

© Health Research and Educational Trust
DOI: 10.1111/1475-6773.12035
RESEARCH ARTICLE

The Effect of Pay-for-Performance in Nursing Homes: Evidence from State Medicaid Programs

Rachel M. Werner, R. Tamara Konetzka, and Daniel Polsky

Objective. Pay-for-performance (P4P) is commonly used to improve health care quality in the United States and is expected to be frequently implemented under the Affordable Care Act. However, evidence supporting its use is mixed with few large-scale, rigorous evaluations of P4P. This study tests the effect of P4P on quality of care in a large-scale setting—the implementation of P4P for nursing homes by state Medicaid agencies.

Data Sources/Study Setting. 2001–2009 nursing home Minimum Data Set and Online Survey, Certification, and Reporting (OSCAR) datasets.

Study Design. Between 2001 and 2009, eight state Medicaid agencies adopted P4P programs in nursing homes. We use a difference-in-differences approach to test for changes in nursing home quality under P4P, taking advantage of the variation in timing of implementation across these eight states and using nursing homes in the 42 non-P4P states plus Washington, DC as contemporaneous controls.

Principal Findings. Quality improvement under P4P was inconsistent. While three clinical quality measures (the percent of residents being physically restrained, in moderate to severe pain, and developed pressure sores) improved with the implementation of P4P in states with P4P compared with states without P4P, other targeted quality measures either did not change or worsened. Of the two structural measures of quality that were tied to payment (total number of deficiencies and nurse staffing) deficiency rates worsened slightly under P4P while staffing levels did not change.

Conclusions. Medicaid-based P4P in nursing homes did not result in consistent improvements in nursing home quality. Expectations for improvement in nursing home care under P4P should be tempered.

Key Words. Quality of care, pay-for-performance, nursing home quality, long-term care

The use of pay-for-performance (P4P) to improve health care quality has become commonplace in the United States. P4P is based on the principle that provider payment should be determined by quality of care rather than

intensity of care. Accordingly, P4P provides financial rewards to providers who perform well on accepted measures of quality and shifts emphasis toward the quality rather than the quantity of care (Robinson 2001). As the principle of aligning payment to improve quality of care is difficult to dispute, P4P programs have been implemented in many health care settings (Centers for Medicare and Medicaid Services 2003; Centers for Medicare and Medicaid 2005, 2007; Rosenthal, Landon et al. 2006) and under the Affordable Care Act it is expected that P4P will become commonplace.

Despite the proliferation of P4P programs, there is currently little evidence to support their use. Two comprehensive reviews of early evaluations of P4P found mixed evidence in support of the hypothesis that P4P improves quality of care (Petersen et al. 2006; Rosenthal and Frank 2006). There are limitations to prior studies, however. Many of the studies identified in these reviews lacked a rigorous empirical design and most evaluated relatively small-scale programs that enrolled a limited number of providers and patients. Whether large-scale P4P programs are effective in improving quality remains a source of debate. In this study, we evaluate large state-run Medicaid P4P programs in nursing homes.

BACKGROUND

Over 1.5 million people reside in U.S. nursing homes at a cost of over \$120 billion per year (Kaiser Family Foundation 2007). Nursing homes generally serve two populations—long-stay residents (who typically receive non-skilled care such as assistance with activities of daily living) and postacute residents (who receive rehabilitative care following an acute-care hospitalization). While long-stay care is typically aimed at chronically ill individuals who spend the remainder of their lives in nursing homes (2 years on average), post-acute care typically requires a shorter stay (25 days on average) and is aimed at a healthy discharge to the community.

Address correspondence to Rachel M. Werner, M.D., Ph.D., 1230 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104; e-mail: rwerner@upenn.edu. Rachel M. Werner is with the Center for Health Equity Research and Promotion, Philadelphia VAMC and the Division of General Internal Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA. R. Tamara Konetzka, Ph.D., is with the Department of Health Studies, University of Chicago, Chicago, IL. Daniel Polsky, Ph.D., is with the Division of General Internal Medicine, University of Pennsylvania School of Medicine, Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA.

Despite this frequent use and high cost of nursing home care, quality of care in nursing homes has long presented a policy challenge (Institute of Medicine 1986). Major regulatory policies aimed at improving nursing home quality were implemented in 1987 under the Nursing Home Reform Act or the Omnibus Budget Reconciliation Act (OBRA), a congressional act that mandated extensive regulatory controls. As a result of OBRA, each Medicare- or Medicaid-certified nursing home is inspected at least once every 15 months and is required to submit a comprehensive assessment of each chronic-care resident at least once per quarter. While researchers found that OBRA led to improved quality (Kane et al. 1993; Shorr et al. 1994; Castle et al. 1996; Fries et al. 1997; Mor et al. 1997; Snowden and Roy-Byrne 1998), a follow-up report by the Institute of Medicine in 2000 concluded that significant problems remain (Wunderlich and Kohler 2000).

With regulation failing to fully reform nursing home quality, efforts have recently turned toward market-based reforms designed to improve quality of care. One prominent example of this is P4P. In an early experiment with P4P in nursing homes in 1980, 36 nursing homes in San Diego were randomized to receive financial incentives tied to the patient outcome of improved functional or health status while in the nursing home. An evaluation of this effort found that residents of the experimental nursing homes were more likely to go home or to a lower level nursing home, and less likely to be hospitalized or to die than were people in the control nursing homes (Norton 1992). Despite this success, P4P was not widely implemented in nursing homes until the last decade.

