

Sustainability of Quality Improvement Following Removal of Pay-for-Performance Incentives

Justin K. Benzer, PhD^{1,2}, Gary J. Young, PhD^{1,3}, James F. Burgess Jr PhD^{1,2}, Errol Baker, PhD^{1,2}, David C. Mohr, PhD^{1,2}, Martin P. Charns, DBA^{1,2}, and Peter J. Kaboli, MD, MS^{4,5}

¹Center for Organization, Leadership, and Management Research (COLMR) at the VA Boston Healthcare System (152 M), Boston, MA, USA;

²Boston University School of Public Health, Boston, MA, USA; ³Northeastern University Center for Health Policy and Healthcare Research, Boston, MA, USA;

⁴Comprehensive Access and Delivery Research and Evaluation (CADRE) Center at the Iowa City VA Healthcare System, Iowa City, IA, USA;

⁵Department of Internal Medicine, University of Iowa Carver College of Medicine, Iowa City, IA, USA.

BACKGROUND: Although pay-for-performance (P4P) has become a central strategy for improving quality in US healthcare, questions persist about the effectiveness of these programs. A key question is whether quality improvement that occurs as a result of P4P programs is sustainable, particularly if incentives are removed.

OBJECTIVE: To investigate sustainability of performance levels following removal of performance-based incentives.

DESIGN, SETTING, AND PARTICIPANTS: Observational cohort study that capitalized on a P4P program within the Veterans Health Administration (VA) that included adoption and subsequent removal of performance-based incentives for selected inpatient quality measures. The study sample comprised 128 acute care VA hospitals where performance was assessed between 2004 and 2010.

INTERVENTION: VA system managers set annual performance goals in consultation with clinical leaders, and report performance scores to medical centers on a quarterly basis. These scores inform performance-based incentives for facilities and their managers. Bonuses are distributed based on the attainment of these performance goals.

MEASUREMENTS: Seven quality of care measures for acute coronary syndrome, heart failure, and pneumonia linked to performance-based incentives.

RESULTS: Significant improvements in performance were observed for six of seven quality of care measures following adoption of performance-based incentives and were maintained up to the removal of the incentive; subsequently, the observed performance levels were sustained.

LIMITATIONS: This is a quasi-experimental study without a comparison group; causal conclusions are limited.

CONCLUSION: The maintenance of performance levels after removal of a performance-based incentive has implications for the implementation of Medicare's value-based

purchasing initiative and other P4P programs. Additional research is needed to better understand human and system-level factors that mediate sustainability of performance-based incentives.

KEY WORDS: inpatients; physician incentive plans; quality improvement; quality indicators; reimbursement; incentive; salaries; fringe benefits.

J Gen Intern Med 29(1):127–32

DOI: 10.1007/s11606-013-2572-4

© Society of General Internal Medicine 2013

INTRODUCTION

Pay-for-Performance (P4P) has become a central strategy for improving the quality of health care in the US, Canada, and the UK, and such programs have become widely adopted among private and public health insurance programs over the last decade.^{1–3} Of particular note, the Patient Protection and Affordable Care Act (ACA) mandates the adoption of P4P (i.e., value-based purchasing) for hospitals and physicians participating in the Medicare program. Although P4P programs vary markedly in their design, two common features are: 1) defined performance goals for selected quality measures, and 2) associated financial incentives that can be targeted to institutions, individuals or both.

Despite the growing prevalence of P4P programs, numerous questions persist about their effectiveness in improving quality of care, particularly about sustainability once the incentive is removed. While some studies of P4P demonstrate positive improvements in quality of care,⁴ other studies report disappointing results, as documented in several reviews of the literature.⁵ Moreover, among studies that do indicate improvements in quality measures, almost no attention has been paid to whether such improvements are sustainable over time,⁶ especially if the performance goals and incentives are removed.

The sustainability of performance levels is a key consideration, as it may not be desirable or practical to

Electronic supplementary material The online version of this article (doi:10.1007/s11606-013-2572-4) contains supplementary material, which is available to authorized users.

