Pay for performance (P4P) aims to use reimbursement to incentivize providers to deliver high-quality services. This article examines P4P in five Medicaid nursing home programs: Iowa, Minnesota, Oklahoma, Utah, and Vermont. It describes each program and draws lessons regarding program participation, financing, measurement, administration, and development. Findings highlight the importance of obtaining stakeholder input, both initially and on an ongoing basis. Findings also highlight the need to provide opportunities for acceptance and learning by phasing in programs slowly, beginning with performance measurement, followed by public reporting and, finally, by introducing P4P incentives. Funding P4P using new appropriations, incorporating multiple quality measures and domains, and relying on existing data sources when possible were deemed important; so, too, was allowing programs to evolve over time to account for innovations in quality measurement.

Because nursing homes (NHs) are a heavy burden on the public purse, they pose particular challenges to public administrators and regulators. Much of their funding comes from Medicaid, the second-largest and fastest-growing budget item for states (MedPac 2012), making up 22.3 percent of state expenditures in 2010 (NASBO 2011). Roughly 17 percent of this—4 percent of all state revenues—was spent on NHs (Kaiser Family Foundation 2012a, 2012b). Medicare also spends 6 percent of its budget on NHs, despite its significant coverage limitations (MedPac 2012). Thus, in view of the projected increases in demand for long-term care as a result of the aging of the population, mechanisms to contain NH costs and ensure the appropriate expenditure of public monies are critically important.

Potentially, government has substantial leverage: a full 54 percent of NH funding comes from either Medicaid (32 percent) or Medicare (22 percent) (MedPac 2012). However, policy tools historically used in the NH sector have proved disappointing. Initially, these focused on penalizing facilities that failed to meet federal standards, through NH inspections conducted by state officials. As in other sectors, however, focus has shifted toward market-based tools: making better information available to allow informed choice among facilities and creating financial incentives to improve quality (Konetzka and Werner 2010). Both strategies depend on robust quality indicators that have been developed and tested over the last 20 years (Castle and Ferguson 2010). The informed choice strategy is manifested in the creation of state and federal report cards, which publish information on NH quality, including staffing, deficiencies, and clinical quality indicators deriving from a federally mandated resident assessment instrument (Castle and Lowe 2005; Miller and Mor 2008).

However, it is the latter strategy that is the focus of this study: motivating quality improvement through better alignment of financial incentives (Arling, Job, and Cooke 2009). NH pay for performance (P4P) departs from traditional fee-for-service payment methods, in which providers are paid on the basis of the frequency or volume of the services that they provide, regardless of outcomes. In contrast, under P4P, provider reimbursement is tied to specific outcome-based quality measures, using methodologies that are currently evolving. Most often, P4P supplements a facility’s daily rate, based on points awarded across a range of parameters; less frequently, facilities are retroactively paid bonuses based on points earned (Werner, Konetzka, and Liang 2010). In addition, rewards may accrue not only to positive outcomes but also to the processes and structures that set the stage for quality.

This market-based approach is consistent with trends in public administration more generally, which seek to use financial incentives to improve the quality of publicly funded services. Although much attention is focused on rewards at the individual level (as in merit pay for teachers) (Perry, Engbers, and Jun 2009), performance-based contracting also occurs at the organizational level, in fields as various as substance abuse treatment (Shen 2003), welfare-to-work programs (Heinrich and Choi 2007), and international...
An effective P4P system also acknowledges and minimizes the potential for negative spillovers (Roland 2012)—that is, strategic behaviors aimed at maximizing rewards rather than improving quality. For example, a poor choice of measures may result in “teaching to the test” or “gaming.” Cherry-picking is another challenge: consequently, health care has focused on methods of risk adjustment, whereby organizations are paid more for treating riskier individuals (which may result in “upcoding,” i.e., the exaggeration of acuity levels). Rewarding good performers also diverts attention (and, possibly, resources) from less successful but essential service providers, which often serve less advantaged clients, for whom quality improvement may be most critical (Casalino et al. 2007, Rosenthal et al. 2005).

Despite potential challenges, “value-based purchasing” is attracting increased interest in health care. Indeed, the Patient Protection and Affordable Care Act (ACA) specifically mandates P4P for NHs and home health care providers and requires pilot testing of P4P for long-term care hospitals, rehabilitation hospitals, and hospice. P4P in NHs is not a new strategy: it was first tested in 1980 through a randomized controlled experiment with 36 San Diego facilities (still the only formal evaluation of NH P4P) (Norton 1992; Weissert et al. 1983). However, the Centers for Medicare and Medicaid Services revived interest in P4P by launching the NH Value-Based Purchasing Demonstration in 2009, which tests P4P within Medicare (White et al. 2009). Results are as yet unavailable and, in any case, will have limited relevance, as most P4P programs operate under Medicaid. A study by Werner, Konetzka, and Liang (2010) found 14 states with planned or existing Medicaid NH P4P programs, with 3,050 facilities enrolled in the nine programs operational in 2007. These programs vary considerably both in terms of the chosen performance measures and in the size and type of financial incentive used (Arling, Job, and Cooke 2009; Briesacher et al. 2009; Mollot, Rudder, and Samji 2008; Werner, Konetzka, and Liang 2010).

This article addresses the prevailing knowledge gap about NH P4P by reporting results of in-depth case studies on its adoption and administration in five states—Iowa, Minnesota, Oklahoma, Utah, and Vermont—with a goal of informing the design and implementation of P4P programs more generally.

