Effects of Physician-targeted Pay for Performance on Use of Spontaneous Breathing Trials in Mechanically Ventilated Patients

Ian J. Barbash^{1,2}, Francis Pike², Scott R. Gunn², Christopher W. Seymour², and Jeremy M. Kahn^{1,2,3}

¹Division of Pulmonary, Allergy, and Critical Care Medicine and ²Clinical Research, Investigation, and Systems Modeling of Acute Illness Center, Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; and ³Department of Health Policy and Management, University of Pittsburgh Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania

Abstract

Rationale: Pay for performance is an increasingly common quality improvement strategy despite the absence of robust supporting evidence.

Objectives: To determine the impact of a financial incentive program rewarding physicians for the completion of daily spontaneous breathing trials (SBTs) in three academic hospitals.

Methods: We compared data from mechanically ventilated patients from 6 months before to 2 years after introduction of a financial incentive program that provided annual payments to critical care physicians contingent on unit-level SBT completion rates. We used Poisson regression to compare the frequency of days on which SBTs were completed among eligible patients and days on which patients were excluded from SBT eligibility among all mechanically ventilated patients. We used multivariate regression to compare risk-adjusted duration of mechanical ventilation and in-hospital mortality.

Measurements and Main Results: The cohort included 7,291 mechanically ventilated patients with 75,621 ventilator days. Baseline daily SBT rates were 96.8% (hospital A), 16.4% (hospital B), and 74.7% (hospital C). In hospital A, with the best baseline performance, there was no change in SBT rates, exclusion rates, or duration of mechanical ventilation across time periods. In hospitals B and C, with lower SBT completion rates at baseline, there was an increase in daily SBT completion rates and a concomitant increase in exclusions from

eligibility. Duration of mechanical ventilation decreased in hospital C but not in hospital B. Mortality was unchanged for all hospitals.

Conclusions: In hospitals with low baseline SBT completion, physician-targeted financial incentives were associated with increased SBT rates driven in part by increased exclusion rates, without consistent improvements in outcome.

Keywords: mechanical ventilation; health care quality; critical care; pay for performance

At a Glance Commentary

Scientific Knowledge on the Subject: Although pay for performance is an increasingly common quality improvement strategy, evidence of its impact on clinical processes and outcomes is mixed and few data are specific to critical care.

What This Study Adds to the Field: The implementation of a physician-targeted financial incentive for the completion of spontaneous breathing trials in mechanically ventilated patients was associated with an increase in trial completion rates in low-performing intensive care units. However, these results were driven in part by increased exclusion rates and were not associated with consistent improvements in patient outcomes, suggesting that the program may have changed documentation patterns rather than improved patient care.

(Received in original form July 25, 2016; accepted in final form December 6, 2016)

Supported by U.S. National Institutes of Health grants T32HL007563 and F32HL132461 (I.J.B.) and grant K24HL133444 (J.M.K.).

Author Contributions: Study concept and design: all authors; data analysis: I.J.B., F.P., and J.M.K.; drafting and revision of manuscript: I.J.B., C.W.S., S.R.G., and J.M.K.; critical appraisal and approval for submission: all authors.

Correspondence and requests for reprints should be addressed to Jeremy M. Kahn, M.D., M.S., Critical Care and Health Policy and Management, University of Pittsburgh, Scaife Hall, Room 602-B, 3550 Terrace Street, Pittsburgh, PA 15213. E-mail: jeremykahn@pitt.edu

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Am J Respir Crit Care Med Vol 196, Iss 1, pp 56–63, Jul 1, 2017

Copyright © 2017 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201607-1505OC on December 12, 2016

Internet address: www.atsjournals.org

For patients receiving invasive mechanical ventilation, the regular performance of a daily spontaneous breathing trial (SBT) is associated with improved clinical outcomes, including shorter duration of mechanical ventilation and length of stay in the intensive care unit (ICU) (1, 2). On the basis of these studies, multiple professional society guidelines recommend daily SBTs to improve patient-centered outcomes in the ICU (3, 4). However, despite this evidence many mechanically ventilated patients do not receive daily SBTs (5). Consequently, there is an opportunity to improve patient outcomes by increasing adherence to this evidence-based care practice.