Since 2002, a number of state Medicaid agencies have implemented P4P programs based on the quality of chronic care delivered using financial incentives tied to Medicaid payment (Kane et al. 2007; Werner et al. 2010). In addition, the Centers for Medicare and Medicaid Services (CMS) has implemented a nursing home P4P demonstration project using financial incentives tied to Medicare payment (Abt Associates Inc. 2006). Despite the recent relative proliferation of P4P programs in nursing homes, it remains unknown how these efforts affect quality of care. Our objective is to test for changes in nursing home quality under Medicaid-sponsored P4P programs in nursing homes.

SETTING

Between 2002 and 2009, eight states adopted Medicaid-sponsored P4P programs in nursing homes (see Table 1), all of which primarily targeted

quality of care for long-stay (or chronic care) residents. The details of these programs have been previously described (Werner et al. 2010). Briefly, most states use a payment model based on a point system that is translated into per diem add-ons. For each measure included in the payment model, each nursing home is evaluated and earns points based on either its ranking compared with other nursing homes in the state or, in one state, whether it has achieved a target level of performance. The earned points are summed across all measures and translated into a per diem add-on for all Medicaid resident days, where nursing homes with more points receive higher add-ons. The total possible bonus amount varied across states from an add-on valued at about half a percent of the per diem rate to over 5 percent with bonuses in most states ranging between 3 and 4 percent of the per diem rates. State Medicaid agencies spent between 0.4 and 1.8 percent of their total Medicaid nursing home budgets on P4P programs in 2008, totaling between 1 and 18 million dollars (Werner et al. 2010).

Most states base payment in part on rates of regulatory deficiencies and staffing ratios. In some states, facilities are only eligible for bonuses if they achieve a predetermined performance level on regulatory compliance and staffing ratios. In others, nursing homes earn points based on their regulatory compliance and staffing ratios. In addition, four of the eight states also base their payment on clinical quality measures.

These programs were implemented at various times across states between 2002 and 2009 (see Table 1). All states continued their P4P programs through 2009 with the exception of Minnesota, where the P4P program ran for 2 years, from October 2006 to September 2008.

METHODS

We use a difference-in-differences approach to test for changes in nursing home quality with the implementation of P4P in eight states, taking advantage of the variation in timing of implementation across these states and using nursing homes in the 42 non-P4P states plus Washington, DC as contemporaneous controls for nursing home quality.

Study Sample, Study Time Period, and Data

The study sample focuses on long-stay (or chronic-care) nursing home residents, the focus of the P4P programs we evaluate over the period from 2001 to

2009. We identify long-stay residents in the minimum data set (MDS), which contains detailed clinical data collected at regular intervals (usually quarterly) for every resident in a Medicare- and/or Medicaid-certified nursing home. The MDS includes data on residents' health, physical functioning, mental status, and psycho-social well-being and are used by nursing homes to assess the needs and develop a care plan for each resident. The MDS is also used by state Medicaid agencies to measure clinical quality and determine P4P incentive payment for clinical quality of care in the four states that include clinical quality measures in their incentives.

Because MDS includes data on both long-stay and short-stay nursing home residents, we identify long-stay residents as those having at least one quarterly or annual assessment in addition to an admission (or prior quarterly/annual) assessment. This ensures that the resident has been in the nursing home for at least 90 days, the cutoff we use to distinguish short-stay from long-stay residents.

Each long-stay resident is assessed at least quarterly in the MDS, including on admission, annually, quarterly, and for a significant status change. To avoid overweighting sicker residents who may have more frequent assessments, we limit each resident to only one assessment per quarter in our final dataset. Following conventions set out by the CMS for measuring nursing home quality (Nursing Home Quality Initiative 2004), if a resident has more than one assessment per quarter, we choose the most recent assessment in that quarter. Finally, also following CMS conventions, we do not include admission assessments, as patient outcomes on admission cannot be attributed to the admitting nursing home's quality of care.

We test for changes under P4P over two time periods—in the 1 year after P4P was implemented in each P4P state and over the 2 years post-P4P implementation. We compare the post-P4P period to the 1 year prior to P4P implementation in each P4P state. For comparison states we include the entire time period of 2001–2009. Thus, each P4P state is compared with non-P4P states only over the 2 or 3 years surrounding P4P implementation in that state.

We use data from the MDS to measure resident-level performance metrics of clinical outcomes and to risk adjust these outcomes. We supplement resident-level data from MDS with facility-level data from the Online Survey, Certification, and Reporting (OSCAR) dataset, which is collected by state surveyors at all certified nursing homes at least once every 15 months. We use the OSCAR data to measure facility-level performance metrics, including deficiency citations and staffing ratios as well as time-varying facility characteristics.

Table 2: The Facility-Level Mean (and Standard Deviation) of Nursing Home Quality Measures Included in P4P Programs and as Outcome Variables in Analyses in States with and without P4P Programs

	<i>States with P4P</i>	<i>States without P4P</i>
% Residents who		
had bladder catheter inserted	6.1 (4.3)	6.2 (4.3)
were physically restrained	8.9 (8.1)	10.5 (10.1)
had moderate to severe pain	13.7 (9.3)	10.7 (7.6)
had falls	10.4 (3.7)	8.9 (3.8)
developed pressure sores	12.3 (6.7)	14.3 (6.8)
had unexplained weight loss	9.1 (3.8)	9.7 (4.1)
Regulatory deficiencies		
Total number	5.6 (5.2)	6.7 (6.0)
Number of immediate jeopardy	0.04 (0.36)	0.06 (0.50)
Staffing ratios		
Total staff hours per resident day	3.1 (5.5)	4.3 (16.3)
RN + LPN hours per resident day	1.1 (2.4)	1.7 (10.0)

Note. All quality measures are summarized prior to P4P implementation, in 2001.