**Methods**

Based on our survey of the limited literature and conversations with key informants, we identified critical characteristics of Medicaid NH P4P programs, all of which speak to the issues outlined earlier. These include their complexity, level of stakeholder participation, financing arrangements, administration, and development. Using available data on the 14 states implementing P4P (Werner, Konetzka, and Liang 2010), Iowa, Minnesota, Oklahoma, Utah, and Vermont were selected to ensure maximum variation among these parameters (see table 1). One of the most important factors is complexity, that is, the range of factors taken into account and used to calculate rewards. Among the selected states, Iowa and Minnesota were considered complex; Oklahoma, moderately complex; and Utah and Vermont, relatively simple. Program sustainability is also an issue: Oklahoma, Utah, and Vermont all had active P4P programs, while the Iowa and Minnesota programs had been suspended because of the fiscal crisis.
specific strategies used to operationalize and reward performance, is another important parameter because it identifies priorities for quality improvement. Administration refers to basic program design attributes, including system complexity, data sources, and reward structures; these parameters speak to the feasibility and fairness of the P4P system. Development refers to program evolution, including phase-in, modification, and public reporting.

Data for the case study comparisons derive from two primary sources: key informant interviews and state-provided documents. In-depth, semistructured interviews were conducted over the telephone, investigating subjects’ perceptions of each state’s P4P program. Respondents were selected based on their knowledge of each state’s P4P program, focusing on administrators and consultants responsible for implementing NH payment policy. Interviews were conducted between December 16, 2010, and January 7, 2011, and lasted approximately 60 minutes. Conversations were recorded (with each respondent’s permission) and transcribed. In all, 11 individuals with expert knowledge were identified in the five states and agreed to interviews. Each transcript was subsequently coded to identify recurring themes, using inductive coding methods. These rely on a close reading of those transcripts, where the researcher breaks the material into naturally occurring segments or chunks, and then systematically and iteratively identifies concepts and themes organizing them (Miles and Huberman 1994).

Interviews focused on soliciting respondents’ experiences with the P4P programs in their state. Interviewees were asked to comment on the design, implementation, and effectiveness of their state’s P4P program. Interviewees were also asked about factors that may have facilitated or impeded program development and administration. Key barriers and critical success factors were identified as well.

<table>
<thead>
<tr>
<th>Implementation year</th>
<th>Iowa</th>
<th>Minnesota</th>
<th>Oklahoma</th>
<th>Utah</th>
<th>Vermont</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program design</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Program implementation</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Program modification</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Percent participating</td>
<td>0</td>
<td>0</td>
<td>98%</td>
<td>80%</td>
<td>25%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Financing</th>
<th>Funded with new/existing money</th>
<th>Existing</th>
<th>Existing</th>
<th>New</th>
<th>New</th>
<th>Existing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program active/suspended</td>
<td>Suspended</td>
<td>Suspended</td>
<td>Active</td>
<td>Active</td>
<td>Active</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Staffing</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer satisfaction</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Inspection performance</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Clinical quality indicators</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Person-centered/quality of life</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Efficiency</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Access</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Employee satisfaction</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Quality improvement plan</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Administration</th>
<th>System complexity</th>
<th>Complex</th>
<th>Complex</th>
<th>Moderate</th>
<th>Simple</th>
<th>Simple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relies primarily on existing data</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Industry-wide benchmarks used</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Composite quality index used</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Reward structure</td>
<td>% Per diem rate</td>
<td>% Per diem rate</td>
<td>% Per diem rate</td>
<td>Lump sum</td>
<td>Lump sum</td>
<td></td>
</tr>
<tr>
<td>Rewards improvement</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Funds allocated competitively</td>
<td>None</td>
<td>Some</td>
<td>None</td>
<td>Some</td>
<td>All</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Development</th>
<th>Program phased in slowly</th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program modified over time</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Program linked to public reports</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

| aDistributed on a lump-sum basis at the end of the rate year.

**Findings**

The P4P systems of Iowa, Minnesota, Oklahoma, Utah, and Vermont are discussed in turn. A detailed review of each state’s P4P program is provided, along with lessons learned.

**Iowa: The Importance of Stakeholder Involvement**

Iowa’s NH P4P program demonstrates considerable learning from experience, based on extensive stakeholder involvement. It shifted from a system that began in 2002 by awarding facilities based on prior performance to a system that required continued “good behavior” on the part of NHs. It also changed from a system that rewarded results to one that emphasized process by requiring facilities to collect and use resident satisfaction data. Despite these steps, financing has proved a fatal problem: because of the state budgetary situation and the resulting lack of funding and staff resources, program implementation and the state’s P4P workgroup have been suspended. Although the state had a potential source of new revenue...
through its adoption of a provider tax, those revenues were diverted toward improved compensation for direct care staff—indicating the state’s political priorities.

Initially, facilities received points based on measures derived from cost reports (e.g., nursing hours, employee retention), state Department of Inspections and Appeals data (e.g., deficiency free surveys, regulatory compliance), and the state Long-Term Care Ombudsman’s office (e.g., complaint resolution). For most measures, facilities could receive up to 10 points based on performance relative to established benchmarks, earning an additional point for scoring higher than the fiftieth percentile on a resident satisfaction survey. The latter instrument was administered independently by an outside entity. Credit was contingent on a response rate of at least 35 percent. This approach encouraged providers to seek information about resident experiences whether or not they anticipated good results. Overall rewards were based on the prior year’s performance and calculated as a percentage of the per diem rate, so that those scoring 3–4 points would earn an additional 1 percent; 5–6 points, 2 percent; and 7 or more points, 3 percent. Those scoring fewer than 3 points received no reward.