Pay for performance (P4P) is one strategy to change provider behavior and speed translation of evidence into practice (6). Traditional payment models reimburse physicians for the volume and complexity of care they provide. In contrast, under P4P providers are also reimbursed based on the quality of care they provide. P4P programs are increasingly common in hospital medicine (7–11). However, studies show inconsistent results, with many programs failing to improve the quality of care (8, 11). One possible explanation for these results is that many inpatient-based P4P programs are directed to hospitals rather than to the clinicians who are directly responsible for implementing evidence-based care (6). Thus, an alternative to hospital-level P4P programs is programs that target individual clinicians, thereby creating a more direct incentive for quality improvement. However, the impact of physician-targeted P4P on critical care processes and outcomes is unknown.

To address this knowledge gap, we evaluated the impact of a quality improvement program based on a physician-targeted financial incentive for the performance of daily SBTs in multiple ICUs in three academic hospitals of an integrated health system. We examined the impact of the program both on processes of care, that is, SBT completion rates on eligible ventilator days and documentation of exclusions of eligibility on all ventilator days, and clinical outcomes, that is, duration of mechanical ventilation (DMV), ICU length of stay, and mortality. Some of the results of this study have been previously reported in the form of an abstract (12).

Methods

Study Design and Setting

We performed a retrospective, before-and-after interrupted time series analysis of a quality improvement intervention at three UPMC academic hospitals; UPMC is a 14-hospital integrated health care delivery system in western Pennsylvania. The three study hospitals are termed hospitals A, B, and C. Hospital A contains 8 subspecialty ICUs and 132 ICU beds. Hospital B contains a single mixed medical–surgical ICU with 14 ICU beds. Hospital C contains 3 subspecialty ICUs with 49 ICU beds. Each ICU is managed by a unique physician and nursing director, who together are responsible for implementing quality improvement activities and who receive regular reports on processes of care and clinical outcomes for patients admitted to their ICUs. All ICUs operate under a closed or mandatory intensivist consult staffing model with daily interdisciplinary rounds, overnight in-house intensivist staffing, and regular participation of trainees in critical care.

Data Sources

We obtained ICU-level data on processes of care from an existing quality improvement data registry. These data included counts of the number of ventilator days, number of ventilator days on which the mechanically ventilated patient was deemed eligible for an SBT by the medical team, and number of eligible ventilator days on which the ventilated patient received an SBT. We obtained patient-level clinical data from an existing ICU patient registry derived from the UPMC electronic health record (Cerner PowerChart; Cerner Corporation, Kansas City, MO). These data included demographic information, physiological variables for risk adjustment, documented reasons for SBT eligibility exclusion, and clinical outcomes including in-hospital mortality, ICU length of stay, and DMV.

Intervention

Starting in July 2012, UPMC instituted a physician-targeted P4P program tied to processes of care for ICU patients. The incentive focused on three aspects of care: (1) daily performance of SBTs in eligible mechanically ventilated patients, (2) rate of central line–associated bloodstream infections (CLABSI), and (3) rate of

ventilator-associated pneumonia (VAP). An annual quality bonus equal to approximately 7.5% of each ICU physician's annual base salary was divided evenly among these three performance measures. Each quality indicator was measured at the ICU level. Nursing and physician directors were given quarterly reports on their ICU's performance for the preceding quarter, including performance relative to other ICUs in the three hospitals. Physician compensation depended on whether or not the ICU in which they worked met the performance benchmarks over the course of the year. The cutoffs for the incentive payments were 90% completion among eligible patient days for SBT and a unit-level rate less than the National Nosocomial Infections Surveillance System median for CLABSI and VAP. This analysis examines only SBTs because CLABSI rates were nearly zero across all time periods and VAP surveillance changed to ventilator adverse event surveillance midway through the evaluation time period, precluding the systematic evaluation of these measures over time.

Patients and Variables

We included patients at least 16 years old admitted to ICUs at the three hospitals from January 2012 to June 2014. In the event of multiple ICU admissions for an individual patient, we included only the first ICU admission from the first hospitalization during the study period to preserve independence of observations. We excluded patients for whom the ICU admission source was missing, as this was a key component of risk adjustment.

The exposure variable was the time period of admission. The baseline period ran from January 2012 to June 2012. We started the study period in January 2012 rather than July 2011 to avoid cross-contamination with a prior quality improvement program involving nurse-led screening for quality measures, for which the effects had not yet stabilized (13). The incentive began in July 2012, and we included a 2-month washout period; thus, Postintervention Year 1 ran from September 2012 to June 2013, and Postintervention Year 2 ran from July 2013 to June 2014 (Figure 1).

