



Impact of Financial Incentives on Early and Late Adopters among US Hospitals: observational study

Igna Bonfrer,^{1,2} Jose F Figueroa,^{1,3,4} Jie Zheng,¹ E John Orav,^{3,4,5} Ashish K Jha^{1,3,6}

¹Department of Health Policy and Management, Harvard T H Chan School of Public Health, 42 Church St, Cambridge, MA 02138, USA

²Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, Netherlands

³Department of Medicine, Harvard Medical School, Cambridge, MA, USA

⁴Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

⁵Department of Biostatistics, Harvard T H Chan School of Public Health, Cambridge, MA, USA

⁶Department of General Internal Medicine, VA Boston Healthcare System, Boston, MA, USA

Correspondence to: A Jha
ajha@hsph.harvard.edu

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2018;360:j5622
<http://dx.doi.org/10.1136/bmj.j5622>

Accepted: 17 November 2017

ABSTRACT

OBJECTIVE

To examine how hospitals that volunteered to be under financial incentives for more than a decade as part of the Premier Hospital Quality Incentive Demonstration (early adopters) compared with similar hospitals where these incentives were implemented later under the Hospital Value-Based Purchasing program (late adopters).

DESIGN

Observational study.

SETTING

1189 hospitals in the USA (214 early adopters and 975 matched late adopters), using Hospital Compare data from 2003 through 2013.

PARTICIPANTS

1 371 364 patients aged 65 years and older, using 100% Medicare claims.

MAIN OUTCOME MEASURES

Clinical process scores and 30 day mortality.

RESULTS

Early adopters started from a slightly higher baseline of clinical process scores (92) than late adopters (90). Both groups reached a ceiling (98) a decade later. Starting from a similar baseline, just below 13%, early and late adopters did not have significantly ($P=0.25$) different mortality trends for conditions targeted by the program (0.05% point difference quarterly) or for conditions not targeted by the program (-0.02% point difference quarterly).

CONCLUSIONS

No evidence that hospitals that have been operating under pay for performance programs for more than a decade had better process scores or lower mortality than other hospitals was found. These findings suggest that even among hospitals that volunteered to participate in pay for performance programs, having additional time is not likely to turn pay for performance programs into a success in the future.

Introduction

OECD (Organisation for Economic Co-operation and Development) countries use financial incentives to improve the quality of healthcare.¹ These incentives are increasingly being used in low and middle income countries.² The USA and the UK are probably the furthest ahead, but other countries are monitoring the experiences of these countries to determine what they might do. Providing incentives to improve the performance of hospitals has become common in the USA over the last decade.³⁻⁵ The Premier Hospital Quality Incentive Demonstration (HQID), a voluntary program run by the Centers for Medicare and Medicaid Services,⁶⁻⁷ ran from 2003 to 2009. It became the model for the national Hospital Value-Based Purchasing (HVBP) program, adopted after the passage of the Affordable Care Act,⁸⁻⁹ which has run since 2011 (details on both programs are shown in web appendices 1 to 3). The hospitals that were invited by Centers for Medicare and Medicaid Services and then participated in the HQID (early adopters) have been under financial incentives to improve quality for more than a decade and are likely to have a comparative advantage, having had more time to refine their delivery of care and strategically focus on improving patient outcomes. Previous evaluations of the HQID and the HVBP program showed limited impact on process measures but no improvements in patient outcomes or in cost reduction.¹⁰⁻²⁰ However, advocates argue that it takes time for hospitals to make meaningful improvements and that we need patience to better understand how delivery of care under pay for performance programs changes care.

Improving outcomes is difficult; it can require changes to workflows, restructuring the way providers are paid, and alignment of information technology systems.²¹⁻²² The national HVBP program, which attaches incentives directly to outcomes of Medicare beneficiaries, has been in effect for a few years, and early evidence suggests that it has had little impact on patient outcomes.¹⁸ One might expect that most hospitals have not yet been able to equip themselves adequately to make meaningful improvements to patient care. However, the early adopters of pay for performance programs, the Premier HQID hospitals, have been under financial incentives since 2003 (except for 2009 to 2010 when the HQID ended and before the HVBP program was implemented).²³⁻²⁴ These early adopters have likely had enough time to make the difficult structural changes which are necessary to improve outcomes. Given that early adopters volunteered to be in the HQID, they likely represent the best case scenario of how much improvement we might expect over a longer period. Whether the early adopters have outpaced late adopters or whether

WHAT IS ALREADY KNOWN ON THIS TOPIC

Previous studies have found that pay for performance programs have had limited impact on process measures and no impact on patient outcomes

No study has examined whether early adopters of pay for performance programs (ie, Premier Hospital Quality Incentive Demonstration) out performed late adopters of pay for performance programs (ie, Hospital Value-Based Purchasing)

WHAT THIS STUDY ADDS

Clinical process scores or 30 day mortality for Medicare beneficiaries were not found to be better at hospitals that have been operating under pay for performance programs for more than a decade

Pay for performance programs as currently implemented are unlikely to be successful in the future, even if their timeframes are extended

the HQID had spillover effects on the late adopters is unclear – and empirical evidence would be helpful for healthcare policymakers as the new administration makes changes to, or replaces, the Affordable Care Act.