Received January 11, 2013

Revised May 17, 2013

Accepted July 18, 2013

Published online August 9, 2013

maintain performance-based incentives indefinitely. For example, removal of performance-based incentives may seem warranted when performance for a measure rises to the upper end of the performance scale (i.e., topped out) (e.g., aspirin in acute myocardial infarction), or is otherwise at a level that is not likely to be exceeded in the current clinical environment. In the design of Medicare's hospital value-based purchasing (VBP) program, considerable debate regarding topped out measures occurred, resulting in a decision to exclude process measures that have attained this status.⁷ Another reason for removing performance-based incentives is to expand the reach of a P4P program to a new range of clinical conditions or areas of focus. Performance-based incentives may be routinely removed from some measures for specified time periods and assigned to other measures to limit the total number of performance measures under evaluation at any point in time.

Despite the importance of sustainability for current healthcare policy, there is limited research on the sustainability of performance levels.⁸⁻¹⁰ There has been no research on the removal of incentives for inpatient medicine quality measures such as those included in Medicare's VBP. A multi-year P4P initiative within the Veterans Health Administration (VA) that included adoption and removal of performance-based incentives for selected quality measures provides the opportunity to conduct such research using a quasi-experimental study design. The objective of this study was to evaluate the empirical support for a hypothesis that performance gains realized during a P4P program would decrease after the removal of the performance-based incentives.

METHODS

Setting

The Veterans Health Administration (VA), US Department of Veterans Affairs, is the largest integrated healthcare provider in the US, with more than 8.5 million enrollees in 2012. VA medical centers are organized into regional networks managed by a network director, use a common electronic medical record, and a P4P quality measurement and reporting system.¹¹ For purposes of P4P, VA's central office sets performance goals in consultation with clinical leaders and reported performance scores to medical centers quarterly. As such, this system-level intervention entailed both public reporting and financial incentives. With respect to public reporting, performance data were available to both clinicians and managers quarterly, and were also included in publicly available annual reports. Performance bonuses were distributed, based on the attainment of performance goals, to both regional network and facility-level senior managers, who, in turn, had discretion to distribute bonus

payments to front-line clinicians and other employees. The unit of analysis for the study was the VA medical center (N=128).

Performance Measures

Since 2004, VA has tracked over 30 performance measures relevant to acute coronary syndrome (ACS), heart failure (HF), and pneumonia (PNU). Following The Joint Commission standards, sampling has been conducted for all patients with these three conditions. Performance measures were developed based on published scientific evidence and established clinical guidelines. The performance measurement system for these quality measures is standardized and includes specified data collection protocols (Appendix A; available online). Performance measure guideline adherence is measured through VA's External Peer Review Program, an independent chart review of randomly selected patients who meet specified inclusion and exclusion criteria. Goals for each performance measure are set annually by the VA central office, and incentives are awarded based on achievement of those goals. Performance goals have also been raised for some measures as the mean performance level has risen over time. In addition, for seven of these measures, the performance-based incentives were removed between 2007 and 2009, but continued to be measured and reported for at least a year. Although no explicit criteria existed for the removal of incentives, high performance level was likely a factor. For six of the seven performance measures, mean performance was over 90 % prior to removal of the incentives. We focused on these seven quality measures, and for each indicated the percent of hospitalized patients who satisfied the inclusion/exclusion criteria and received guideline concordant care.

Acute Coronary Syndrome (ACS)

- **Cardiology Involvement:** High or moderate-high risk patients with cardiology involvement within 24 hours of arrival, or if acute myocardial infarction (AMI) during inpatient stay, within 24 hours of initial electrocardiogram (ECG) or first positive troponin, whichever is earlier.
- **Troponin Returned:** First troponin result returned within 60 min of order.
- **Diagnostic Catheterization:** High or moderate-high risk patients who received a diagnostic catheterization prior to discharge.

Heart Failure (HF)

- **ACE-I or ARB:** For patients with ejection fraction less than 40 %, presence of an angiotensin-converting

enzyme inhibitor (ACE-I) or angiotensin receptor blocker (ARB) prior to admission (i.e., a continuous care metric targeting the quality of care in the outpatient setting).

- **Weight Monitoring:** Documentation of instruction for monitoring weight prior to admission (i.e., a continuous care metric targeting the quality of care in the outpatient setting).

