, 2009; NYS DOH, 2004–2008b). We assessed performance based on rates of annual HbA1c, lipid, and eye exam testing, and on levels of lipid control. Because we had only plan-level data, we tested for differences in performance on QARR performance between Hudson and comparator plans using a plan-level regression that accounts for the panel nature of the data (generalized estimating equations with binomial family and logistic link function). We also examined an alternative model that adjusted for baseline performance (2003) to account for the possibility that improvement opportunities might be different for lower-performing plans than higher-performing plans.

Second, we estimated and interrupted time series model using Hudson's health plan enrollment and claims data. Specifically, our quasi-experimental design looked for a break in trend after the P4P program was implemented for all of our outcome variables assuming an underlying quadratic trend in the data. We restricted our analysis to adults who were continuously enrolled in Hudson for at least 6 months. The primary dependent variables in this analysis were the same as those in the QARR-based study with respect to annual HbA1C testing, LDL cholesterol testing, and retinal exams (these were the ones most reliably captured by billing data). Although diabetes-related utilization was not incentivized by the P4P program, we also examined it to provide additional insight into the mechanism that may engender changes in diabetes care quality (e.g., whether the intervention period was associated with increases in the frequency of office visits). We examined emergency department and inpatient care for diabetes on the theory that reductions might occur as a result of improved management of diabetic patients. The unit-of-analysis for quality measures, which are annual by construction, was the patient year. The unit-of-analysis for utilization measures was the patient–month. In these models, we included age, sex, a non-nested race/ethnicity variable, whether patients had 2 or comorbidities according to Elixhauser's well-established method based on International Classification of Disease, Ninth edition (ICD-9)-codes (Elixhauser et al., 1998), type of practice, and primary language spoken. We used patient-level logistic regression models that account for clustering within provider as well as serial correlation in the data. The statistical significance of differential effects by age category, race/ethnicity and comorbidities were tested through simulation and bootstrapping of standard errors. Interaction effects were examined for race/ethnicity, age, type of clinic, and presence of comorbidities.