However, in 2008, several news sources revealed serious deficiencies in facilities that had qualified for incentive payments during the previous year, which continued to be paid in the next year. The legislature subsequently modified the program to pay the incentive payments retroactively as a lump-sum payment, using the per diem add-on methodology at the end of the fiscal year, contingent on “good behavior.” End-of-year payments were reduced by 25 percent for each subsequent serious deficiency and withheld altogether for unaddressed deficiencies. Iowa’s experience, therefore, highlights potential drawbacks of using historical data to reward facilities. Moreover, the state’s responsiveness to media reports suggests the issue’s political sensitivity.

In 2010, the state overhauled the P4P program, though its provisions have yet to be funded and implemented. Moving forward, facilities will receive up to 100 points based on measures in four quality domains: quality of life (25 points), quality of care (59 points), access (8 points), and efficiency (8 points). The quality of life domain includes person-directed care (e.g., dining activities, resident choice) and resident satisfaction (e.g., resident/family survey, complaint resolution). The quality of care domain includes state inspection (e.g., deficiency-free surveys, regulatory compliance), staffing (e.g., nursing hours, turnover), and nationally reported quality measures (e.g., high-risk pressure ulcers, physical restraints). Facilities receiving serious inspection deficiencies will be ineligible to participate, and their performance on quality measures will be reported publicly.

As before, retroactive incentive payments will be based on continued “good behavior.” Those scoring 0–50 points will not receive additional payment; those receiving 51–60 points, an additional 1 percent; 61–70 points, 2 percent; and so on, up to 5 percent. State officials expected that the add-ons will range from $1.25 to $6.25 per patient day; extra payment must be used to support direct care staff through wages, benefits, and training.

The overhaul included a new emphasis on process: specifically, on collecting and using resident satisfaction data. Initially, facilities could choose whether to conduct resident satisfaction surveys; if they did, they were required to use a uniform 10-question Resident Opinion Survey and were rewarded for scoring over the median. Under the new, yet-to-be-implemented system, facilities must collect resident satisfaction data, but they are no longer required to use that particular survey tool. Again, points are awarded only if the facility’s response rate is at least 35 percent or higher, regardless of results.

Another process element included in the new system is activities improving resident quality of life. Facilities can self-certify the presence of desired activities and processes by filling out forms and providing evidence (such as one month’s worth of menus and meal times). Verifying this information, however, is a challenge, and this perhaps explains why facilities that receive national accreditation in person-centered care will automatically receive all possible points in this area. In general, the subjective nature of quality of life has made it more difficult to measure than other quality dimensions.

To save costs and encourage participation, P4P program developers aimed to allow providers to use current administrative systems to report performance wherever possible. Developers also aimed to use industry-wide benchmarks when awarding points, so that any facility could qualify for extra payment. This would not have been possible if, say, points were awarded only to those scoring above a certain percentile on a particular quality measure.

Iowa’s experience demonstrates the importance of having broad discussions, with extensive stakeholder involvement, throughout program implementation and development. To design the state’s initial P4P system, the state convened a workgroup comprised of industry representatives, advocacy groups, state agency personnel, legislative staff, and other interested parties. This group met annually thereafter to review and update the system, eventually redesigning the state’s P4P program altogether at the direction of the legislature. Other participants included the state’s Long-Term Care Ombudsman and survey and certification staff, as well as outside consultants (providing perspectives based on other states’ experiences). Clarifying program goals early on was seen as essential, giving stakeholders a sense of the stakes involved, as was engaging stakeholders early on and checking in with them frequently. Continued interaction kept everyone on the same page; eased communication among the state, providers, and other participants; and promoted buy-in and consensus over program attributes.

Iowa’s P4P system also demonstrates considerable evolution over time. It began by rewarding facilities during the rate year on the basis of 10 measures derived from the previous year. It was then modified so that payments would be received at the end of the year, contingent on continued positive performance. Subsequently, the system was revised to account for new thinking about quality measurement and what kinds of facility behaviors should be rewarded. All of this work is for naught, however, unless sustainable financing can be secured.
Minnesota: The Importance of Quality Measurement

The Minnesota system demonstrates a progressive shift toward more complex reward calculations. The state bases awards on a facility’s quality score, which ranges from 0 to 100. In the first year, 2006, this score was generated from measures readily derived from state administrative data systems, including 24 clinical quality indicators (40 points), direct staff retention (25 points), direct staff turnover (15 points), use of pool staff (10 points), and survey deficiencies (10 points). The payment was calculated as a percentage of a facility’s base rate: in the first year, those scoring 0–40 did not receive an add-on; those scoring 100 received a 2.4 percent add-on; and those scoring 41–99 received an add-on proportionate to the summary quality score.