Dependent Variables: Nursing Home Quality

Our dependent variables are a set of performance metrics used by states to determine P4P payments. First, we include a binary, resident-level indicator of clinical outcomes, following CMS’s technical specifications on the construction of these measures (Morris et al. 2003; Nursing Home Quality Initiative 2004). A list of the specific clinical outcomes included in state P4P programs is included in Table 1. We also include facility-level regulatory deficiencies, measured in two ways: the total number of deficiencies for a nursing home in a given year and the number of immediate jeopardy deficiencies. Finally, we include facility-level staffing ratios, measured as staff hours per resident day for two groups of staff—all direct-care staff and skilled staff (registered nurse plus licensed practical nurse). We follow standard procedures to calculate staffing ratios, assuming that each full-time equivalent staff member works 70 hours in a 2-week period and dividing the staffing hours per day by the number of residents in the facility (Abt Associates Inc. 2001). The mean values of these dependent variables are summarized in Table 2.

Covariates

For all resident-level analyses of changes in patient outcomes under P4P, we include resident-level characteristics as covariates, defined by CMS technical

specifications for each quality measure (Morris et al. 2003; Nursing Home Quality Initiative 2004). In addition, because these measures are minimally risk adjusted (which may adversely affect nursing homes caring for more severely ill residents), we also include age, gender, and race as well as additional outcome-specific clinical characteristics defined as risk adjusters in prior work (Berlowitz et al. 2001; Minnesota Department of Human Services 2007; Mukamel et al. 2008; Li et al. 2009) as covariates in all analyses. A list of resident-level covariates is included in Appendix SA1.

For all facility-level analyses of changes in resident- and facility-level measures, we include the following facility-level characteristics as covariates in all analyses: a facility's percent of residents covered by Medicare, percent of residents covered by Medicaid, ownership; whether the facility is hospital-based, part of a chain; and its total number of beds. For these facility-level analyses we also included facility-level summaries of resident characteristics, including each facility's mean age, percent female, percent in each racial and ethnic group, and mean Cognitive Performance Scale (Morris et al. 1994), ADL scale (Morris et al. 1999), and Clinically Complex Scale (Kidder et al. 2002).

Empirical Specifications

For changes in clinical outcomes, we estimate the following resident-level specification using a linear probability model:

$$\text{Quality}_{i,j,t} = \alpha \text{P4P}_{i,j,t} + \beta X_{i,j,t} + \gamma X_{j,t} + \varphi_j + \tau_t + \varepsilon_{i,j,t}$$

where i indexes residents, j indexes nursing homes, and t indexes time quarter. We estimate quality (clinical outcomes in this case) as a function of a P4P variable (which equals 1 if the resident is in a nursing home after P4P is implemented and zero otherwise), a vector of resident-level covariates, a vector of facility-level covariates, facility fixed effects, quarterly time fixed effects, and a mean zero random error component. The P4P indicator variable, in combination with the facility and time fixed effects, gives the difference-in-difference estimate of the effect of P4P on the outcome of interest. Thus, α represents the within-facility change in quality in P4P states compared with non-P4P states, after P4P was implemented compared to before.

To measure the 1-year effects of P4P, we define P4P as equal to zero in the 1 year prior to P4P in P4P states, equal to 1 in the 1 year after P4P implementation, and missing otherwise in P4P states. To measure the 2-year effect, in each P4P state P4P equals 1 in the 2 years after P4P implementation, zero

before, and missing otherwise. In non-P4P states, the P4P variable is zero and time invariant.

We estimate this equation for the six clinical quality measures defined in Table 1. In addition to defining P4P as being time specific according to whether we are interested in 1- or 2-year effects, we also define the P4P variable to be measure specific as not all states included each clinical outcome in their P4P programs (and half of the states did not include any measure of clinical outcomes). Thus, for example, for the measure of the percent of residents who had a bladder catheter inserted, P4P equals 1 if a resident is in Oklahoma after 7/2007 and in Minnesota between 10/2006 and 9/2008 (for the 2-year effect) and equals zero for all other states and in the 1 year prior to P4P implementation in Oklahoma and Minnesota.

For changes in number of deficiencies and staffing ratios, we estimate facility-level models (using a negative binomial model for the count of deficiencies and a linear model for staffing ratios). These regressions follow the same form as that described in the above equation but include resident-level covariates as means at the facility level.

In addition to testing for a P4P effect across all P4P states combined, to account for possible variation in P4P programs across states we also test state by state for a P4P effect. To do this, for each of our outcomes we run the above regression but we include one P4P state at a time, comparing each state to the 42 non-P4P states in our sample plus Washington, DC.

Robust standard errors were used to account for nonindependence of observations from the same facility in all regressions (Huber 1967; White 1980).