The outcome variables included both process measures and clinical outcomes. Process measures included the number of ventilator days on which the patients were excluded from SBT eligibility and the

Figure 1. Overview of study periods.

number of eligible ventilator days on which SBTs were performed. Physicians and respiratory therapists determined eligibility prospectively and documented the reason(s) for exclusion in the electronic record. Clinical outcomes included DMV, ICU length of stay, and in-hospital mortality.

Demographic and clinical variables included age, sex, race, ICU admission source (emergency department, postanesthesia care unit, procedure unit, step-down unit, ward, and clinic/other), key comorbidities defined using the *International Classification of Diseases, Ninth Revision, Clinical Modification* diagnosis codes in the manner of Elixhauser (congestive heart failure, cancer, and liver disease [14]), presence of mechanical ventilation on admission to the ICU, and severity of illness as measured using a modified electronic SAPS (Simplified Acute Physiology Score) 3 score, as performed previously (13).

Analysis

To characterize the patient cohort we summarized patient demographic and clinical characteristics across time periods, comparing them by χ^2 or analysis of variance statistics as appropriate.

To test for changes in the completion of SBTs and documentation of SBT eligibility associated with the intervention we used Poisson regression, accounting for the number of eligible ventilator days (for SBT completion) and total number of ventilator days (for SBT eligibility). The exposure variable was time period of ICU admission, with each postintervention year compared with the reference baseline time period. We did not adjust for severity of illness or other patient-level variables in these analyses, because these factors by themselves should not affect the receipt of an SBT if the patient is otherwise eligible. We also analyzed the number of ventilator days on which SBTs were completed among all

patients as an indirect indicator of the contribution of changes in documentation to observed changes in SBT completion. For hospitals containing multiple ICUs, we accounted for within-ICU correlation by including an ICU-fixed effect with robust standard errors (15).

To test for changes in outcomes we used multivariable regression, specifying a logistic model for mortality and linear models for ICU length of stay and DMV. To reduce the influence of long-stay outliers, we censored ICU length of stay and DMV at 30 days. In addition, because both ICU length of stay and DMV data are skewed, we log transformed them before analysis. Consequently, the regression coefficients can be interpreted as a percent change. The

exposure variable was time period of admission relative to the intervention; each postintervention time period was compared with the baseline time period. Risk-adjustment variables included age (as linear splines), comorbidities (as indicator covariates), ICU admission source (as an indicator covariate), and the modified SAPS 3 score (as linear splines). For hospitals that contained multiple ICUs, we again included ICU-fixed effects with robust standard errors.

In testing the effects of the intervention on processes and outcomes we *a priori* planned both a pooled analysis using data from all hospitals and a stratified analysis examining each hospital separately. However, preliminary analyses demonstrated substantial variation in baseline SBT completion rates and significant interaction between the hospital and the effects of the intervention, indicating that there was no single pooled effect. Consequently, we present only the stratified analysis in this report.

To better understand the potential influence of temporal trends, we created graphs with month-specific data, stratified by hospital. We created separate graphs for

Figure 2. Patient selection schema. Adm = admission; Hosp = hospital; ICU = intensive care unit; MV = mechanical ventilation.

SBT completion rates, SBT eligibility exclusion rates, risk-adjusted DMV, risk-adjusted ICU length of stay, and risk-adjusted in-hospital mortality. Adjusted outcomes were obtained by indirect standardization, using the previously described risk-adjustment models.

To better understand the degree to which observed changes in SBT eligibility rates represented better documentation of true ineligibility or the use of nonspecific exclusions of otherwise eligible patients as a way to exclude these patients and increase completion rates, we categorized exclusion criteria into eight groups and created graphs of the frequency of each exclusion criterion within each study period, stratified by hospital. We created a hierarchical schema for exclusion criteria, the full details of which are provided in the online supplement. We hypothesized that an increase in nonspecific, “discretionary” exclusions would be suggestive of the use of exclusion criteria to artificially increase completion rates, a process known as exception reporting.

All analyses were performed with Stata 14.0 (StataCorp, College Station, TX). All tests were two-tailed, and we considered a *P* value less than 0.05 to be statistically significant. The intervention evaluated in this study was designated as a quality improvement by UPMC, as its primary intention was to improve patient care within the UPMC system. Our use of patient records to evaluate the impact of the intervention was reviewed and approved by the UPMC Quality Improvement Committee (project No. 249) under the authority of the University of Pittsburgh Institutional Review Board.