As we continue to shift to value based payments in healthcare,²⁵ it is crucial to understand the long term effects of financial incentives on quality of care and effectiveness of HVBP programs. Therefore, we used national Medicare claims data from 2003 to 2013 to answer three questions. First, how do early adopters that had participated in the HQID perform under the HVBP program compared with similar hospitals that did not participate in the HQID (late adopters) in terms of clinical process scores? Second, how do early adopters compare with late adopters on 30 day mortality outcomes for the three target conditions: acute myocardial infarction, congestive heart failure, and pneumonia? And third, are there any spillover effects for 30 day mortality on non-target conditions?

Methods

Data and matching

We identified all 2702 American acute care hospitals participating in Hospital Value-Based Purchasing (HVBP) programs in 2013 using publicly available Hospital Compare data. Of these hospitals, 233 had also voluntarily participated in the Hospital Quality Incentive Demonstration (HQID) from 2003 through 2009, and we defined these hospitals as early adopters (for details see web appendix 4). We obtained baseline hospital characteristics from the 2003 American Hospital Association Annual Survey.²⁶

Previous work has shown that different hospital types are important to consider when examining patient outcomes under pay for performance programs.²⁷ To limit the potential bias arising from differences in observable hospital characteristics between early and late adopters, including volunteering to participate in the HQID, we applied coarsened exact matching.²⁸ This matching method, like other matching methods, prunes observations from the data. The remaining data have a better balance between the treated and control groups for the baseline of the outcome measure and for the following observed hospital characteristics: size, region, ownership, teaching status, location, presence of an intensive care unit, and safety net status (defined as the top 25% of hospitals with the largest disproportionate share index). The advantage of coarsened exact matching over other matching methods is that the resulting data are exactly balanced and do not need to be controlled further because both groups have exactly the same observed hospital characteristics (for details see web appendix 4).²⁸

Data on clinical process scores (range from 0 to 100) for each hospital were obtained from Hospital Compare.^{29 30} Clinical process scores were available for the three HVBP program target conditions (acute myocardial infarction, congestive heart failure, and pneumonia) from 2004 through 2014 (latest available year) (see web appendix 1 and 4). The clinical process score used in this study is the measure that the Centers

for Medicare and Medicaid Services uses to determine the clinical process score of hospitals. These clinical process scores assess whether “what is known to be ‘good’ medical care has been applied.”^{31 32} These process measures indicate whether or not a healthcare provider gives the recommended care to patients with a particular condition.³³

Using the matched dataset, we identified the subset of Medicare beneficiaries admitted to the hospital from the 100% Medicare inpatient fee-for-service claims for 2003 through 2013 (latest available year at time of analysis). We included 1 371 364 patients with any of the three HVBP program target conditions or a selected non-target condition (stroke, gastroenteritis and esophagitis, gastrointestinal bleed, urinary tract infection, metabolic disorder, arrhythmia, and renal failure) (see web appendix 4).

Statistical analysis

Using the dataset of matched hospitals and their associated patients, we performed a segmented linear regression analysis to estimate differences in trends over time for outcome measures for both early and late adopters. Using these segmented linear models, we assessed the trends in outcome measures over three periods: the HQID period (fourth quarter of 2003 to fourth quarter of 2009), the pre-HVBP period (first quarter of 2010 to second quarter of 2011), and the HVBP period (third quarter 2011 to fourth quarter of 2013), and the associated years for the clinical process scores which are based on annual as opposed to quarterly data. We expected differences in trends to increase over time as the duration of the exposure to incentives increases.

Clinical process scores

Using the annual hospital level data, we estimated the relation between early adopter status and clinical process scores. The linear segmented regression model (see web appendix 4 for a formal description of all models used) included a binary variable to determine whether a hospital was an early adopter and further corrected for the underlying annual time trend in the HQID, pre-HVBP, and HVBP periods as well as the interactions between the time and early adopter variables. We included a stratum fixed effect that represented the strata of comparable hospitals based on the coarsened exact matching and a hospital random effect following the Centers for Medicare and Medicaid Services methods to account for correlation between patients within each hospital and for correlation over time. We weighted our analyses for hospital volume. We confirmed robustness of our findings using a hospital fixed effect, controlling for unobserved hospital characteristics that were time invariant (not shown).