Pneumonia (PNU)

- **Timely Antibiotic:** Initial antibiotic dose administered no earlier than 15 min prior to or no later than 240 min following hospital arrival.
- **Pneumococcal Immunization:** Receipt of pneumococcal immunization prior to admission (i.e., a continuous care metric targeting the quality of care in the outpatient setting).

Patient Sample

313,600 VA patient records were peer reviewed between FY2004 and FY2010 across the seven measures. Sample sizes for a single year ranged from 3,588 for HF: ACE-I or ARB in FY2010 to 13,777 for HF: Weight monitoring in FY2009. For each performance measure, the average numbers of patients sampled per facility per quarter are reported in Table 1.

Statistical Analyses

Quarterly performance data were obtained from FY2004 to FY2010 from VA administrative data. Each measure was a percentage score representing the number of patients meeting the performance criteria divided by the total number of eligible patients. Medical centers served as their own controls in analyses. Missing data were an issue for

between four to nine study sites. These sites had more than 50 % missing data, whereas all other sites averaged 1 % missing data. Sensitivity analyses with and without the high missing data sites demonstrated that conclusions would not differ based on the decision to include or exclude sites. The sites with substantial missing data were excluded. For the remaining sites, we imputed missing data using maximum likelihood estimation during analyses. Latent growth models implemented with MPLUS Version 5.2 were used to estimate slopes across years. A piecewise latent growth model was used for each performance measure to estimate an intercept and slopes for each year in the model, accounting for autocorrelations across time periods (Appendix B; available online). A significant slope indicates that the rate of change is significantly different from zero. For example, PNU: Timely Antibiotic was measured from FY2005 to FY2009 so analyses estimate the slopes for each of the 5 years. This model permits evaluation of changes in performance between years where the performance goal changed, years where the performance goal remained constant, and years during which the performance goal was removed. A significant negative slope in the year following incentive removal indicates that performance was not sustained. Power is a concern because the absence of a significant negative slope will be interpreted as sustained performance. Thus, we performed power calculations. Analyses had 86 % power to detect whether the slope was at least -2 % in the year following incentive removal.

RESULTS

Table 1 presents the introduction and removal of the performance-based incentives for each performance measure. Only two measures (PNU: Timely Antibiotic and ACS: Diagnostic Catheterization) had a true baseline period where reporting for the measure occurred before the adoption of performance-based incentives. Performance-

Table 1. Performance Goals for Each Quality Measure (FY2004–FY2010)

Measure	N	FY2004	FY2005	FY2006	FY2007	FY2008	FY2009	FY2010
Acute Coronary Syndrome								
Cardiology Involvement	17	N/A	82 %	87 %	87 %	91 %	none	none
Troponin Returned	15	N/A	75 %	85 %	89 %	95 %	none	none
Diagnostic Catheterization	15	Baseline	90 %	90 %	none	N/A	N/A	N/A
Heart Failure								
ACE-I or ARB	11	N/A	90 %	90 %	90 %	none	none	none
Weight Monitoring	25	90 %	90 %	90 %	90 %	none	none	N/A
Pneumonia								
Timely Antibiotic	18	N/A	Baseline	73 %	80 %	none	none	N/A
Pneumococcal Immunization	20	85 %	85 %	85 %	85 %	none	none	N/A

Percentage scores indicate the performance standard for that year. N/A indicates that the measure was not assessed that year. "Baseline" indicates the measure was collected, but was not incentivized for that year. "None" indicates that the performance-based incentive was removed. N= Average number of patients per hospital per quarter

based incentives were removed between 2–4 years following adoption.

Rates of change for each measure are shown by quarter in Fig. 1 for the latent growth models. The three ACS measures are displayed together in Fig. 1a, the two HF measures in Figure 1b, and the two PNU measures in Fig. 1c. Each line represents the overall trend for a single year regarding the rate of change after removal of the performance-based incentives, as indicated by the arrows in Figure 1.