Results

The final cohort included 7,291 unique ICU admissions and 75,621 ventilator days (Figure 2). There were 4,888 patients and 56,725 ventilator days in hospital A; 311 patients and 2,105 ventilator days in hospital B; and 2,101 patients and 16,791 ventilator days in hospital C. Patient characteristics were generally similar across time periods, although over time there were small decreases in the proportion of patients admitted from the emergency department, small changes in the proportion of patients mechanically

ventilated on admission, and small decreases in unadjusted ICU length of stay over time (Table 1). Only 145 patients (2.0%) were long-stay outliers with clinical outcome data censored at 30 days.

SBT Completion and Eligibility Exclusion Rates

Baseline and postintervention SBT completion rates and SBT eligibility exclusion rates are shown in Table 2 and Figure 3. In hospital A, with the highest baseline performance, SBT completion rates remained stable over time (96.8% at baseline, 97.0% in Year 1, and 97.7% in Year 2). On the basis of these results, all physicians in hospital A received the incentive payments in each postintervention year. In hospital B, with the lowest baseline performance, SBT completion rates increased from 16.4 to 26.0% in Year 1 (IRR, 1.59; *P* = 0.02), although not enough for the physicians to qualify for the incentive payments. In the next year, SBT completion rates increased to 93.7% (IRR, 5.72; *P* < 0.001) and the physicians qualified for the incentive

payments. In hospital C, with intermediate baseline performance, SBT completion rates increased from 74.7% at baseline to 83.4% in the first postintervention year (IRR, 1.11; *P* < 0.01) although not enough for the physicians to qualify for the incentive payments. In the next year, SBT completion rates further increased to 91.7% (IRR, 1.22; *P* < 0.001) and the physicians qualified for the incentive payments.

Nearly all significant increases in SBT completion rates were accompanied by significant increases in SBT eligibility exclusions (Table 2 and Figure 3). For example, the major SBT completion rate increase in hospital B occurred with a concurrent increase in eligibility exclusion rates from 39.8% at baseline to 69.1% by the second postintervention year (IRR, 1.74; *P* < 0.001). This pattern occurred in both years for hospital C. When analyzed with respect to total ventilator days regardless of eligibility, SBT completion rates increased over time in hospital B but remained unchanged in hospitals A and C (see Table E1 and Figure E1 in the online supplement), suggesting that

Table 1. Patient Characteristics

	Baseline	Postintervention Year 1	Postintervention Year 2	<i>P</i> Value*
Number of patients	1,625	2,631	3,035	
Age, yr, mean (SD)	61.1 (16.5)	60.7 (16.4)	60.6 (16.1)	0.58
Female, n (%)	768 (47.3)	1,239 (47.1)	1,370 (45.1)	0.23
Race/ethnicity, n (%)				0.21
White	1,358 (83.6)	2,205 (83.8)	2,536 (83.9)	
Black	215 (13.2)	358 (13.6)	426 (14.1)	
Hispanic	2 (0.1)	3 (0.1)	1 (<1)	
Other	18 (1.1)	21 (0.8)	13 (0.4)	
Unknown	32 (2.0)	44 (1.7)	45 (1.5)	
Key comorbidities, n (%)				
CHF	119 (7.3)	199 (7.6)	270 (8.9)	0.09
Cancer	150 (9.2)	239 (9.1)	309 (10.2)	0.33
Liver disease	184 (11.3)	317 (12.0)	349 (11.5)	0.73
ICU admission source, n (%)				<0.01
Clinic/direct/other	201 (12.4)	399 (15.2)	482 (15.9)	
ED	706 (43.4)	1,069 (40.6)	1,227 (40.4)	
Ward	297 (18.3)	498 (18.9)	532 (17.5)	
PACU	112 (6.9)	129 (4.9)	163 (5.4)	
Procedure unit	252 (15.5)	440 (16.7)	542 (17.9)	
Step-down unit	57 (3.5)	96 (3.6)	89 (2.9)	
MV on admission, n (%)	1,283 (79.0)	2,050 (77.9)	2,454 (80.9)	0.02
DMV, d, mean (SD)	4.8 (5.9)	4.6 (5.7)	4.6 (5.8)	0.33
ICU LOS, d, mean (SD)	6.0 (6.8)	5.8 (6.5)	5.4 (6.3)	0.01
In-hospital mortality, n (%)	321 (19.8)	521 (19.8)	623 (20.5)	0.74

Definition of abbreviations: CHF = congestive heart failure; DMV = duration of mechanical ventilation; ED = emergency department; ICU = intensive care unit; LOS = length of stay; MV = mechanical ventilation; PACU = postanesthesia care unit.