30 Day mortality

Using quarterly patient level data, we estimated a similar model for mortality within 30 days after admission, standardized for age, gender, and

comorbidities. We examined 30 day mortality data for each quartile for the combination of target conditions, for the three target conditions separately, and for the non-target conditions over the fourth quarter of 2003 to the fourth quarter of 2013. In addition, the segmented regression models included a correction for seasonality; the target condition for which a patient was admitted; the underlying time trend in the HQID, pre-HVBP, and HVBP periods by quarter; and patient characteristics (age, sex, race, and comorbidities) (see web appendix 4).

Differences in trends

Based on the estimates from the different linear segmented regressions, we compared differences between early and late adopters for each of the three periods and subsequently compared these differences for the HVBP period with those in the pre-HVBP period to determine whether the outcomes improved differently over time for early and late adopters.

All analyses were performed using software packages SAS version 9.4 or Stata version 14.0.

Patient involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

Results

Late adopters were frequently dropped during the matching process for being dissimilar to the early adopters with regards to the following characteristics: small size; located in the north east and west; for profit or public; non-teaching; small rural and large rural; and safety net hospital status. This resulted in a matched dataset of 1189 hospitals that are mostly medium or large, private not for profit, and based in urban areas. Table 1 shows the baseline hospital characteristics before and after matching. The associated sample of individuals consists of 263 088 patients admitted to a hospital classed as an early adopter, and 1 108 276 admitted to a late adopter. Table 2 shows the patient characteristics before and after matching. As we would expect from a matching on hospital characteristics, the patient characteristics across early and late adopters still differed after matching, mostly in terms of race, with early adopters caring for a larger share of white patients (87.4% v 84.9%, $P<0.001$). All observed patient characteristics were controlled for in the regression models.

Clinical process scores were incentivized during both the Hospital Quality Incentive Demonstration (HQID) and the current Hospital Value-Based Purchasing (HVBP) periods. Figure 1 shows that early adopters started from a slightly higher baseline clinical process score in 2004. Table 3 shows that early adopters had an average score of 91.5 versus 89.9 for late adopters in the HQID period for the combined target conditions. Improvements among the early adopters were smaller

Table 1 | Hospital characteristics at baseline. Values are percentages unless stated otherwise

Characteristics	Before matching			After matching			Not matched*
	Early adopters	Late adopters	P value	Early adopters	Late adopters	P value	
No	233	2469		214	975		1492
Size:							
Small	12.9	25.1	<0.001	13.1	13.1	1.00	34.1
Medium	61.8	61.0		64.0	64.0		51.7
Large	25.3	13.9		22.9	22.9		14.2
Region:							
North east	13.7	18.0	0.002	14.0	14.0	1.00	14.8
Midwest	22.7	24.8		23.4	23.4		21.9
South	50.2	37.8		50.0	50.0		39.8
West	13.3	19.4		12.6	12.6		23.7
Ownership:							
For profit	0.0	18.0	<0.001	0.0	0.0	1.00	29.4
Private not for profit	91.0	66.2		91.6	91.6		49.6
Public	9.0	15.8		8.4	8.4		21.0
Teaching status:							
Non	59.2	70.3	0.001	59.3	59.3	1.00	9.2
Major	14.6	9.3		14.5	14.5		19.3
Minor	26.2	20.4		26.2	26.2		71.5
Location:							
Urban	75.5	63.8	<0.001	77.6	77.6	1.00	55.3
Suburban	6.4	4.0		4.7	4.7		5.7
Large rural	14.6	22.4		14.0	14.0		25.0
Small rural	3.4	9.8		3.7	3.7		14.0
Has intensive care unit	89.7	83.2	0.010	91.6	91.6	1.00	71.1
Safety net hospital	21.9	26.4	0.130	19.6	19.6	1.00	31.3
Medicare patients	44.4	43.4	0.209	44.5	43.6	0.370	42.2
Medicare to non-Medicare ratio	0.8	0.8	0.209	0.8	0.8	0.370	0.7
Mean age (years) of Medicare patients	78.7	78.8	0.130	78.7	78.9	0.051	78.7

* Twenty one of the non-matched hospitals had missing data and are not included

Table 2 | Patient characteristics at baseline. Values are percentages unless stated otherwise