We test the robustness of our results in several ways. First, we test the same empirical model in a dataset composed of a balanced panel of nursing homes over the study period (i.e., excluding nursing homes that enter or exit the market during our study period). We do this to confirm that our results are not driven by entry or exit. Second, we test the sensitivity of our results to our definition of non-P4P or control states by excluding states from the control group that have P4P but where the P4P does not target the quality indicator of interest. For example, in this robustness check we exclude Colorado, Georgia, Iowa, Kansas, Ohio, and Utah from the control group when testing the effect of P4P on use of bladder catheters (whereas in our main specification these states are included in the control group). Although the large number of observations in the non-P4P states likely overwhelms any differing quality changes in the non-P4P states, it is possible that states with P4P targeting some areas but not others experience positive or negative spillovers to nontargeted

quality, making them a potentially inappropriate control. Third, because regulatory deficiencies and staffing are used as a minimum standard to receive incentives in some states, whereas other states tie payment directly to performance on deficiencies and staffing levels, we separately tested whether the effect of P4P in states with direct financial incentives for these two performance measures by removing states with indirect incentives (or minimum standards) from the analyses. We thus excluded Colorado, Georgia, and Utah from regressions of changes in total deficiencies and excluded Georgia from regressions of changes in staffing ratios. Finally, we tested whether our results are sensitive to our choice of all 42 states plus Washington, DC as a comparison group. To do this, we chose one neighboring control state for each P4P state. We selected the neighboring state that most closely matched the P4P state on their average levels and change in study outcomes during the pre-P4P period.

RESULTS

A total of 17,579 nursing homes were included in the study, of which 3,513 (20 percent) were located in a state with P4P (Table 3). These nursing homes cared for a total of 5,681,244 residents during the study period, of which 950,173 (16.7 percent) were in states that implemented P4P and accounted for 44,174,667 observations in our data in total (and 7,571,737 or 17 percent in states that implemented P4P). Nursing homes in states that implemented P4P programs were similar to those in states without P4P with respect to most facility characteristics and patient characteristics (Table 3). However, facilities in non-P4P states had a higher portion of private-pay residents (and hence a lower percentage of Medicare and Medicaid residents). In addition, a higher proportion of residents in P4P states was black and Hispanic compared with non-P4P states. However, residents were similar between P4P and non-P4P states with respect to measures of disability and clinical complexity.

Changes in quality of care under P4P 1 and 2 years after P4P implementation are summarized in Table 4. After 1 year, three of the targeted resident-level outcomes improved in P4P states compared with non-P4P states, controlling for secular trends. Pay-for-performance was associated with a decline in the percent of residents who were physically restrained of 0.5 percentage points (on a base of 9 percent); a decline in the percent of residents in moderate to severe pain of 0.5 percentage points (on a base of 14 percent); and a decline in the development of pressure sores of 0.3 percentage points (on a

Table 3: Characteristics of Nursing Homes and Nursing Home Residents in P4P and Non-P4P Facilities

	<i>States with P4P</i>	<i>States without P4P</i>
Nursing home characteristics		
Number of nursing homes	3,513	14,066
Percent Medicaid, mean (SD)	63.6 (22.0)	57.8 (20.0)
Percent Medicare, mean (SD)	13.1 (13.6)	10.5 (11.5)
Ownership, %		
Government	5.8	5.4
Not for profit	25.7	31.0
For profit	68.5	63.6
Hospital-based, %	6.2	6.5
Chain, %	54.0	56.3
Total beds, mean (SD)	113.6 (69.3)	91.3 (53.8)
Resident characteristics		
Number of residents	950,173	4,731,071
Number of assessments	7,571,737	36,602,930
Female, %	71.1	70.9
Race, %		
White	80.9	88.8
Black	13.2	9.6
Hispanic	4.0	0.8
Other	1.9	0.8
Age, mean (SD)	80.4 (13.5)	80.9 (13.2)
Cognitive performance scale, mean (SD)	2.9 (1.8)	2.9 (1.6)
Activities of daily living scale, mean (SD)	11.2 (4.7)	11.5 (4.9)
Clinically complex scale, mean (SD)	0.6 (1.0)	0.7 (1.0)

base of 12 percent). However, two clinical quality measures that were targeted by P4P worsened slightly a year after the implementation of P4P: the percent of residents who had a bladder catheter inserted and had unexplained weight loss. These quality declines were small in size (about 0.2 percentage points). There was no statistically significant change in the percent of residents who developed pressure sores with the implementation of P4P. Two years after P4P was implemented, the changes in resident outcomes were in the same direction and similar in magnitude.

Examining facility-level quality measures with the implementation of P4P, the total number of nursing home deficiencies increased under P4P (1 year after P4P implementation the incidence rate of total deficiencies increased by 1.12 and of immediate jeopardy deficiencies by 1.97). Changes in total staff hours per resident day and skilled staff hours per resident day were

Table 4: Results from Regressions Showing Effect of P4P Implementation on Nursing Home Quality

	Resident-Level Outcomes				Facility-Level Outcomes					
	Catheter Inserted	Physically Restrained	Pain	Falls	Pressure Sores	Weight Loss	Total Number of Deficiencies	Immediate Jeopardy Deficiencies	Total Staff HPRD	RN + LPN HPRD
One year following P4P implementation										
P4P	0.0016* (0.0008)	-0.0050*** (0.0012)	-0.0050*** (0.0017)	0.0019 (0.0014)	-0.0032** (0.0016)	0.0026* (0.0015)	0.1178*** (0.0094)	0.6794*** (0.1119)	-0.0148 (0.0161)	-0.0029 (0.0085)
Constant	0.153*** (0.0022)	0.0274** (0.0112)	0.175*** (0.0036)	0.0575*** (0.0013)	0.146*** (0.0027)	0.0684*** (0.0017)	1.3948*** (0.0350)	-0.1698 (0.2589)	3.170*** (0.163)	1.359*** (0.107)
Observations	34,809,913	35,691,015	35,545,537	30,037,910	20,704,519	34,205,962	447,016	73,665	437,288	437,288
Number of nursing homes	14,511	15,257	14,925	14,462	15,122	14,687	16,265	2,452	16,661	16,661
Two years following P4P implementation										
P4P	0.0026*** (0.0010)	-0.0078*** (0.0013)	-0.0061*** (0.0018)	0.0024* (0.0013)	-0.0033** (0.0016)	0.0026* (0.0014)	0.131*** (0.0080)	0.9592*** (0.0874)	-0.0066 (0.0150)	-0.0059 (0.0079)
Constant	0.153*** (0.0021)	0.0277** (0.0111)	0.175*** (0.0036)	0.0574*** (0.0013)	0.146*** (0.0027)	0.0444*** (0.0132)	1.378*** (0.0347)	-0.2021*** (0.2570)	3.166*** (0.161)	1.352*** (0.105)
Observations	34,975,704	35,982,151	35,771,523	30,174,254	20,880,951	34,364,577	456,916	74,708	447,809	447,809
Number of nursing homes	14,514	15,265	14,929	14,465	15,132	14,691	16,351	2,510	16,721	16,721
Resident covariates	X	X	X	X	X	X	X			