*P values are for analysis of variance or χ^2 testing using global comparisons across time, rather than comparing each postintervention year with baseline.

Table 2. Process Measures by Time Period and Hospital

	Baseline	Postintervention Year 1	Postintervention Year 2
Hospital A			
SBT eligibility exclusion			
Ventilator days	12,648	20,292	23,785
Excluded days	7,895	12,037	15,312
Exclusion rate	62.4%	59.3%	64.4%
IRR (95% CI)	1.0 (reference)	0.95 (0.92–0.99)*	1.02 (0.99–1.05)
SBT delivery			
Eligible days	4,753	8,255	8,473
Days received	4,602	8,004	8,280
Rate	96.8%	97.0%	97.7%
IRR (95% CI)	1.0 (reference)	1.00 (0.99–1.01)	1.02 (0.99–1.05)
Hospital B			
SBT eligibility exclusion			
Ventilator days	284	1,001	820
Excluded days	113	352	567
Exclusion rate	39.8%	35.2%	69.1%
IRR (95% CI)	1.0 (reference)	0.88 (0.72–1.09)	1.74 (1.42–2.12) [†]
SBT delivery			
Eligible days	171	650	253
Days received	28	169	237
Rate	16.4%	26.0%	93.7%
IRR (95% CI)	1.0 (reference)	1.59 (1.18–2.14) [†]	5.72 (4.43–7.39) [†]
Hospital C			
SBT eligibility exclusion			
Ventilator days	4,401	6,267	6,123
Excluded days	2,069	3,263	3,589
Exclusion rate	47.0%	52.1%	58.6%
IRR (95% CI)	1.0 (reference)	1.10 (1.02–1.18) [†]	1.24 (1.15–1.33) [†]
SBT delivery			
Eligible days	2,332	3,004	2,534
Days received	1,742	2,504	2,324
Rate	74.7%	83.4%	91.7%
IRR (95% CI)	1.0 (reference)	1.11 (1.04–1.18) [†]	1.22 (1.14–1.30) [†]

Definition of abbreviations: CI = confidence interval; IRR = incidence rate ratio; SBT = spontaneous breathing trial.

* $P < 0.05$.

[†] $P < 0.01$.

there was a contribution from an increase in the total number of SBTs performed in hospital B, but primarily a change in documentation practices in hospital C.

Patterns of Eligibility Exclusions

There was an increase in the frequency of eligibility exclusions over time, most prominently in hospital B (Figure E2). There was no obvious change in the pattern of eligibility exclusions within a given hospital, and notably no increase in the frequency of nonspecific, discretionary eligibility exclusion criteria (e.g., “MD Order/Other”).

Clinical Outcomes

Adjusted changes in outcomes over time are shown in Table 3 and Figure E3. In hospital A, with consistently high SBT completion rates, there were no changes in outcomes in Year 1. In Year 2, DMV and mortality were

unchanged and ICU LOS slightly decreased (β , -0.09 ; 95% CI, -0.17 to -0.01 ; $P = 0.02$). In hospital B, with the lowest baseline performance and a large increase in SBT completion over time, there were no statistically significant changes in DMV, ICU length of stay, or mortality over time. In hospital C, with intermediate baseline performance and modest increases in SBT completion, there were small decreases in both DMV (β , -0.18 ; 95% CI, -0.28 to -0.09 ; $P < 0.001$) and ICU LOS (β , -0.15 ; 95% CI, -0.26 to -0.03 ; $P = 0.01$) in Postintervention Year 2, but no changes in adjusted mortality in either year.

Discussion

A multicenter evaluation of a quality improvement program using physician-

targeted financial incentives for daily SBTs in mechanically ventilated patients led to varied changes in care processes. There was an increase in SBT completion rates among eligible patients along with a simultaneous increase in the proportion of patients deemed ineligible for an SBT. The magnitude of these effects was greatest in the hospital with the lowest baseline SBT completion rates and smallest in the hospital with the highest baseline SBT completion rates. The magnitude of the effects also increased after the unit physicians did not qualify for the incentive in the first program year. In an exploratory evaluation of patterns in eligibility exclusions, we found an increase in overall exclusions but no obvious change in the pattern of exclusions, and no dramatic rise in the use of “discretionary” reasons for exclusion. The effects on clinical outcomes were mixed, with modest improvements in DMV and ICU length of stay that were not necessarily associated with concurrent improvements in SBT rates.