The overall mean score changes for the period where performance-based incentives were adopted and the period where performance-based incentives were removed are summarized in Table 2. Prior to the removal of incentives, we found that performance significantly improved for six of the seven measures. The most dramatic improvement occurred with the PNU: Timely Antibiotic measure, where performance improved from 64 % to 82 % in 2 years following the adoption of performance-based incentives. The only measure that did not demonstrate significant improvement was the heart failure: ACE-I measure.

Results did not support the hypothesis that performance decreased after incentives were removed. Six of the seven measures did not demonstrate a significant slope in the year following incentive removal. The seventh measure, weight monitoring, demonstrated a significant positive slope in the year following incentive removal. However, a significant negative slope was observed in the following year and a non-significant slope in the third post-removal year. Given that the design provides adequate power to detect changes in performance, results indicate that performance was sustained for all measures following removal of incentives.

DISCUSSION

In this observational cohort study evaluating P4P over 7 years in 128 VA hospitals, we found evidence of improvement in performance measures following the adoption of performance-based incentives, and that after removal of the incentives, performance neither further improved nor deteriorated. As the US makes a substantial investment in P4P, both financially and intellectually, it is imperative that researchers capitalize on opportunities to learn about the potential effectiveness of such programs on quality of care. Current national policy discussions involve both use of quality measures and choices regarding when to retire measures. Our findings have important implications for Medicare's value-based purchasing program as we focused on the same types of hospital inpatient measures included in the Medicare program.

Our study contributes to a growing literature on P4P for which there is a lack of consistent evidence regarding the effectiveness of such programs. The mixed findings in the

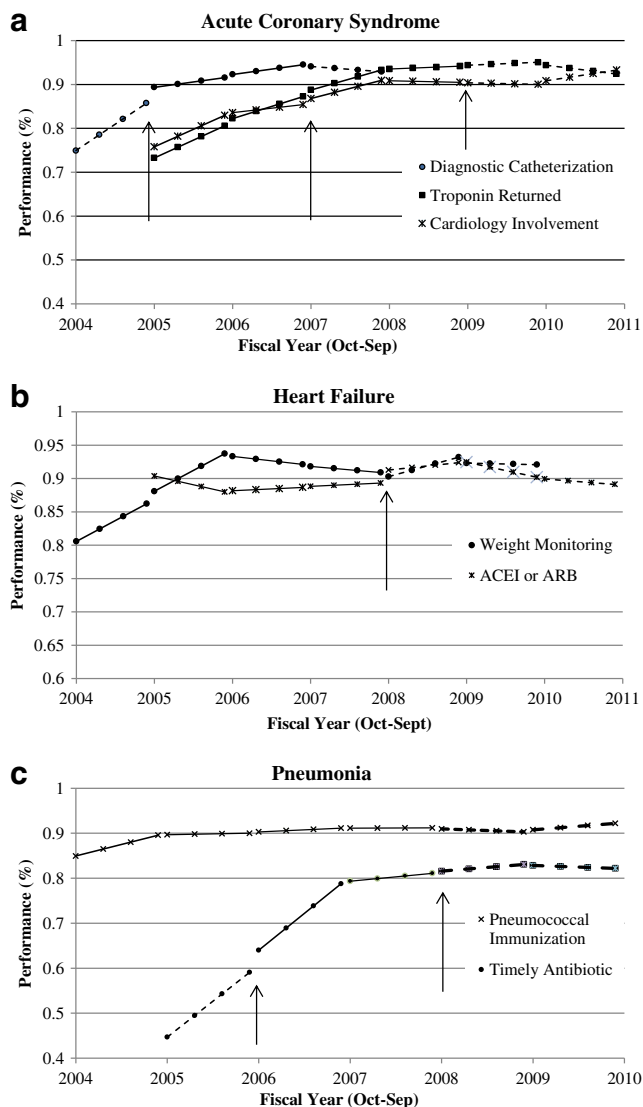


Figure 1. Graph of latent growth model analyses for seven performance measures. Dependent variable is the number of patients who receive guideline-adherent care divided by the number of eligible patients. Trend lines are estimated for each year to demonstrate how the trend changes over time. Arrows indicate the point at which incentives are either introduced or removed. Dashed lines indicate periods in which performance-based incentives were removed. Significant slopes are indicated by larger point size lines.