Characteristics	Before matching			After matching		
	Early adopters	Late adopters	P value	Early adopters	Late adopters	P value
No	275 354	1 428 540		263 088	1 108 276	
Mean age (years)	78.6	78.7	0.04	78.8	78.7	0.32
Male	41.3	41.4	0.89	40.5	41.0	0.08
Race:						
White	86.8	84.7	<0.001	87.4	84.9	<0.001
Black	9.9	11.6		9.5	11.4	
East Asian	1.0	1.0		1.0	1.0	
Native American	0.6	0.7		0.6	0.7	
Hispanic	1.2	1.4		1.0	1.3	
Other	0.6	0.7		0.6	0.7	
Comorbidities:						
Chronic pulmonary disease	26.3	26.9	0.11	27.0	27.3	0.50
Congestive heart failure	14.7	15.3	0.01	14.7	15.3	0.01
Depression	5.6	5.5	0.55	5.8	5.4	0.01
Diabetes	25.4	25.3	0.77	25.6	25.5	0.67
Hypertension	46.6	45.4	0.01	46.2	45.3	0.01
Ischemia heart disease	23.7	23.4	0.54	23.0	22.5	0.23
Liver disease	1.0	1.0	0.33	1.0	1.0	0.04
Renal failure	7.1	7.5	0.06	6.9	7.3	0.02

during the HQID period (difference -0.21 , 95% confidence interval -0.31 to -0.11), although early adopters continued to perform at a slightly higher level than the late adopters (-0.55 , -1.01 to -0.10) during the pre-HVBP period. Over the HVBP period, early and late adopters no longer differed in their clinical process scores. In the HVBP period, the increase in the clinical process score reached a ceiling where early and late adopters approach the same level: 98.5 versus 98.2 with a difference of -0.27 (95% confidence interval -0.77 to 0.22). Comparing the difference in the differences in the trend for the HVBP period with the pre-HVBP period, we found that there is no significant difference ($P=0.19$) across early and late adopters (last two columns of table 3). Estimates for the individual target conditions confirmed these patterns (web appendix 5).

30 Day mortality

Figure 2 shows that mortality fell for both early and late adopters during the study period. Both groups started from a similar baseline (14.9% and 14.8% for the early and late adopters in the fourth quarter of 2003) and ended at the same rate of 9.9% for both groups in the fourth quarter of 2013. Table 4 shows that the average

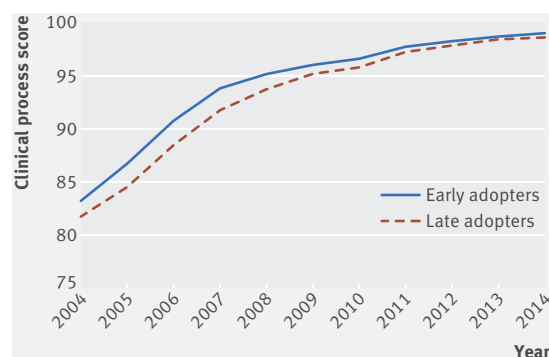


Fig 1 | Clinical process scores for target conditions

mortality is slightly higher among late adopters for each period. This pattern is not confirmed by the underlying mortality from individual target conditions (web appendix 6). For the non-targeted conditions, we found no significant differences ($P=0.48$) in average mortality by period. In formal testing with the linear segmented regression models, we found that the reduction in mortality was comparable for early and late adopters during the HQID period (difference 0.00% points, 95% confidence interval -0.01 to 0.01). The reduction in mortality (-0.04% points for early adopters and -0.02% points for late adopters) slowed down after the abolition of the financial incentives. There continued to be no noticeable differences in mortality reductions between early and late adopters over the pre-HVBP (difference -0.02% points, 95% confidence interval -0.06 to 0.02) and HVBP (0.02 , -0.02 to 0.07) period. Comparing the difference in the differences in the trends for the HVBP versus pre-HVBP period, we found no noticeable effect (0.05 , -0.03 to 0.13). This suggests that the HQID did not have a meaningful effect on mortality through 2013, even though the hospitals had a decade of experience and volunteered to participate in the demonstration. These results, as expected, did not lead to spillover effects on mortality for the non-targeted conditions. The reduction in mortality from non-targeted conditions during the HQID period was comparable across early and late adopters (0.00 , 0.00 to 0.01). After the abolition of the financial incentives, there continued to be no noticeable differences in mortality reductions between early and late adopters (0.00 , -0.03 to 0.03) nor were these differences present over the HVBP period (-0.02 , -0.05 to 0.01). Comparing the difference in the differences in the trends for the HVBP with the pre-HVBP period, we found no noticeable effect (-0.02 , -0.07 to 0.03). Finally, when examining the individual target conditions, we found qualitatively similar effects (web appendix 7).