continued

Table 4. Continued

	Resident-Level Outcomes				Facility-Level Outcomes					
	Catheter Inserted	Physically Restrained	Pain	Falls	Pressure Sores	Weight Loss	Total Number of Deficiencies	Immediate Jeopardy Deficiencies	Total Staff HPRD	RN + LPN HPRD
Facility covariates	X	X	X	X	X	X	X	X	X	X
Facility fixed effects	X	X	X	X	X	X	X	X	X	X
Time fixed effects	X	X	X	X	X	X	X	X	X	X

Note. The coefficient on the P4P variable indicates the within-facility difference in outcomes for P4P states compared with non-P4P states after P4P was implemented compared to before. All regressions used a linear model except when the outcome was number of deficiencies for which a negative binomial model was used. Robust standard errors are in parentheses.

*** $p < .01$, ** $p < .05$, * $p < .1$.

HPRD, hours per resident day; LPN, licensed practical nurse; P4P, Medicaid nursing home pay for performance; RN, registered nurse.

small and not statistically different from zero, 1 or 2 years after P4P was implemented.

When examining the effect of P4P in a state-by-state analysis (i.e., each P4P state compared with all control states), we find that the effect of P4P was variable across states (Table 5). While P4P had a consistent effect on quality in Georgia (improving the three targeted resident-level outcomes and decreasing the number of deficiencies), the effect of P4P was inconsistent in the remaining seven states.

Our findings did not change substantially using a balanced panel of nursing homes, using a stricter definition of being a non-P4P state, or examining changes under P4P only in states with direct incentives for some measures rather than minimum standards (see Appendix SA3). When using one neighboring state as a control state for each P4P state, the results changed for some quality measures in some states, but these changes were unpredictable and did not change the overall finding of no consistent effect of P4P on nursing home quality (see Appendix SA4).

DISCUSSION

Although there has been significant hope that P4P can be a valuable tool to improve health care quality, accumulating evidence has found that P4P often fails to achieve this goal. However, few prior studies have used a rigorous research design and evaluated large-scale P4P programs. This study evaluates large nursing home P4P programs implemented by state Medicaid agencies using a differences-in-differences design, comparing nursing home quality before and after the implementation of P4P and controlling for secular trends in quality using states that did not have P4P in nursing homes.

Although P4P is increasingly being used to improve quality of care in nursing homes, we find little evidence to support its use. Over the period 2002–2009, P4P was implemented by eight state Medicaid programs. Four states tied incentive payments to performance on clinical quality measures, seven states tied incentives to deficiency rates, and six states tied incentives to staffing ratios. We find that although quality improved for three of the nine measures we examined (use of physical restraints, pain control, and pressure sores), nursing home quality did not consistently improve in any of the states that implemented P4P.

These Medicaid-sponsored P4P programs may have been minimally effective because the incentives targeted Medicaid patients and, in particular,

Table 5: State-by-State Regressions Showing Effect of P4P Implementation on Nursing Home Quality in Each State (in the One Year Following P4P Implementation)

	Resident-Level Outcomes					Facility-Level Outcomes				
	Catheter Inserted	Physically Restrained	Pain	Falls	Pressure Sores	Weight Loss	Total Number of Deficiencies	Immediate Jeopardy Deficiencies	Total Staff HPRD	RN + LPN HPRD
Colorado	-0.0031 (0.0030)	-0.0035 (0.0045)	-0.0030 (0.0021)	-0.0007 (0.0016)	-0.0026 (0.0049)	-0.0002 (0.0017)	0.0469* (0.0282)	-1.739** (0.739)	-0.0249 (0.0220)	0.0137 (0.0108)
Georgia	-0.0051** (0.0021)	-0.0073*** (0.0027)			-0.0071*** (0.0027)		-0.0711*** (0.0198)	0.560 (0.368)		
Iowa							0.0534** (0.0250)	0.275 (0.373)	0.0916*** (0.0226)	0.0322*** (0.0111)
Kansas								0.0106 (0.0228)		0.0171 (0.0104)
Minnesota	0.0031*** (0.0010)	0.0025* (0.0013)	-0.0030 (0.0021)	-0.0007 (0.0016)	0.00006 (0.0019)	-0.0002 (0.0017)	0.0153 (0.0162)	-0.596** (0.287)	-0.0450** (0.0210)	-0.0304*** (0.0100)
Ohio							0.0489*** (0.0138)	-0.187 (0.266)	-0.0300** (0.0138)	-0.0147** (0.0067)
Oklahoma	0.00056 (0.0019)	-0.0182*** (0.0032)		0.0063*** (0.0025)	-0.0002 (0.0045)	0.0078*** (0.0029)	0.161*** (0.0187)	0.675*** (0.186)	-0.0253 (0.0245)	0.0131 (0.0119)
Utah							0.0125 (0.0445)	-0.859** (0.421)		
All states	0.0016* (0.0008)	-0.0050*** (0.0012)	-0.0050*** (0.0017)	0.0019 (0.0014)	-0.0032** (0.0016)	0.0026* (0.0015)	0.1178*** (0.0094)	0.679*** (0.112)	-0.0148 (0.0161)	-0.0029 (0.0085)