In the second year, staffing turnover was replaced by the direct care staffing level, and a dimension measuring resident satisfaction/quality of life was introduced (based on a standardized interview conducted by an independent contractor with a random sample of facility residents). Weighting also changed: points devoted to staffing were reduced from 50 to 35 to allow the assignment of 20 points to the new satisfaction/quality dimension. Rather than comparing relative performance among facilities, points were awarded on the basis of fixed standards, with scores above or below certain thresholds yielding greater rewards. In addition, both resident satisfaction/quality of life and the clinical quality indicators were risk adjusted to account for acuity levels within the populations served.

The biggest impact on the program, however, was the inability to secure ongoing funding for awards. Although the quality add-on paid under the program was initially significant, budgetary constraints reduced payments from an average 1 percent supplement in year one, to an average of 0.13 percent in year two, to nothing in year three, when the program was suspended. Despite being suspended, the program continues to have an impact because the quality information developed under it is still collected and reported publicly through the Minnesota NH Report Card.

Alongside the 2006 implementation of P4P, Minnesota introduced its Nursing Facility Performance-Based Payment Program. This initiative allows facilities to receive incentive payments of up to 5 percent of their base payment rate for projects lasting from one to three years. Facilities apply competitively, proposing innovative projects aiming to improve quality or efficiency or to increase successful diversions or discharges to a home- or community-based alternative. Incentive payments, which are based on project scope and complexity, are incorporated into the winning facilities’ rates and range from $0.32 to $13.32 per patient day (Cooke et al. 2010). During the program’s first two years, 158 of the state’s 383 NHs were funded to implement 45 projects, most of which address clinical quality and technology, while others pertain to psychosocial issues, home- and community-based services, and organizational change. Funding appropriated for this program increased from $1.2 million in 2007 to $6.7 million in state 2011, approximately 1.5 percent of the state’s Medicaid NH budget.

In short, Minnesota’s example illustrates the need for a sustainable funding strategy. Otherwise, payouts can vary year to year depending on the state budgetary situation and other factors. In Minnesota, these payouts ranged from an average of 1.0 percent in the first year, 0.13 percent in the second year, and nothing subsequently. The size of the incentive payment declined because, once awarded, it became permanently incorporated into the base rate and unavailable for future add-ons. Some suggest that Minnesota’s dormant P4P system could be reactivated by a redistribution of existing resources through a rate cut, although others believe that this approach would be too punitive. Another strategy would be to tie resurrection of the state’s P4P program to, say, a cost-of-living adjustment, with 1.0 percent of that adjustment going into the quality add-on and the remaining 2.0 percent into an across-the-board increase. This strategy, though appealing, is unlikely in the current fiscal environment, where the latest adjustment was zero.

The Minnesota experience also highlights the tension between focusing on outcomes when distributing quality-related rewards and rewarding the structure and processes that necessarily precede those outcomes. Despite system developers’ preference for rewarding performance primarily on the basis of outcomes, a structural measure (staffing) was given prominence because of insistence from stakeholder groups, even though developers felt that programs that effectively measure clinical and quality outcomes do not have as much need to measure process and structure. Moreover, although system developers were most interested in incentivizing absolute performance, they also valued promoting sustained improvement over time and, as such, incorporated measures that rewarded both absolute performance and year-on-year improvements.

Developing solid quality measures was seen as a central goal in Minnesota’s efforts to improve NH quality. Although Minnesota’s strategy for measuring resident satisfaction/quality of life has been criticized for being too detailed, lengthy, and costly, program administrators felt that systematic measurement—with a valid and reliable instrument—trumps expediency. Program administrators therefore chose to employ an outside contractor, using a standard, in-depth instrument.

Risk adjustment was deemed especially critical for ensuring that facilities serving disproportionately high-risk residents were not penalized by P4P. Thus, it might be reasonable to look beyond the clinical quality indicators and satisfaction/quality of life measures to risk-adjust other outcomes, such as inspection deficiencies, which are statistically more likely in larger facilities.

Minnesota’s P4P program further highlights the importance of in-depth and continuous involvement on the part of key stakeholders, including consumer advocacy groups, the NH associations, labor unions, the department of health, and the ombudsman’s office. In addition, a panel of nursing directors was convened to provide advice on particular quality measures. Lessons from other states’ experiences also contributed. By actively engaging stakeholders, key constituency groups were able to voice approvals and objections, leading eventually to agreement on basic system design. While there was disagreement about program details related to measurement,
weighting, and scoring, many of these differences were ironed out over time.

The Minnesota case also suggests the utility of phasing in P4P slowly, beginning with transparent performance measurement, adding the release of public report cards, and culminating in P4P. Report cards not only provide consumers, discharge planners, and other referral staff with the information necessary to compare facilities, they also provide data to benchmark facility performance. Additionally, report cards rely on a strong set of quality measures: once developed, identified, and publicly disclosed, they can be tied to payment and refined over time.

**Oklahoma: The Importance of Supplemental Financing**

Oklahoma’s voluntary P4P program, which went into effect in July 2007, has been extremely successful in recruiting participation: only 2 percent of facilities opted out. The state offered a significant incentive for facilities to take the first step. To participate, facilities only need sign and execute a P4P contract amendment, which commits them to filing all required forms and cost reports and submitting the requisite monthly administrative and annual survey data. Initially, participants received a 1.0 percent incentive payment for doing so. (These are made quarterly as a percentage of the Medicaid per patient day payment rate.) Subsequently, facilities could earn up to 10 points on the basis of 10 quality indicators. For every two points earned, facilities receive an additional 1.0 percent; at first, facilities received a point for each of these measures for scores exceeding the median. Beginning in 2010, however, thresholds were established for each measure, allowing all facilities to receive points if they perform well enough on a particular indicator. The add-on ranges from 1 percent to 5 percent, amounting to an extra $1.09 to $5.45 per day. The program’s entire annual budget was reported to be $560,000.