Table 3 | Trends in clinical process scores over time. Values are clinical process scores (range 0 to 100) unless stated otherwise

Period	Average clinical process score			Yearly change in clinical process scores			Difference in differences in trends HVBP–pre-HVBP (95% CI)	P value
	Early adopters	Late adopters	Difference (95% CI)	Early adopters	Late adopters	Difference in trends (95% CI)		
Target conditions*								
HQID	91.5	89.9	−1.59 (−1.98 to −1.20)	2.44	2.65	−0.21 (−0.31 to −0.11)	0.24 (−0.12 to 0.59)	0.19
pre-HVBP	97.3	96.7	−0.55 (−1.01 to −0.10)	0.38	0.70	−0.32 (−0.52 to −0.12)		
HVBP	98.5	98.2	−0.27 (−0.77 to 0.22)	0.53	0.61	−0.08 (−0.28 to 0.12)		

HQID=Hospital Quality Incentive Demonstration; HVBP=Hospital Value-Based Purchasing.

*Acute myocardial infarction, congestive heart failure, and pneumonia

Discussion

While hospitals should be rewarded for better outcomes, not just for the number of inputs, the impact of pay for performance programs has been limited and disappointing. However, there has been a longstanding view that it may take time to make meaningful improvements to care under financial incentive programs. We examined how a group of hospitals, having effectively been under a pay for performance program for more than a decade, fared under the current Hospital Value-Based Purchasing (HVBP) program compared with control hospitals with far less experience in a pay for performance program. We found that despite substantial time, the early adopter Premier Hospital Quality Incentive Demonstration (HQID) hospitals did not have better outcomes and had only marginally better adherence to clinical process measures compared with a group of controls over time. Clinical process scores reached a ceiling for both groups, suggesting that the comparative advantage of the early adopters reduced over time, allowing the late adopters to catch up.

We found, contrary to our hypothesis, that the improvements in clinical process scores were smaller for the early adopters during the HQID period, even though they did perform at a slightly higher level. We further found no noticeable difference in the trends across early and late adopters for the HVBP versus pre-HVBP period. It might have been the case that the definition and communication of specific, measurable processes to provide good quality care as part of the highly visible HQID, resulted in non-participating hospitals also improving their processes, because of a more intrinsic motivation to provide good quality care. The inevitable interaction between early and late

adopters might have further led to spillover effects by healthcare personnel teaching each other about standards and approaches to improve quality.

Our findings have important implications for the way policymakers should approach pay for performance programs. These findings provide evidence that having additional time is not likely to turn these programs into a success, at least as far as patient outcomes are concerned. Even for clinical processes, while HQID hospitals began with better performance at baseline, by the end of the study period, the gap between early and late adopters was gone – presumably because of a ceiling effect. This suggests that there is a need to change the measures for clinical process scores.

The limited effects after more than a decade of financial incentives might be explained by different factors. First, the incentives are very small (see web appendix 1 for size and timing of performance bonuses), and given that these incentives only cover a small set of conditions and that hospitals receive revenue from a multitude of payers, this modest incentive is diluted, limiting its impact.³⁴ However, hospital margins tend to be small, usually only a few percentage points, so one might assume that these Medicare payments could still motivate many hospitals.⁷ Second, the program was extremely complex (in terms of its structure and how hospitals were incentivized), making it more difficult for hospitals to engage meaningfully in the program. Previous work has shown that when the incentivized measure is clear and simple (eg, hospital readmissions in the Hospital Readmission Reduction Program),³⁵ it is possible to observe the impact of financial incentives on performance even within a year after implementation. Finally, given that studies suggest that people tend to discount future gains when the delay in receiving the benefit increases, it is possible that having to wait until the end of the year to receive bonuses or penalties might have reduced the impact.³⁶

Weaknesses of this study

There are limitations to this study. First, observable hospital characteristics used for coarsened exact matching and observable patient characteristics used as covariates in the linear segmented regression models might not sufficiently capture unobserved differences between the early adopters who voluntarily joined the demonstration and the matched late adopters. This might lead to an overestimation of the effect of having more time to adapt to financial incentives because part of this effect might be driven by unobserved

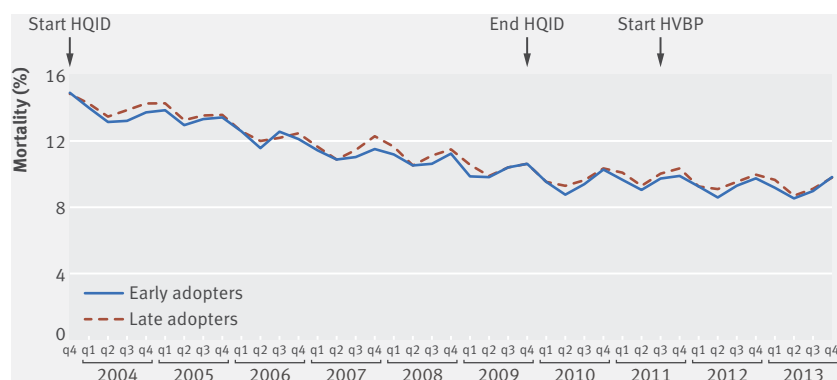


Fig 2 | Standardized 30 day mortality for target conditions. HQID=Hospital Quality Incentive Demonstration; HVBP=Hospital Value-Based Purchasing