Note. The Coefficient (Standard Error) Shown is for the P4P Variable, Indicating the within-Facility Difference in Outcomes for each P4P State Compared with All Non-P4P States after P4P was Implemented Compared to before. Each Coefficient Shown is derived from a Separate Regression. All regressions used a linear model except when the outcome was number of deficiencies for which a negative binomial model was used. Robust standard errors are in parentheses.

*** $p < .01$, ** $p < .05$, * $p < .1$.

HPRD, hours per resident day; LPN, licensed practical nurse; P4P, Medicaid nursing home pay for performance; RN, registered nurse.

high-Medicaid facilities. Under these programs, facilities with the largest number of Medicaid patient days were eligible for the largest bonus payments. Although larger incentives are generally thought to be more effective at motivating quality improvement, high-Medicaid facilities also often have the worst financial performance (Mor et al. 2004). Prior work has found that baseline financial performance predicts a nursing homes response to market-based incentives (Park and Werner 2011). Thus, because that the largest incentives targeted the facilities that were least able to respond and the smallest incentives targeted that facilities that would have been better able to improve quality under this program (low-Medicaid facilities), it is possible that the program had little net effect.

There are a number of other possible reasons for these disappointing findings. First, the incentives themselves may have been too small to effectively motivate changes in performance, particularly for the measures of staffing as staffing increases are very costly. It is intuitively logical that larger incentives may be more effective than smaller ones in P4P. Prior research has found that hospitals eligible for larger P4P incentives had a larger response to P4P (Werner et al. 2011). Ideally, bonus payment should exceed the marginal cost of quality improvement. In the case of nursing homes, P4P incentives were relatively small. In most states, the maximum possible bonus payment ranged between 3 and 4 percent of the per diem rates and in only one state (Oklahoma) did the maximum possible bonus exceed 5 percent. In addition, the amount of bonus actually paid was even smaller, below 2 percent of the total Medicaid nursing home budget in every state (Werner et al. 2010). It is possible that on average these bonuses were not high enough to motivate nursing homes to invest in quality improvement. However, in this study, the one state that achieved the most consistent success in response to P4P (Georgia), the size of the bonus was relatively small, at 1 percent of the per diem rate, suggesting that there were other nonfinancial factors behind the success of this program.

There may be ways to get more of a return without increasing the size of the award. Most nursing homes received annual bonuses for their performance. However, more frequent feedback on performance in the form of quarterly or even monthly payments may increase attention to performance in these areas because it provides frequent positive reinforcement (Thaler 1981; Kirby 1997). Although this approach may improve performance, the feasibility of frequent reporting may be limited, particularly at small nursing homes where the number of patients may be too small to get frequent reliable estimates of performance.

Another reason that the current P4P programs may have failed to consistently achieve quality improvement is that the incentives were paid to the nursing home, rather than to the individual staff members of the nursing home. Whether to target P4P incentives toward organizations or individuals remains unknown, as there are arguments to be made in support of each incentive target. Providing payments to organizations, rather than individuals, has conceptual appeal as quality deficits are thought to be system based. Payments to organizations can be used to help improve system failures by investing in large-scale approaches to quality improvement that would be expensive or infeasible for individuals to implement. Targeting payments to organizations may also be helpful if individuals are risk averse. Organizations, however, are difficult to motivate and to hold accountable for direct effects on patient care. Thus, providing payments to individuals may have a larger effect (Rosenthal and Dudley 2007). Bonus payments based on facility-level performance could be at least partially redistributed to managers or front-line providers (directly or by facility managers themselves) to increase personal motivation to improve performance.

Although our analytic approach, testing for changes in performance when P4P was implemented and controlling for secular changes in performance using non-P4P states, is well suited to answer questions about the effect of P4P on performance, there are several possible limitations to this approach. First, this approach treats all P4P programs as the same. While the main features of these P4P are similar across states, including higher per diem payment for higher performance on a set of performance measures, there are other ways in which these programs may differ. These might include factors such as a state's commitment to quality improvement and the state's ability to effectively communicate the incentive program to nursing homes. If these factors vary with time (or are correlated with the implementation of P4P), this will bias our results. However, we control for most state-level differences using our difference-in-differences approach. In addition, the multistate approach we use allows us to address the effectiveness of P4P over a wide range of settings rather than drawing conclusions about P4P based on a single state where the result may be an outlier rather than a true effect. In addition, the reduced form approach we use does not allow us to examine why some states, such as Georgia, were more successful than others. Second, our difference-in-differences approach assumes that non-P4P states make an appropriate counterfactual to what would have happened to performance in P4P states had P4P not been implemented. However, these P4P states were not randomly selected into these programs and the state's decision to adopt a P4P program may have

