

# Are Performance-Based Payment Structures More Strongly Correlated with Better Health Outcomes Compared to Solely Fee-For-Service Payment for Surgeons?

Matthew George<sup>1</sup> and John Anagnost<sup>#</sup>

<sup>1</sup>Cistercian Preparatory School

<sup>#</sup>Advisor

## ABSTRACT

For years, the healthcare industry has compensated its surgeons with a simple fee-for-service model. With more payment structures becoming prevalent, we decided to review one specific type of payment structure, pay-for-performance, and its impact on patient health outcomes. We hypothesized that, on average, across all specialties, performance-based contracts would yield better health outcomes compared to fee-for-service payments alone. We also hypothesized that performance-based contracts might pressure physicians to get results and could lead to mistakes and a worse quality of care. We reviewed nine papers from 2006-2014, of which the majority were conducted in the United States, with one each from Italy and Germany. Our review included three literary/systematic reviews, three before & after studies, and two cross-sectional analyses. Many papers did not provide strong evidence of the effect of pay-for-performance on health outcomes, although examined papers agree on certain things. A common theme we found was the pay-for-performance increased documentation and the number of procedures done, but with minimal conclusions about outcomes. This may result from poor or non-standardized metrics being used to measure and report “performance.” However, two studies found that pay-for-performance improved patient health outcomes, despite limitations. For example, Brosig-Koch et al., (2013) found that P4P was not cost-effective from a solely financial standpoint.

## **Introduction**

For years, the healthcare industry has compensated its providers predominantly using a single model. Whether it be a general practitioner or surgeon, they are usually paid in a fee-for-service format. No matter their performance on a provided service, providers are paid a single fee corresponding to the volume of services they provide. Recently, researchers have contemplated whether there is merit to a performance-based contract for providers. We hypothesize that, on average, across all surgical specialties, performance-based contracts would yield better health outcomes compared to fee-for-service payments. It is also possible, but less probable in our view, that the pressure of having to perform well to be paid can lead to a worse outcome for the patient and potentially lowering the quality of care.

This paper reviews the literature published about the merits of performance-based contracts compared to current fee-for-service models. We reviewed nine papers from 2006-2014. The majority were conducted in the United States, with one each from Italy and Germany. Some papers only reviewed English-speaking papers in their analysis. We had many types of studies in our review. We had three literary/systematic reviews, three before & after studies, and two cross-sectional analyses.

## Methods

We defined an inclusion criteria for the examined studies to be included in our review. Although the inclusion criteria was purposefully kept broad, there were key variables that we established as necessary, which include the following: 1) written in English, 2) academic journal, 3) pay-for-performance (P4P) as an explanatory variable if a research study and either a 4) systematic review, literature review, or research study (i.e., observational or experimental). Studies were removed from being considered in our paper if these inclusion conditions were not met. We also defined specific search words which we believed would generate results aligning with our inclusion criteria. These include “pay-for-performance,” “pay-for-performance & quality of healthcare,” “pay for performance & fee for service,” and “pay for performance & healthcare.” To improve the scientific rigor (i.e., to reduce bias) of our search methodology, we employed two reviewers to examine studies independently.

## Systematic & Literature Reviews (Petersen et al., 2006; Van Herck et al., 2010; Rosenthal & Dudley 2007)

We analyzed three literary/systematic reviews. Two of them (Petersen et al., 2006; Rosenthal & Dudley 2007) examined the link between pay-for-performance and the quality-of-care, while one (Van Herck et al., 2010) observed how designing pay-for-performance contracts could yield certain outcomes that are associated with quality of care. We'll discuss them in the order they were mentioned, starting with Petersen et al., 2006.