Table 4 | Trends in mortality over the three periods. Values are percentages unless stated otherwise

Period	Average mortality			Quarterly change in mortality				P value
	Early adopters	Late adopters	PP Difference (95% CI)	Early adopters	Late adopters	PP Difference in trends (95% CI)	PP Difference in difference in trends pre-HVBP v HVBP (95% CI)	
Target conditions*								
HQID	12.2	12.5	0.27 (0.03 to 0.50)	-0.19	-0.19	0.00 (-0.01 to 0.01)	0.05 (-0.03 to 0.13)	0.25
pre-HVBP	9.8	10.0	0.29 (0.01 to 0.56)	-0.04	-0.02	-0.02 (-0.06 to 0.02)		
HVBP	9.4	9.7	0.28 (0.01 to 0.56)	-0.05	-0.07	0.02 (-0.02 to 0.07)		
Non-target conditions								
HQID	9.1	9.2	0.09 (-0.08 to 0.27)	-0.11	-0.11	0.00 (0.00 to 0.01)	-0.02 (-0.07 to 0.03)	0.48
pre-HVBP	7.4	7.6	0.14 (-0.06 to 0.34)	-0.06	-0.06	0.00 (-0.03 to 0.03)		
HVBP	7.2	7.3	0.15 (-0.04 to 0.35)	-0.04	-0.02	-0.02 (-0.05 to 0.01)		

PP=Percentage point; HQID=Hospital Quality Incentive Demonstration; HVBP=Hospital Value-Based Purchasing.

*Acute myocardial infarction, congestive heart failure, and pneumonia

characteristics that make those who voluntarily joined different from those who did not (ie, primed to be successful). However, we found virtually no effects so an upwards bias seems unlikely, and we also have a relatively good comparison group so there is no obvious reason to expect large unobserved differences between both groups of acute care hospitals and their patients. Second, generalization to the general population is limited because this study only looks at patients aged 65 and older. However, Medicare covers more than 55 million people and accounts for more than 20% of the total American health spending.³⁷ Third, given the other interventions being implemented simultaneously by Centers for Medicare and Medicaid Services, especially the public reporting under the Hospital Compare initiative for process measures (since 2005) and of 30 day mortality (since 2008), it is difficult to determine the pure effect of pay for performance financial incentives. This might have led to an underestimation of the effects studied here. However, because these interventions were implemented across all hospitals studied, both early and late adopters, we would expect an effect on the outcome levels but not on the differences in the trends across both groups. In other words, the average clinical process scores across early and late adopters might have changed as a result of public reporting, but we expect these changes to be similar across both groups, therefore not affecting the differences in the trends across both groups. Related to this, because the indicators for the HQID were published, late adopters might have already been aware of quality measures to be used for the targeted conditions, in turn already improving their clinical process scores. Next, when comparing trends in clinical processes of HQID and non-HQID hospitals, given that clinical process measures are top coded, it is possible that it may represent an artifact of top coding as most hospitals may be approaching their asymptote.¹⁴ Finally, we could not study the cost effectiveness of either of the two pay for performance programs. So far there have been mixed findings on the effectiveness of HVBP programs to simultaneously improve quality and control costs.³⁸

Conclusion

We found that hospitals that have been under financial incentives for more than a decade have not been able to reduce patient mortality more than the late adopters, which had only been under financial incentives for less than three years. The HVBP program as currently structured, is not living up to the promise advocates originally envisioned. Given its cost, policymakers in the USA should consider one of two things: revise the current program or potentially end it. Given major efforts to shift more and more payments toward a value based framework, it is more likely that the program will be revised rather than completely stopped. There are a few ways that this may be done. First, the program should consider increasing the incentives. Currently, the small amount of money at stake is arguably not enough to change the way hospitals are doing business. Second, policy makers should focus on a few measures that matter most to patients (mortality, patient experience, and functional status). The current structure includes numerous measures that are difficult to track and measure by hospital leaders, and therefore, hard to improve. Given the growing worldwide interest in pay for performance programs and the unclear American health policy agenda,³⁹ these findings should be considered by policymakers when assuming that programs like these simply need more time to have a meaningful effect.