been endogenous to nursing home quality. For example, it is possible that states that were more concerned about nursing home quality both had better quality at baseline and were more likely to adopt quality improvement programs such as P4P. If nursing homes enrolled in P4P were higher quality to begin with, they may have had less room to improve under P4P, dampening its effect. Our descriptive analyses of states with and without P4P (Tables 2 and 3) suggest that P4P states were similar to non-P4P states in many ways that we can observe, although there were some differences in baseline nursing home quality. However, for the three quality measures where we document improvement under P4P (restraint use, moderate to severe pain, and development of pressure sores), differences in baseline quality between nursing homes in states that did and did not implement P4P were mixed. P4P states performed better at baseline for the measures of restraint use and development of pressure sores, while non-P4P states had a lower percentage of residents in moderate to severe pain. This mixed relationship between baseline quality and quality improvement under P4P suggests that differences in baseline quality did not systematically affect our results. Finally, in the cases where quality improved, we do not address whether quality truly improved or whether the documentation of these outcomes changed. We also do not assess whether other, nontargeted areas of care improved.

To our knowledge, this work provides the only comprehensive evaluation of P4P in nursing homes and one of the only evaluations of P4P on such a large scale. These results highlight the need to carefully design P4P programs to encourage changes in provider quality. This may include experimenting with targeting payments at individuals instead of facilities, combining P4P with other incentives to improve care coordination and patient outcomes, or using larger financial incentives in future P4P programs. Although our results are disappointing from a policy perspective, they can help inform the future use of P4P in nursing homes. In the meantime, expectations for the effect of P4P on improving quality of care in nursing homes should be tempered.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This research was funded by a grant from the National Institute on Aging (R01 AG034182-01). Rachel Werner was supported in part by a VA HSR&D Career Development Award. The authors gratefully thank Chris Wirtalla for his outstanding programming and research assistance.

Disclosures: None.

Disclaimers: None.

REFERENCES

- Abt Associates Inc. 2001. *Report to Congress: Appropriateness of Minimum Nurse Staffing Ratios in Nursing Homes Phase II Final Report*. Baltimore, MD: Centers for Medicare and Medicaid Services.
- Abt Associates Inc. 2006. *Quality Monitoring for Medicare Global Payment Demonstrations: Nursing Home Quality-Based Purchasing Demonstration*. Baltimore, MD: Centers for Medicare and Medicaid Services.
- Berlowitz, D. R., G. H. Brandeis, J. J. Anderson, A. S. Ash, B. Kader, J. N. Morris, and M. A. Moskowitz. 2001. "Evaluation of a Risk-Adjustment Model for Pressure Ulcer Development Using the Minimum Data Set." *Journal of the American Geriatrics Society* 49 (7): 872–6.
- Castle, N. G., B. S. Fogel, and V. Mor. 1996. "Study Shows Higher Quality of Care in Facilities Administered by ACHCA Members." *Journal of Long Term Care Administration* 24 (2): 11–6.
- Centers for Medicare and Medicaid. 2005. "Medicare Begins Performance-Based Payments for Physicians Groups: New Demonstration Program Tests Financial Incentives for Improved Quality and Coordination in Large Group Practices" [accessed January 16, 2008]. Available at <http://www.cms.hhs.gov/apps/media/press/release.asp?Counter=1341>
- Centers for Medicare and Medicaid. 2007. "Medicare Announces Plans for Home Health Pay for Performance Demonstration" [accessed on September 12, 2008]. Available at http://www.cms.hhs.gov/apps/media/press_releases.asp
- Centers for Medicare and Medicaid Services. 2003. "Premier Hospital Quality Incentive Demonstration" [accessed January 16, 2008]. Available at http://www.cms.hhs.gov/HospitalQualityInits/35_HospitalPremier.asp
- Fries, B. E., C. Hawes, J. N. Morris, C. D. Phillips, V. Mor, and P. S. Park. 1997. "Effect of the National Resident Assessment Instrument on Selected Health Conditions and Problems." *Journal of the American Geriatrics Society* 45 (8): 994–1001.
- Huber, P. J. 1967. "The Behavior of Maximum Likelihood Estimates under Non-Standard Conditions." In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, edited by L. M. Le Cam and J. Neyman, pp. 221–33. Berkeley, CA: University of California Press.
- Institute of Medicine. 1986. *Improving the Quality of Care in Nursing Homes*. Washington, DC: National Academies Press.
- Kaiser Family Foundation. 2007. "Medicaid and Long-Term Care Services and Supports" [accessed August 20, 2008]. Available at http://www.kff.org/medicaid/upload/2186_05.pdf
- Kane, R. L., G. Arling, C. Mueller, R. Held, and V. Cooke. 2007. "A Quality-Based Payment Strategy for Nursing Home Care in Minnesota." *The Gerontologist* 47 (1): 108–15.