literature suggest that the effectiveness of P4P likely depends on contextual factors that researchers have yet to fully explicate with conceptual frameworks and empirical testing. In this vein, the particular implementation of P4P in the VA and the nature of the VA system may have improved the likelihood of performance sustainability. As noted, performance-based incentives in the VA are awarded to facilities and their managers, who decide whether and how to distribute them to clinicians. This type of incentive arrangement is similar to those established by Medicare and private health plans for purposes of contracting with

Table 2. Overall Change in Performance Measures from Initial to Final Measurement Among VA Facilities

Measure	Adoption of performance-based incentives				Removal of performance-based incentives			
	First quarter	Last quarter	Diff	P value	First quarter	Last quarter	Diff	P value
Acute Coronary Syndrome								
Cardiology Involvement	74 (24)	94 (13)	20	< 0.001	90 (16)	91 (22)	1	0.93
Troponin Returned	74 (29)	96 (8)	22	< 0.001	94 (13)	92 (18)	-2	0.35
Diagnostic Catheterization	91 (21)	95 (12)	4	0.03	94 (17)	93 (19)	-1	0.26
Heart Failure								
ACE-I or ARB	89 (23)	92 (15)	3	0.26	90 (17)	89 (19)	-1	0.50
Weight Monitoring	80 (23)	92 (13)	12	< 0.001	91 (11)	90 (16)	-1	0.66
Pneumonia								
Timely Antibiotic	64 (21)	82 (19)	18	< 0.001	81 (13)	85 (17)	4	0.06
Pneumococcal Immunization	85 (14)	92 (13)	7	< 0.001	89 (10)	92 (13)	3	0.05

Values represent mean percentage scores with standard deviation of the percentage in parentheses; Diff indicates the absolute difference in mean percentage scores; p values indicate whether mean scores significantly changed overall across each measurement period; paired t-tests (two-tailed) conducted to provide an overall summary of the change, results are consistent with latent growth models

accountable care organizations (ACO). In most such programs, ACOs are also the unit of accountability for performance-based incentives, and ACO senior managers have discretion as to whether and how incentive payments are distributed to front-line clinicians.

Limitations to our study include a relatively small number of performance measures for investigation, a brief post-incentive removal period of between 1 and 3 years, and the absence of a comparison group. The current paper indicates that once hospitals achieve a high level of performance, it may be possible to sustain that high performance after incentives are removed. However, this study does not indicate how performance may change if incentives are removed before a high level of performance is reached. Further, the absence of a comparison group limited our ability to isolate the effects of the performance-based incentives from other factors, such as a secular trend or public reporting that may have contributed to changes in the performance measures during the study period. In this vein, some evidence exists indicating that public reporting of performance measures alone can lead to performance improvements in hospitals.¹²⁻¹⁵ Although a study that compared the performance effects of combining financial incentives and public reporting to public reporting alone found that incentives raise performance levels above those obtained from just public reporting, the added increase was quite modest.¹⁶ As such, it is possible that the VA would have experienced similar patterns, though perhaps not at identical levels, of performance improvement and sustainability from reporting the performance of its facilities on the selected measures even without offering performance-based incentives.

Future research, perhaps using mixed methods, should address how incentives are most effectively implemented and how incentives may have unintended positive or negative effects in complex health care delivery systems.¹⁷ In general, sustainability of quality improvements may depend on changes in clinical systems that

do not consistently add to the workload of busy clinical staff. However, sustainability may also vary depending on who receives the incentives. Incentives targeted toward physicians may cause them to focus their efforts on patient-level clinical issues related to the performance measures, whereas incentives targeted toward managers may cause them to focus their efforts on system issues. Increased effort by clinical staff may be needed to improve performance initially, but changes in clinical systems may be required for the improvements to be sustainable.

In summary, this study found that performance improvements that occurred in VA medical centers for three common conditions (i.e., ACS, HF, and PNU) were sustained for up to 3 years after performance-based incentives were removed. These sustained improvements may represent adoption of new standards of care that were driven by P4P and, once adopted, the incentive was no longer necessary to maintain a high level of quality. If these findings can be reproduced, they could help guide the adoption and discontinuation of P4P measures.