We thank the discussant and participants of the ASHEcon biennial conference in Philadelphia, participants of the Annual Research Meeting of AcademyHealth in Boston, and participants of the Health Economics seminar at Erasmus University Rotterdam for their valuable insights.

Contributors: All authors contributed to the design and conduct of the study; data collection and management; interpretation of the data; and preparation, review, or approval of the manuscript. Data analyses were performed by IB and JZ. IB, JFF, and AKJ are the guarantors.

Funding: IB received funding from a Rubicon Fellowship provided by the Netherlands Organization for Scientific Research. The funders had no role in the study design; collection, analysis, and interpretation of data; writing of the report; or decision to submit the article for publication.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organization for the submitted work; no financial

relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Data sharing: No additional data are available.

Transparency: The lead author (IB) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- 1 Roland M, Guthrie B. Quality and Outcomes Framework: what have we learnt? *BMJ* 2016;354:i4060. doi:10.1136/bmj.i4060
- 2 Van de Poel E, Flores G, Ir P, O'Donnell O. Impact of Performance-Based Financing in a Low-Resource Setting: A Decade of Experience in Cambodia. *Health Econ* 2016;25:688-705. doi:10.1002/hec.3219
- 3 Rosenthal MB, Fernandopulle R, Song HR, Landon B. Paying for quality: providers' incentives for quality improvement. *Health Aff (Millwood)* 2004;23:127-41. doi:10.1377/hlthaff.23.2.127
- 4 Cutler DM. Payment reform is about to become a reality. *JAMA* 2015;313:1606-7. doi:10.1001/jama.2015.1926
- 5 Roadmap for Implementing Value Driven Healthcare in the Traditional Medicare Fee-for-Service Program. Baltimore: 2016. Centers for Medicare & Medicaid Services. (Accessed 01/21, 2016, at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/VBPRoadmap_OEA_1-16_508.pdf).
- 6 Premier Hospital Quality Incentive Demonstration. Baltimore: 2015. Centers for Medicare & Medicaid Services. (Accessed 01/21, 2016, at <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalPremier.html>).
- 7 Cashin C. United States: Hospital quality incentive demonstration. In: Cashin C, Chi Y, Smith P, Borowitz M, Thomson S, eds. *Paying for Performance in Health Care, Implications for health system performance and accountability*. 1st ed. Open University Press, 2014: 287-300.
- 8 The Hospital Value-Based Purchasing (HVBP) Program. Baltimore: 2016. Centers for Medicare & Medicaid Services. (Accessed 07/28, 2016, at <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/HVBP/Hospital-Value-Based-Purchasing.html>).
- 9 Hospital Value-Based Purchasing. Baltimore: 2015. Centers for Medicare & Medicaid Services. (Accessed 01/21, 2016, at <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/HVBP/Hospital-Value-Based-Purchasing.html>).
- 10 Eijkenaar F, Emmert M, Scheppach M, Schöffski O. Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy* 2013;110:115-30. doi:10.1016/j.healthpol.2013.01.008
- 11 Lindenauer PK, Remus D, Roman S. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med* 2007;356:486-96. doi:10.1056/NEJMsa064964
- 12 United States Government Accountability Office. Hospital Value-Based Purchasing. Initial Results Shows Modest Effects on Medicare Payments and No Apparent Change in Quality-of-Care Trends. GAO-16-9. Washington D.C.: Government Accountability Office, 2015. (Report to Congressional Committees). <http://www.gao.gov/assets/680/672899.pdf>
- 13 Ryan AM, Burgess JF, Jr, Pesko MF, Borden WB, Dimick JB. The early effects of Medicare's mandatory hospital pay-for-performance program. *Health Serv Res* 2015;50:81-97. doi:10.1111/1475-6773.12206
- 14 Werner RM, Kolstad JT, Stuart EA, Polsky D. The effect of pay-for-performance in hospitals: lessons for quality improvement. *Health Aff (Millwood)* 2011;30:690-8. doi:10.1377/hlthaff.2010.1277
- 15 Ryan AM, Blustein J, Casalino LP. Medicare's flagship test of pay-for-performance did not spur more rapid quality improvement among low-performing hospitals. *Health Aff (Millwood)* 2012;31:797-805. doi:10.1377/hlthaff.2011.0626
- 16 Ryan AM, Krinsky S, Maurer KA, Dimick JB. Changes in Hospital Quality Associated with Hospital Value-Based Purchasing. *N Engl J Med* 2017;376:2358-66. doi:10.1056/NEJMsa1613412