Our findings have important implications for the use of physician-targeted P4P programs in the ICU. First, the increase in SBT completion among eligible patients and a simultaneous increase in the documentation of eligibility exclusions is consistent with the phenomenon of exception reporting, which occurs when performance is artificially improved by reducing the number of patients in the denominator of a measure (16, 17). Exception reporting is an important effect of P4P, with potentially adverse and beneficial implications. If exception reporting occurs because of gaming and discretionary exclusion of otherwise eligible patients, then programs may not actually improve quality but still incur the opportunity costs of running the program, thus decreasing value. Conversely, if P4P programs encourage appropriate exception reporting, then these programs could improve overall quality and value by avoiding inappropriate treatment of patients to whom a given quality measure does not apply. Mechanistically, although the primary targets of this incentive program (attending ICU physicians) were not responsible for bedside documentation of SBTs, as leaders of the ICU teams they were likely involved in daily discussion of SBTs on rounds, which may have led to improved documentation on the part of respiratory therapists. Of note, although we

Figure 3. Spontaneous breathing trial (SBT) performance and eligibility by hospital. (A) There was an increase in SBT performance on eligible patient days in hospitals B and C, corresponding with (B) an increase in the proportion of ventilator days excluded from SBT eligibility in those hospitals. Vertical stripe indicates the run-in period. Vertical solid line indicates break between Postintervention Years 1 and 2. Horizontal dashed line in A represents 90% SBT completion threshold for receiving the incentive. Vent = ventilator.

found evidence of increased exception reporting, our exploratory analysis of the reasons for exclusions did not provide any evidence of outright gaming. The potential for exception reporting in this program highlights the importance of objective, unbiased quality measurement in P4P (18). Whenever possible, P4P programs should not rely on provider documentation to assess quality.

Second, the observed heterogeneity of effects across hospitals demonstrates the possibility that P4P can often simply reward high-performing hospitals rather than actually leading to improved quality (19). In this case, hospital A, with the highest baseline SBT completion rate and lowest baseline SBT eligibility rate, was the target of a previous quality improvement program that included nurse-led screening and

prompting for SBTs, which perhaps led to hospital A's high performance (13). This observation highlights the importance of selecting measures for which performance is low, and for rotating measures to ensure that all hospitals have an opportunity to improve and to prevent hospitals from "teaching to the test" by focusing on some performance measures at the expense of others (20).

Third, while the incentive pay was targeted to individual intensivist physicians, these physicians received the incentive based on overall unit performance; there are a variety of potential approaches to targeting individual physicians with financial incentives, some of which are less direct (as in this case). While it is possible that an incentive that evaluated performance at the individual physician level could yield an alternative result, the collaborative and increasingly shift-based nature of critical care would create a significant challenge in measuring and correctly attributing outcomes with such an incentive structure. Future work should address the optimal approach to structuring physician incentives in the ICU.

Finally, the inconsistent link between changes in process measures and changes in clinical outcomes highlights the need to continuously evaluate P4P programs to confirm that they are having their intended effect. In the current study, the disconnect between process and outcome is likely the result of exception reporting driving changes in SBT completion. However, our findings also call into question the link between isolated processes of care (such as SBTs) and patient outcomes in the ICU, where the determinants of clinical outcomes are complex. Indeed, although most clinical trial data indicate that daily SBTs are associated with improved outcomes (1, 2), quality improvement initiatives focusing on SBTs do not consistently translate into improvements in outcomes (5, 21). Moreover, even hospital B, which increased SBT compliance primarily by increasing the number of SBTs, failed to see meaningful improvements in outcome. This discrepancy may reflect the difference between trial efficacy and real-world effectiveness, as similarly found in parenteral versus enteral nutritional support in the ICU (22).