To develop and administer the system, the state contracted with a private vendor, My InnerView. Data for two items—quality of life and resident/family satisfaction—derive from My InnerView’s family and resident satisfaction surveys. Data for another two items—employee satisfaction and system wide culture change (i.e., resident-centered care)—derive from My InnerView’s employee satisfaction survey. Response rates for employee surveys must be at least 30 percent; a facility’s resident/family surveys may also be disqualified if response rates are less than 65 percent. Remaining indicators are reported through the My InnerView Quality Profile, which includes information on staffing, state inspection compliance, and clinical measures (e.g., falls, catheters, physical constraints, pressure ulcers). These measures and scores are available online to both NHs and the general public.

Oklahoma’s experience highlights the importance of developing and maintaining good relations with providers. Because of the ongoing relationship between the state and the associations, a large proportion of the NH industry was reportedly on board with and even contributed to writing the policy eventually approved by the state legislature. There are also several ongoing taskforces involved in the program that continue to ensure stakeholder input. Program administrators explained that extensive industry involvement explained the limited resistance to requirements such as the My InnerView resident, family, and employee surveys.

Adoption of P4P in Oklahoma was eased because the state drew on new money when implementing the program, at least initially. The state could do so because, when implementation took place, extra funding was available through the state’s annual cost-of-living adjustments. These additional funds were used to fund the P4P program rather than distributed equally among facilities. Thus, facilities did not have to take a reduction in rates for the system to be put into place, which helped with buy-in. Recently, funding for the P4P program has remained static as a result of prevailing budget difficulties. In the unlikely event that additional money becomes available, the state would consider increasing the incentive payments even more.

Program administrators indicated that minimal additional data-collection requirements have also served to further industry involvement. With the exception of the resident and employee surveys, most of the data used—the clinical measures, deficiencies, staffing—are taken from existing sources. Furthermore, data reporting is performed online, which saves a lot of time administratively. The relative simplicity of Oklahoma’s reimbursement system has helped with administration as well.

As noted, the state delegated development and administration of P4P to My InnerView, which works under general guidelines included in the program’s authorizing legislation. My InnerView was chosen largely because the state perceived it to be a well-equipped management company with the expertise necessary to choose the specific quality measures that fell within the guidelines established by the state. The state also delegated the administration of uniform resident and employee satisfaction surveys to My InnerView rather than allowing facilities to choose their own vendors. This delegation of authority promoted uniformity in quality measurement for both survey content and data collection.

Oklahoma’s experience also demonstrates the utility of phasing in P4P over time. As noted, before implementing the program fully, facilities were given additional reimbursement for filling out the new surveys, which worked well to encourage initial participation. Facilities thus had the opportunity to acculturate to the My InnerView system, particularly with respect to developing the routines necessary to execute whatever new data-entry requirements were put into place. Facilities then began receiving payment based on their quality scores, after which higher thresholds for receiving additional funds were used to fund the P4P program rather than distributed equally among facilities. This delegation of authority promoted uniformity in quality measurement for both survey content and data collection.

Oklahoma’s experience also highlights the importance of phased implementation and the benefits of slowly phasing in a P4P program, which in part explains its estimated 80 percent
participation rate. It began as a basic program in 2004, when facilities simply needed to show that they had a quality improvement plan and a measurement strategy. Although the program became more complex, it remains relatively straightforward. For instance, a facility with a serious inspection violation is now disqualified from participating, while one with a substandard quality of care citation becomes eligible for only half the potential award. The state also now requires facilities to contract with independent third parties to conduct customer satisfaction surveys and, when items rate below average on that survey, demonstrate action plans addressing them. Additionally, facilities must have a plan for culture change (i.e., resident-centered care) and an employee satisfaction program in place.

These requirements were introduced incrementally: in year one, facilities needed only to show that they had a program in place; in year two, they needed to demonstrate progress on a measure related to that program. This phase-in proved particularly beneficial for less sophisticated facilities by encouraging them to create something more meaningful than what may have existed previously. The slow phase-in was also beneficial where measurement of customer satisfaction was concerned. Initially, the state simply required facilities to show evidence of a patient satisfaction survey. Then, the following year, the state required facilities to demonstrate that one of the performance indicators targeted was related to their survey results.

The program has also benefited from new monies. Although the state planned to move ahead with P4P whether or not new money was made available, additional federal dollars became available when the state adopted a provider tax. The QI1 award pool, as the state’s P4P program is known, is currently $1.0 million annually, and individual facility payments, which are distributed on the basis of Medicaid patient days, range from $3,000 to $30,000 per year.

Utah has also established a separate pool of money ($4,275,900 in 2011) for a second initiative, known as the QI2 program. This targets capital improvements that improve quality. Facilities that make such improvements, and provide evidence that they paid for them, receive a fixed amount per Medicaid certified bed. The state offers facilities several capital improvement options: a nurse call system, vans, heavy-duty lifts, electronic health records, bathing systems, quality training, and dining systems. Options will be continuously added to the QI2 menu, giving facilities many ways to earn an award. This program evolved in a stepwise fashion as well, beginning with a small pot of money before expanding to the $4 million, nine-item program operating today.