- Kane, R. L., C. C. Williams, T. F. Williams, and R. A. Kane. 1993. "Restraining Restraints: Changes in a Standard of Care." *Annual Review of Public Health* 14: 545–84.
- Kidder, D., M. Rennison, H. Goldberg, D. Warner, B. Bell, L. Hadden, J. Morris, R. Jones, and V. Mor (2002). MegaQI Covariate Analysis and Recommendations: Identification and Evaluation of Existing Quality Indicators that are Appropriate for Use in Long-Term Care Settings. Contract No. 500-95-0062 TO #4, Abt Associates Inc.
- Kirby, K. 1997. "Bidding on the Future: Evidence against Normative Discounting of Delayed Rewards." *Journal of Experimental Psychology: General* 126: 54–70.
- Li, Y., X. Cai, L. G. Glance, W. D. Spector, and D. B. Mukamel. 2009. "National Release of the Nursing Home Quality Report Cards: Implications of Statistical Methodology for Risk Adjustment." *Health Services Research* 44 (1): 79–102.
- Minnesota Department of Human Services. 2007. "Minnesota Quality Indicators and Adjusters—Detailed" [accessed December 14, 2011]. Available at http://www.dhs.state.mn.us/main/groups/aging/documents/pub/dhs_id_051942.pdf
- Mor, V., O. Intrator, B. E. Fries, C. Phillips, J. Teno, J. Hiris, C. Hawes, and J. Morris. 1997. "Changes in Hospitalization Associated with Introducing the Resident Assessment Instrument." *Journal of the American Geriatrics Society* 45 (8): 1002–10.
- Mor, V., J. Zinn, J. Angelelli, J. M. Teno, and S. C. Miller. 2004. "Driven to Tiers: Socioeconomic and Racial Disparities in the Quality of Nursing Home Care." *Milbank Quarterly* 82 (2): 227–56.
- Morris, J. N., B. E. Fries, D. R. Mehr, C. Hawes, C. Phillips, V. Mor, and L. A. Lipsitz. 1994. "MDS Cognitive Performance Scale." *Journal of Gerontology* 49 (4): M174–82.
- Morris, J. N., B. E. Fries, and S. A. Morris. 1999. "Scaling ADLs within the MDS." *Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 54 (11): M546–53.
- Morris, J. N., T. Moore, R. Jones, V. Mor, J. Angelelli, K. Berg, C. Hale, S. Morris, K. M. Murphy, and M. Rennison. 2003. *Validation of Long-Term and Post-Acute Care Quality Indicators*. Baltimore, MD: Centers for Medicare and Medicaid Services.
- Mukamel, D. B., L. G. Glance, Y. Li, D. L. Weimer, W. D. Spector, J. S. Zinn, and L. Mosqueda. 2008. "Does Risk Adjustment of the CMS Quality Measures for Nursing Homes Matter?" *Medical Care* 46 (5): 532–41.
- Norton, E. C. 1992. "Incentive Regulation of Nursing Homes." *Journal of Health Economics* 11 (2): 105–28.
- Nursing Home Quality Initiative. 2004. *Quality Measures Resource Manual: Enhanced Set of Quality Measures*. Baltimore, MD: Centers for Medicare and Medicaid Services.
- Park, J., and R. M. Werner. 2011. "Changes in the Relationship between Nursing Home Financial Performance and Quality of Care under Public Reporting." *Health Economics* 20: 783–801.
- Petersen, L. A., L. D. Woodard, T. Urech, C. Daw, and S. Sookanan. 2006. "Does Pay-for-Performance Improve the Quality of Health Care?" *Annals of Internal Medicine* 145 (4): 265–72.

- Robinson, J. C. 2001. "Theory and Practice in the Design of Physician Payment Incentives." *Milbank Quarterly* 79 (2): 149–77.
- Rosenthal, M. B., and R. A. Dudley. 2007. "Pay-for-Performance: Will the Latest Payment Trend Improve Care?" *Journal of the American Medical Association* 297 (7): 740–4.
- Rosenthal, M. B., and R. G. Frank. 2006. "What is the Empirical Basis for Paying for Quality in Health Care?" *Medical Care Research and Review* 63 (2): 135–57.
- Rosenthal, M. B., B. E. Landon, S.-L. T. Normand, R. G. Frank, and A. M. Epstein. 2006. "Pay for Performance in Commercial HMOs." *New England Journal of Medicine* 355 (18): 1895–902.
- Shorr, R. I., R. L. Fought, and W. A. Ray. 1994. "Changes in Antipsychotic Drug Use in Nursing Homes during Implementation of the OBRA-87 Regulations." *Journal of the American Medical Association* 271 (5): 358–62.
- Snowden, M., and P. Roy-Byrne. 1998. "Mental Illness and Nursing Home Reform: OBRA-87 Ten Years Later. Omnibus Budget Reconciliation Act." *Psychiatric Services* 49 (2): 229–33.
- Thaler, R. H. 1981. "Some Empirical Evidence on Time Inconsistency." *Review of Economic Studies* 23: 165–80.
- Werner, R. M., J. T. Kolstad, E. A. Stuart, and D. Polsky. 2011. "The Effect of Pay-for-Performance in Hospitals: Lessons for Quality Improvement." *Health Affairs* 30 (4): 690–8.
- Werner, R. M., R. T. Konetzka, and K. Liang. 2010. "State Adoption of Nursing Home Pay-for-Performance." *Medical Care Research and Review* 67: 364–77.
- White, H. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48: 817–30.
- Wunderlich, G. S., and P. Kohler. 2000. *Improving the Quality of Long-Term Care*. Washington, DC: Division of Health Care Services, Institute of Medicine.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix SA2: Resident-Level Covariates Included in Regression Analyses.

Appendix SA3: Results of Robustness Checks Compared with the Main Specification (as reported in Table 4 of the manuscript). (The table displays the coefficient for the P4P variable, indicating the within-facility difference in outcomes for P4P states compared with non-P4P states after P4P was implemented compared with before [in the 1 year following P4P implementation]).

Appendix SA4: Effect of Using Narrowed Compactor States. (State-

by-state regressions showing effect of P4P implementation on nursing home quality in each state [1 year following P4P implementation]. Each P4P state is paired with the neighboring non-P4P state listed in the table. The coefficient [standard error] shown is for the P4P variable, indicating the within-facility difference in outcomes for each P4P state compared with all non-P4P states after P4P was implemented compared to before. Each coefficient shown is derived from a separate regression).