Our study has several limitations. First, the P4P program involved a single incentive structure within a single health system,

Table 3. Adjusted Patient Outcomes by Time Period

	Postintervention Year 1 vs. Baseline		Postintervention Year 2 vs. Baseline	
	Effect (95% CI)*	P Value	Effect (95% CI)*	P Value
Hospital A				
DMV	−0.03 (−0.10 to 0.04)	0.38	−0.02 (−0.09 to 0.05)	0.56
ICU LOS	−0.04 (−0.13 to 0.04)	0.29	−0.09 (−0.17 to −0.01)	0.02
Mortality	1.22 (0.99 to 1.51)	0.07	1.21 (0.98 to 1.49)	0.08
Hospital B				
DMV	0.27 (0.00 to 0.55)	0.05	0.20 (−0.09 to 0.48)	0.17
ICU LOS	0.33 (0.01 to 0.66)	0.04	0.22 (−0.13 to 0.56)	0.22
Mortality	1.06 (0.38 to 2.98)	0.92	1.11 (0.39 to 3.16)	0.84
Hospital C				
DMV	−0.11 (−0.21 to −0.00)	0.05	−0.18 (−0.28 to −0.09)	<0.001
ICU LOS	−0.07 (−0.19 to 0.05)	0.28	−0.15 (−0.26 to −0.03)	0.01
Mortality	0.77 (0.57 to 1.04)	0.09	0.85 (0.63 to 1.13)	0.27

Definition of abbreviations: CI = confidence interval; DMV = duration of mechanical ventilation; ICU = intensive care unit; LOS = length of stay.

*For mortality, effect is reported as an odds ratio, which is the exponentiated β coefficient for the postintervention year relative to baseline from the logistic regression model. For DMV and ICU LOS, the effect is reported as the β coefficient for the postintervention year relative to baseline from the linear regression models. In these models, DMV and ICU LOS are log-transformed, and therefore the effect roughly translates into a percent change. For example, in hospital C, there was an approximately 18% reduction in DMV in the second postintervention year compared with the baseline period.

which limits the generalizability of these findings. However, the three hospitals were diverse in size and case-mix, and the incentive structure was designed to be both simple and generalizable to other settings. Second, most of the patients and ventilator days were at hospital A, whereas the largest changes occurred at hospitals B and C. In these two hospitals, some clinically significant changes may not have been statistically significant because of small sample sizes. Third, we did not have a concurrent control group, so we are unable to exclude the possibility that the results

were not due to temporal factors unrelated to the P4P intervention. For example, ICU LOS is decreasing over time in hospital A in the absence of changes in duration of ventilation or SBT rates. However, graphical examination of the data does not suggest any obvious preexisting temporal trends that would account for our primary findings. Fourth, because the quality-based financial incentive also addressed aspects of care other than SBT completion, observed variation in clinical outcomes reflects the impact of the multicomponent quality improvement intervention, rather than just

the incentive for SBT completion. Fifth, because our data captured only calendar days of mechanical ventilation, we may have lacked the power to detect more granular changes in duration of mechanical ventilation in our linear regression analyses. Finally, we cannot confirm that the documented reasons for SBT ineligibility were aligned with the clinical characteristics at the patients' bedside, and are therefore unable to directly determine whether changes in exception reporting represent more comprehensive documentation of previously missed ineligibility as opposed to intentional or unintentional "gaming" of the P4P program. However, either scenario would be considered an effect of P4P primarily on documentation rather than meaningful changes in patient care or outcomes.

In summary, we found that a physician-targeted P4P program designed to increase SBT completion in three academic hospitals was associated with an increase in SBT completion rates that was in part driven by increases in exception reporting without meaningful changes in outcomes. These findings highlight the need for P4P programs that are based on objective, reliable measures of quality rather than solely documentation, the need to target areas of known quality deficits, and the need to rotate quality measures over time. Such efforts can help ensure that as P4P programs expand, they are used to stimulate actual improvements in clinical care, rather than isolated shifts in documentation. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