Utah keeps its QI1 and QI2 initiatives relatively flexible, rewarding quality and improvement but permitting facilities to decide which improvements are most important. This flexibility allows each facility to grow at its own rate and assess its own strengths and weaknesses. Providers may choose their own contractor to measure consumer satisfaction, as well as the benchmarks used to formulate action plans addressing survey items rated below average. Providers also have discretion in choosing among capital improvement options, in addition to choosing how they demonstrate commitment to resident-centered care and employee satisfaction.

Although this methodology is fairly straightforward, the discretion that it offers providers has a downside: because each facility establishes its own areas for improvement, state administrators must assess different materials for each facility, which can be time consuming. Furthermore, facilities often delay submitting their applications, leaving administrators with little time to evaluate performance based on the large stack of documents provided. Program administrators suggested that deadlines be staggered, especially in a large state, to avoid such administrative burdens, and that sufficient staff be made available. Undertaking extensive outreach, sending out reminders and e-mail alerts, and attending association meetings might prove helpful as well.

Overall, the program’s high levels of participation can be explained by its relative simplicity and its flexibility—namely, providers’ freedom to select which areas to focus on as the basis for their reward. The availability of financing was another key factor. Industry buy-in was also boosted by participation on the state’s quality committee, which comprises state staff from Medicaid and licensing, resident assessment, and certification.

Vermont: The Importance of Simplicity

Vermont’s P4P initiative, launched in 2004, awards the top five NHs in the state a total of up to $500,000 annually. To be eligible for a quality incentive reward, facilities must be deficiency free on their most recent health and fire safety inspection survey—a criterion that disqualifies most NHs—and must participate in another quality initiative, the Gold Star Employer Program. Facilities that qualify are then ranked according to objective standards of quality to determine winners. If more than five facilities qualify, tied facilities are ranked according to cost efficiency, that is, allowable costs per day. In recent years, approximately a quarter of the NHs in the state participated.

Award amounts are based on the ratio of its Medicaid days to the total Medicaid days for all qualifying facilities. Each winning facility has discretion in how its award is used, though funds must be used to improve quality of life among Medicaid-eligible residents. Typically, the reward is rolled back into the facility to add a garden path or change some structure or to add new programming for the residents.

The state’s voluntary Gold Star Employer Program focuses on best practices for recruitment and retention of direct care staff and has three parts: facilities must conduct a self-assessment of best practices in seven areas (e.g., recruitment, orientation/straining, professional development), develop a plan for implementing a new best practice or significantly expanding an existing one, and document progress toward their best practice goals. Gold Star is jointly sponsored by the Vermont Department of Disabilities, Aging and Independent Living and the Vermont Health Care Association, the leading NH industry advocate in the state.

Vermont’s example highlights two P4P design issues: the question of how much compensation is necessary to motivate change, and the question of the breadth of impact.
top-performing facilities in the state because of the limited funding available. In combination with the Gold Star Employer Program, it has been apparent effective in encouraging high performers to strive toward continuous improvement. Interestingly, it appears that the positive publicity associated with the award’s selectiveness is a stronger motivator than the monetary award. As part of the program, the state holds an awards ceremony at each of the winning homes, and it is a big honor to be recognized.

However, there is little, if any, incentive to motivate improvements among other, lower-performing facilities. That might change if funding became available and program modifications were introduced. Currently, eligibility for the P4P program requires a deficiency-free survey. To broaden the program, state officials suggested that appropriate awards and recognition might be created for facilities that reduce the number of deficiencies or go without a severe deficiency within a specified period of time. Interviewees also suggested that larger states might want to set a slightly lower bar, such as requiring facilities to be free of all major deficiencies, as they have much larger volumes of residents and a greater chance of receiving a citation. This requirement could be made more stringent over time as facilities improve or the survey and certification process evolves.

It should be noted that a long history of collaboration between the state and NH association considerably facilitated implementation of Vermont’s quality incentive initiative. As a small state with only 42 NHs and one NH association, buy-in has been easier than in states with multiple associations representing hundreds of facilities. This close collaboration between the state and facilities resulted in a consensus to focus the program primarily on staffing, under the assumption that the quality and motivation of facility staff is the most important determinant of quality within an NH.

### Discussion

Findings from this review of five states’ P4P programs parallel existing literature on pay for performance in other contexts (see, e.g., Perry 2003). These reinforce the importance of the parameters identified here for analysis, shown in table 1: program participation, financing, measurement, administration, and development. A consistent finding in our study is the need to incentivize engagement—by ensuring that ongoing funding for programs is secured and by devising reward systems that are perceived as fair and workable. This means that state officials should seek to minimize the administrative burden on facilities and include a range of criteria that address different aspects of quality, including some that encourage improvement among low- or middle-tier performers. Practically speaking, a slow phase-in, incorporating provider feedback, will maximize the chances of developing a strong system that enjoys provider support (Perry 2003). Of the states studied, Oklahoma most closely meets these best practices criteria, although its strategies for incentivizing lower-performing providers could be strengthened.

Clearly, the most important factor in any P4P program is the availability of funding to make the awards. However, funding such programs is difficult and requires the political will to prioritize NH quality and identify new funding sources. The alternative—reallocating the existing pot of money for provider reimbursement—creates winners and losers and, likely, political opposition.