Acknowledgements: The work reported herein was supported by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service (IIR 08-067-1) and an Investigator Award in Health Policy to Gary Young from the Robert Wood Johnson Foundation. The authors had full access to and take full responsibility for the integrity of the data. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs. The authors would like to thank Terry Duncan for consultation on implementing time series models in MPLUS.

Conflict of Interest: The authors declare that they do not have any conflicts of interest.

Corresponding Author: Justin K. Benzer, PhD; Center for Organization, Leadership, and Management Research (COLMR) at the VA Boston Healthcare System (152 M), 150 South Huntington Avenue, Boston, MA 02860, USA (e-mail: Justin.benzer@va.gov).

REFERENCES

1. **Rosenthal MB, Frank RG, Li Z, Epstein AM.** Early experience with pay-for-performance: from concept to practice. *JAMA.* 2005;294(14):1788–1793.
2. **Millett C, Gray J, Saxena S, Netuveli G, Majeed A.** Impact of a pay-for-performance incentive on support for smoking cessation and on smoking prevalence among people with diabetes. *CMAJ.* 2007;176(12):1705–1710.
3. **Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, et al.** Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med.* 2006;355(4):375–384.
4. **Chung S, Palaniappan LP, Trujillo LM, Rubin HR, Luft HS.** Effect of physician-specific pay-for-performance incentives in a large group practice. *Am J Manag Care.* 2010;16(2):e35–e42.
5. **Christianson JB, Leatherman S, Sutherland K.** Lessons from evaluations of purchaser pay-for-performance programs: a review of the evidence. *Med Care Res Rev.* 2008;65(6 Suppl):5S–35S.
6. **Glasgow JM, Scott-Caziewell JR, Kaboli PJ.** Guiding inpatient quality improvement: a systematic review of lean and six sigma. *Joint Comm J Qual Patient Saf.* 2010;36(12):533–540.
7. Medicare Program; Hospital Inpatient Value-Based Purchasing Program; Final Rule <http://www.gpo.gov/fdsys/pkg/FR-2011-05-06/html/2011-10568.htm> In: Centers for Medicare & Medicaid Services, ed: Department of Health and Human Services; 2011:26489–547.
8. **Hysong SJ, Khan MM, Petersen LA.** Passive monitoring versus active assessment of clinical performance: impact on measured quality of care. *Med Care.* 2011;49(10):883–890.
9. **Glasgow JM, Davies ML, Kaboli PJ.** Findings from a national improvement collaborative: are improvements sustained? *BMJ Qual Saf.* 2012;21(8):663–669.
10. **Lester H, Schmittiel J, Selby J, Fireman B, Campbell S, Lee J, et al.** The impact of removing financial incentives from clinical quality indicators: longitudinal analysis of four Kaiser Permanente indicators. *BMJ.* 2010;340:c1898.
11. **Kizer KW, Dudley RA.** Extreme makeover: transformation of the veterans health care system. *Annu Rev Public Health.* 2009;30:313–339.
12. **Hannan EL, Kilburn H Jr, Racz M, Shields E, Chassin MR.** Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA.* 1994;271(10):761–766.
13. **Werner RM, Bradlow ET.** Public reporting on hospital process improvements is linked to better patient outcomes. *Health Aff (Millwood).* 2010;29(7):1319–1324.
14. **Hibbard JH, Stockard J, Tusler M.** Hospital performance reports: impact on quality, market share, and reputation. *Health Aff (Millwood).* 2005;24(4):1150–1160.
15. **Smith MA, Wright A, Queram C, Lamb GC.** Public reporting helped drive quality improvement in outpatient diabetes care among Wisconsin physician groups. *Health Aff (Millwood).* 2012;31(3):570–577.
16. **Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, et al.** Public reporting and pay for performance in hospital quality improvement. *N Engl J Med.* 2007;356(5):486–496.
17. **Bokhour BG, Burgess JF Jr, Hook JM, White B, Berlowitz D, Guldin MR, et al.** Incentive implementation in physician practices: a qualitative study of practice executive perspectives on pay for performance. *Med Care Res Rev.* 2006;63(1 Suppl):73S–95S.