- Esteban A, Frutos F, Tobin MJ, Alía I, Solsona JF, Valverdú I, Fernández R, de la Cal MA, Benito S, Tomás R, et al.; Spanish Lung Failure Collaborative Group. A comparison of four methods of weaning patients from mechanical ventilation. *N Engl J Med* 1995;332:345–350.
- Blackwood B, Burns KE, Cardwell CR, O'Halloran P. Protocolized versus non-protocolized weaning for reducing the duration of mechanical ventilation in critically ill adult patients. *Cochrane Database Syst Rev* 2014; 11:CD006904. [Update of *Cochrane Database Syst Rev* 2010;5:CD006904.]
- Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, Sevransky JE, Sprung CL, Douglas IS, Jaeschke R, et al.; Surviving Sepsis Campaign Guidelines Committee Including the Pediatric Subgroup. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med* 2013;41:580–637.
- MacIntyre NR, Cook DJ, Ely EW Jr, Epstein SK, Fink JB, Heffner JE, Hess D, Hubmayer RD, Scheinhorn DJ; American College of Chest Physicians; American Association for Respiratory Care; American College of Critical Care Medicine. Evidence-based guidelines for weaning and discontinuing ventilatory support: a collective task force facilitated by the American College of Chest Physicians; the American Association for Respiratory Care; and the American College of Critical Care Medicine. *Chest* 2001;120(6, Suppl): 375S–395S.
- Robertson TE, Mann HJ, Hyzy R, Rogers A, Douglas I, Waxman AB, Weinert C, Alapat P, Guntupalli KK, Buchman TG; Partnership for Excellence in Critical Care. Multicenter implementation of a consensus-developed, evidence-based, spontaneous breathing trial protocol. *Crit Care Med* 2008;36:2753–2762.
- Kahn JM, Scales DC, Au DH, Carson SS, Curtis JR, Dudley RA, Iwashyna TJ, Krishnan JA, Maurer JR, Mularski R, et al.; American Thoracic Society Pay-for-Performance Working Group. An official American Thoracic Society policy statement: pay-for-performance in pulmonary, critical care, and sleep medicine. *Am J Respir Crit Care Med* 2010;181:752–761.
- Rosenthal MB, Landon BE, Normand S-LT, Frank RG, Epstein AM. Pay for performance in commercial HMOs. *N Engl J Med* 2006;355: 1895–1902.

8. Glickman SW, Ou F-S, DeLong ER, Roe MT, Lytle BL, Mulgund J, Rumsfeld JS, Gibler WB, Ohman EM, Schulman KA, *et al*. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA* 2007;297:2373–2380.
9. Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, Bratzler DW. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med* 2007;356:486–496.
10. Khanduja K, Scales DC, Adhikari NKJ. Pay for performance in the intensive care unit: opportunity or threat? *Crit Care Med* 2009;37:852–858.
11. Kristensen SR, Meacock R, Turner AJ, Boaden R, McDonald R, Roland M, Sutton M. Long-term effect of hospital pay for performance on mortality in England. *N Engl J Med* 2014;371:540–548.
12. Barbash IJ, Seymour CW, Gunn SR, Kahn JM. Impact of a physician-based pay-for-performance program on use of spontaneous breathing trials in mechanically ventilated patients [abstract]. *Am J Respir Crit Care Med* 2016;193:A6446.
13. Kahn JM, Gunn SR, Lorenz HL, Alvarez J, Angus DC. Impact of nurse-led remote screening and prompting for evidence-based practices in the ICU. *Crit Care Med* 2014;42:896–904.
14. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care* 1998;36:8–27.
15. Localio AR, Berlin JA, Ten Have TR, Kimmel SE. Adjustments for center in multicenter studies: an overview. *Ann Intern Med* 2001;135:112–123.
16. Doran T, Fullwood C, Reeves D, Gravelle H, Roland M. Exclusion of patients from pay-for-performance targets by English physicians. *N Engl J Med* 2008;359:274–284.
17. Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, Roland M. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med* 2006;355:375–384.
18. Pronovost PJ, Miller M, Wachter RM. The GAAP in quality measurement and reporting. *JAMA* 2007;298:1800–1802.
19. Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance: from concept to practice. *JAMA* 2005;294:1788–1793.
20. Werner RM, Konetzka RT, Kruse GB. Impact of public reporting on unreported quality of care. *Health Serv Res* 2009;44:379–398.
21. Robertson TE, Sona C, Schallom L, Buckles M, Cracchiolo L, Schuerer D, Coopersmith CM, Song F, Buchman TG. Improved extubation rates and earlier liberation from mechanical ventilation with implementation of a daily spontaneous-breathing trial protocol. *J Am Coll Surg* 2008;206:489–495.
22. Harvey SE, Parrott F, Harrison DA, Bear DE, Segaran E, Beale R, Bellington G, Leonard R, Mythen MG, Rowan KM; CALORIES Trial Investigators. Trial of the route of early nutritional support in critically ill adults. *N Engl J Med* 2014;371:1673–1684.