The additional pressure created by the difficult budget environment during the case study time period meant that this political will was lacking in Iowa and Minnesota, where the programs were suspended for lack of funds. Similarly, the Oklahoma and Vermont programs, while not suspended, were in a holding pattern. Oklahoma was able to sustain its program by funding it through a portion of a planned rate increase and new revenue generated by a provider tax, as did Utah. The commitment to a long-term funding strategy in these two states was critical and reinforced provider support. Thus, to minimize the risk of provider opposition and to promote long-term sustainability, states should consider using “new” dollars to fund P4P rather than reallocating existing dollars. Developing a long-term funding strategy is critical.

Any reimbursement system that relies on performance measures to determine rewards must develop a transparent strategy to determining those rewards and generate widespread agreement about the measures and methods used. Table 1 shows how the case study states assessed performance: the most common quality dimensions included were staffing (Iowa, Minnesota, Vermont); consumer satisfaction (Iowa, Minnesota, Oklahoma, Utah, Vermont); clinical quality indicators (Iowa, Minnesota, Oklahoma); and person-centered care, culture change, or quality of life (Iowa, Minnesota, Oklahoma, Utah). Less frequently considered dimensions included efficiency (Iowa, Vermont), access (Iowa, Oklahoma), employee satisfaction (Oklahoma), and presence of a quality improvement plan (Utah). The incorporation of different quality measures in most case study states suggests that use of a range of measures is preferred because it spreads the risk of poor performance across multiple dimensions, thereby minimizing the chances of unduly penalizing providers that perform well overall while reducing the chances that providers might gain rewards by focusing on a single quality dimension to the exclusion of others; it also minimizes the risk of gaming or outright fraud.

Another key element to P4P success is the level of provider participation in program development and refinement. P4P is a politically charged issue, given the general distrust between the provider community and state officials in most states (Miller 2008, 2010; Sparer 2003). Most providers regard P4P adoption with skepticism, although some elements of the NH industry may oppose P4P more than others, depending on their economic interests. To promote buy-in, all case study states involved key stakeholders in P4P program development. In most cases, the likelihood of consensus around program attributes was improved by establishing a taskforce or workgroup comprising representatives from the NH industry, consumer advocates, and the state rate-setting and survey and certification offices (among other groups). So, too, was discussing with stakeholders the philosophical underpinnings of the planned P4P program. This finding suggests that key to gaining stakeholder acceptance and therefore the chances of program success is engaging industry and other stakeholder representatives early on and throughout the P4P design and adoption process.

Iowa, Minnesota, and Oklahoma focus on rewarding absolute performance. Interestingly, these states do so by pooling NH’s quality scores across multiple domains into a total composite score...
that is used to determine a facility’s reward. This approach contrasts with rewarding facilities separately on the basis of scores achieved in each domain. In addition, the states that employ a composite score also use industry-wide benchmarks as opposed to relative performance against other facilities, thereby decreasing the administrative burden of using the measures. The composite score approach is generally preferred because it evaluates and allocates rewards on the basis of each facility’s actual performance while simplifying the calculation and reporting of program outcomes compared to systems that do so separately for each individual measure.

Although rewarding absolute performance may be desirable to some industry segments, more providers might be willing to participate in P4P if it rewar ded quality improvement (or factors tied to improvement) instead. Rewarding absolute performance generally benefits already high-performing providers and may discourage participation by low- or middle-tier performers that are unlikely to reach absolute-style benchmarks. It is arguable that improvement is most critical among these lower performers, which should be the target of any new dollars invested in NH quality. In addition to achieving secular improvement on quality measures over time, participation by lower-performing providers could be incentivized by rewarding facilities that meet certain procedural hurdles. For example, Utah rewarded facilities adopting a quality improvement plan, while Vermont did so, in part, based on participation in its Gold Star Employer Program. Both states also set a basic bar for participation based on annual inspections. The advantage of the procedural approach is that it is relatively straightforward and simple; the disadvantage is that it rewards facilities independent of outcomes and thus may be most appropriate for a program’s earlier stages. Still, to incentivize low- and middle-level performers while also rewarding good performers, states could reward relative improvement and procedural advances, as well as absolute performance.

Attention to the administrative burden imposed by P4P programs is another factor that has an impact on provider engagement. Especially in the current economic environment, providers are wary of anything that may increase reporting and workload burdens. Thus, if P4P implementation makes an existing reimbursement system more cumbersome, buy-in from key stakeholder groups will be difficult to secure, suggesting that minimizing the administrative burdens associated with the adoption of P4P is particularly important, including permitting providers to use existing data systems to report performance where appropriate.

In general, the P4P programs examined embraced this approach, aiming to maximize the use of existing data sources. Still, all but Vermont made it a priority to collect consumer satisfaction and quality of life information, a potentially burdensome undertaking. States took a variety of approaches to incorporating such measures, including mandating use of a standard, independently administered tool among all facilities (Minnesota, Oklahoma) and permitting facilities to use their own tools (Iowa, Utah). Indeed, there are a variety of ways to incorporate consumer satisfaction measures; these vary in terms of their flexibility and their cost implications for facilities. It is likely that state subsidization of the additional data collection costs, say, by contracting with a vendor, would likely reduce provider resistance while promoting systematic compilation and assessment of the data recorded.