- 17 Jha AK, Joynt KE, Orav EJ, Epstein AM. The long-term effect of premier pay for performance on patient outcomes. *N Engl J Med* 2012;366:1606-15. doi:10.1056/NEJMsa1112351
- 18 Figueroa JF, Tsugawa Y, Zheng J, Orav EJ, Jha AK. Association between the Value-Based Purchasing pay for performance program and patient mortality in US hospitals: observational study. *BMJ* 2016;353:i2214. doi:10.1136/bmj.i2214
- 19 Ryan AM. Effects of the Premier Hospital Quality Incentive Demonstration on Medicare patient mortality and cost. *Health Serv Res* 2009;44:821-42. doi:10.1111/j.1475-6773.2009.00956.x
- 20 Glickman SW, Ou FS, DeLong ER. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA* 2007;297:2373-80. doi:10.1001/jama.297.21.2373
- 21 Rosenthal MB, Landon BE, Howitt K, Song HR, Epstein AM. Climbing up the pay-for-performance learning curve: where are the early adopters now? *Health Aff (Millwood)* 2007;26:1674-82. doi:10.1377/hlthaff.26.6.1674
- 22 Rosenthal MB, Dudley RA. Pay-for-performance: will the latest payment trend improve care? *JAMA* 2007;297:740-4. doi:10.1001/jama.297.7.740
- 23 About the Law. Washington D.C.: 2016. U.S. Department of Health & Human Services. (Accessed 06/30, 2016, at <https://www.hhs.gov/healthcare/about-the-law/read-the-law/>).
- 24 Frequently Asked Questions Hospital Value-Based Purchasing Program. Baltimore: 2012. Centers for Medicare & Medicaid Services. (Accessed 06/30, 2016, at <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/HVBP/HVBP-FAQs.pdf>).
- 25 Care B, Spending S, People H. Improving Our Health Care Delivery System. Baltimore: 2015. Centers for Medicare & Medicaid Services. (Accessed 01/26, 2016, at <https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2015-Fact-sheets-items/2015-01-26.html>).
- 26 Hospital Database. Washington D.C.: 2017. American Hospital Association. (Accessed 08/03, 2017, at <https://www.ahadataviewer.com/about/hospital-database/>).
- 27 Figueroa JF, Wang DE, Jha AK. Characteristics of hospitals receiving the largest penalties by US pay-for-performance programmes. *BMJ Qual Saf* 2016;25:898-900. doi:10.1136/bmjqs-2015-005040
- 28 Iacus SM, King G, Porro G. Causal Inference without Balance Checking: Coarsened Exact Matching. *Polit Anal* 2012;20:1-24. doi:10.1093/pan/mpr013
- 29 Hospital Compare datasets. Baltimore: 2016. Centers for Medicare and Medicaid Services. (Accessed 05/18, 2016, at <https://data.medicare.gov/data/hospital-compare>).
- 30 Hospital Compare Data Archive. Baltimore: 2016. Centers for Medicare and Medicaid Services. (Accessed 05/18, 2016, at <https://data.medicare.gov/data/archives/hospital-compare>).
- 31 Ryan AM, Burgess JF, Jr, Tompkins CP, Wallack SS. The relationship between Medicare's process of care quality measures and mortality. *Inquiry* 2009;46:274-90. doi:10.5034/inquiryjrn1.46.03.274
- 32 Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966;44:166-206. doi:10.2307/3348969
- 33 Inpatient measures. Baltimore: 2017. Center for Medicare and Medicaid Services. (Accessed 07/21, 2017, at <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/InpatientMeasures.html>).
- 34 Damberg CL, Raube K, Teleki SS, Dela Cruz E. Taking stock of pay-for-performance: a candid assessment from the front lines. *Health Aff (Millwood)* 2009;28:517-25. doi:10.1377/hlthaff.28.2.517
- 35 Hospital Readmission Reduction Program. Baltimore: 2017. Centers for Medicare and Medicaid Services. (Accessed 07/22, 2017, at <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html>).
- 36 Eijkenaar F. Key issues in the design of pay for performance programs. *Eur J Health Econ* 2013;14:117-31. doi:10.1007/s10198-011-0347-6
- 37 The Facts on Medicare Spending and Financing. Washington D.C.: 2015. The Henry J. Kaiser Family Foundation. (Accessed 07/11, 2016, at <http://www.kff.org/medicare/fact-sheet/medicare-spending-and-financing-fact-sheet/>).
- 38 Damberg C, Sorbero M, Lovejoy S, Martsoff G, Raaen L, Mandel D. *Measuring Success in Health Care Value-Based Purchasing Programs*. 1. RAND Corporation, 2014.
- 39 Witter S, Fretheim A, Kessy FL, Lindahl AK. Paying for performance to improve the delivery of health interventions in low- and middle-income countries [Review]. *Cochrane Database Syst Rev* 2012;2:CD007899. doi:10.1002/14651858.CD007899.pub2

Appendix: Supplementary materials