Biases built into the reward calculations will also affect the perceived fairness of and engagement in, P4P programs. Two states—Utah and Vermont—rewarded facilities using lump-sum payments distributed on the basis of Medicaid patients served. Three states— Iowa, Minnesota, and Oklahoma—rewarded facilities as a percentage of the per diem rate, whereby the size of the incentive payments received depends, in part, on facility costs. None of the states rewarded facilities on the basis of a fixed-dollar add-on or bonus, whereby the incentive payments were the same across all NHs, regardless of their base levels of payment. The fixed per diem add-on approach is preferred because it is dependent exclusively on the basis of facility performance rather than on how much money facilities happen to be paid.

Another critical factor that may promote successful implementation of P4P is taking an incremental approach to program development: all case study states but Vermont reported modifying their P4P programs over time. Doing so, for example, permitted state officials to expand the number of quality dimensions and measures as providers gained experience with the program. It also permitted state officials to refine program measures to take advantage of and encourage innovations in quality assessment. States should build in flexibility to provide state officials with opportunities to adjust P4P programs, thereby enabling both facilities and the state to take advantage of new knowledge and experience to improve program effectiveness.

Three states—Iowa, Oklahoma, and Minnesota—implemented a consumer report card along with P4P. Initially, making a report card available based on the P4P measures prompted facilities to put the necessary data-reporting processes into place. Subsequently, it provided facilities with the data needed to adjust their behavior while also tracking progress toward program goals. It also permitted state officials to assess and, when necessary, adjust the program over time to ensure the measures and incentives used were still appropriate and relevant. Phasing in P4P slowly, beginning with performance measurement, followed by public report cards and, finally, introducing P4P incentives, maximizes opportunities for stakeholder acceptance and learning. Moreover, an emphasis on measurement ensures that facilities have access to important performance data; provides richer data for report cards and state-level quality monitoring; and, where funding for P4P is available, provides a fair basis for distributing incentive payments.

Limitations
There are several limitations worth noting. First, we studied P4P in only five states. Consequently, our findings may not apply to states facing substantially different economic, political, and programmatic circumstances. Second, the particular interview subjects selected may have biased the results. Because there was no sampling frame, potentially knowledgeable individuals may have been excluded. While we are confident that we spoke with the key P4P program implementers in the five states studied, our impressions may have been dependent, in part, on the specific individuals interviewed. Finally, the interview protocol was designed to acquire detailed information on the particular topics addressed. Although providing a rich source of data, doing so sacrificed breadth for depth. Future research should build on the results reported here by exploring additional issues related to P4P program development and administration that may be of interest.
Conclusion

NH P4P programs can be designed in a variety of different ways, and there is no single design that, as yet, has been demonstrated to achieve best outcomes. More than any other case study state, Oklahoma, with its moderately complex and active P4P system and 98 percent provider participation rate, met the criteria identified as critical to successful P4P adoption, which include active and continued industry engagement combined with a slow phase-in and ongoing program modification. Oklahoma’s program also relies on additional program funding, a range of quality measures (including one focused on person-centered care), existing data sources, a composite quality score, public reporting, and standardized, independent assessment of resident satisfaction.

To date, P4P has been applied primarily to hospitals, physician practices, and individual health care providers, but it has not yielded particularly promising results—uncovering issues associated with measurement, provider participation, behavior alteration, and incentive structure (Arling, Job, and Cooke 2009; Rosenthal and Frank 2006). However, there is good reason to believe that the NH sector may be particularly conducive to the P4P approach because NHs meet several previously identified success factors (Rosenthal and Frank 2006). In contrast to hospitals, for example, care is delivered through a fairly straightforward organizational structure, with fewer providers and transitions involved. Moreover, care is delivered in a single setting and under controlled conditions (Arling, Job, and Cooke 2009) and, unlike in hospital and outpatient settings, a single payer (Medicaid) dominates (Werner, Konetzka, and Liang 2010). Other NH characteristics, however, may minimize the potential of P4P: complex patient populations, comparatively low payment rates, conflicting payer incentives, and ongoing challenges measuring and monitoring quality (Arling, Job, and Cooke 2009; Mollot, Rudder, and Samji 2008; Mor, Miller, and Clark 2010; Werner, Konetzka, and Liang 2010).

Thus, further evidence is needed about NH P4P’s ability to have an impact on quality. To date, few studies have rigorously evaluated its results—unsurprising, given its early stage of development. However, one preliminary study (Buika and Skira 2012) has found promising results, suggesting that rewarding clinical outcomes and focusing on culture change (as Oklahoma does) yields measurable quality improvements. Until stronger evidence is available, early information about success factors for NH P4P will be useful not only to those implementing such programs but also for those thinking about how government can structure financial incentives to ensure value for money more generally. Moreover, it is worth noting the positive spillovers of P4P: its focus on developing the infrastructure and processes necessary to make comparable quality information more widely available and the emphasis on patient-centered care and resident satisfaction.

Acknowledgments

The authors would like to acknowledge Catherine Jo, Rachel Dolin, and Jon Polit for excellent assistance in writing, note taking, and analysis. They would also like to thank Judy Goldberg-Dey, David Grabowski, Pedro Gozalo, and Vincent Mor for their invaluable insights. This work was supported, in part, by the Washington Department of Social and Health Services (Contract no. 1034-12027).

References


Mor, Vincent, Edward A. Miller, and Melissa Clark. 2010. The Taste for Regulation in Long-Term Care. Supplement 4, Medical Care Research and Review 67: 38–64.