In the Petersen et al., 2006 paper, the reviewers concluded that documentation significantly increased in the presence of a financial incentive. Documentation, however, is not a perfect metric of quality of care. Petersen and his team were also able to determine that a bonus of at least 5% of a physician's salary would impact their behavior. They also determined that an intermittent bonus would be more effective at impacting their behavior than a lump sum at the end of the year. The researchers also stated that adverse selection could arise due to these contracts. Adverse selection is when a physician deliberately avoids an extremely sick patient. In this case, physicians would avoid these sick patients because it would derail their chance for a bonus. They were not able to conclude that financial incentives are cost-effective, though. We cannot determine that the results could be applied to a general setting. The researchers speculated that one theoretical advantage of performance pay might be that explicit

financial incentives are provided even when patient demand for healthcare is unresponsive to quality.

The Rosenthal & Dudley 2007 paper was able to draw multiple conclusions. Just like the earlier paper, it was also a literary review. This paper highlights various incentive structures with which P4P can become a successful payment delivery system. It also explores facilitators and barriers/limitations to P4P. The researchers were able to conclude that pay-for-performance could be applied in various ways, such as direct payment, public reporting of performance, administrative simplification programs, etc. The researchers talked about the measures necessary for pay-for-performance programs. They said that to obtain provider-specific results, the cost would be higher. One way that programs evaluate performance is to evaluate insurance claims and the reports of provider quality. Physicians dispute this metric, believing it to be inadequate. For pay-for-performance to be impactful, physicians and other involved parties should find some compromise solutions. The paper also discussed ways of prioritizing disadvantaged patients using pay-for-performance programs. Overall, this paper doesn't evaluate healthcare performance with pay-for-performance. Rather it evaluates different types of P4P programs and the metrics that should be used to evaluate the performance.

In the Van Herck et al., 2010 paper, researchers identified multiple aspects of quality care and observed P4P programs' impacts on each aspect. Pay-for-performance programs were able to increase clinical effectiveness and equity/accessibility of care. Still, they were not shown to increase aspects like coordination and continuity, patient-centeredness, and cost-effectiveness. The effects of P4P could be viewed as either encouraging

or discouraging, depending on whether the P4P program involved minimal quality standards or boosting quality improvement. The researchers proposed that further studies should target P4P programs and the way they define their targets.

## **Before & After Studies (Colais et al., 2013; Long et al., 2012; Shih et al., 2014)**

These studies utilized a before-and-after method to arrive at their results. All three evaluated pay-for-performance programs, and each used a different dependent variable.

The first study that will be examined is Colais et al., 2013, which was conducted in Italy. It monitored P4P and the effect on the proportion of surgeries performed after 48 hours of a hospital arrival. It used an adjusted regression analytical model. The researchers found that the share of patients with hip fractures that had surgery within 48 hours was 11.7% before the introduction of the pay-for-performance act. After the introduction, that share of patients rose to 22.2%. This study, however, does not examine the quality of patient outcomes but focuses on the number of surgeries conducted because of a P4P program. This result is similar to the Petersen et al., 2006 paper, which said that documentation would increase due to P4P incentives.

In the Long et al., 2012 paper, P4P was evaluated by observing its effects on the rate of checks that are processed correctly and the time spent on the check-processing programs. This paper doesn't focus on the healthcare setting, but it's an experimental study that examines how performance incentives can affect outcomes. Initial measurements were taken on the rate of processed checks to be used as a baseline measure. These baseline measures were compared to the measurements taken after the implementation of the P4P and the pay-for-time system. The study had decent internal validity. The researchers concluded that P4P systems produce higher rates of performance and more time on tasks than a pay-for-time system. However, this study has very low generalizability, meaning that the results may not apply to the healthcare setting in which we're discussing our research question.

In the Shih et al., 2014 study, the researchers compared premier hospitals with non-premier hospitals. Premier hospitals are hospitals that use a P4P metric, and non-premier hospitals don't use P4P metrics. In the study, premier hospitals had lower mortality and complication rates for cardiac and orthopedic patients. Between the two types of hospitals, there was no significant difference in mortality for joint replacement operations and CABG (coronary artery bypass graft) surgeries. The strength of the findings is high because the researchers conducted a rigorous study using a control group (non-premier hospitals) which was very similar to the treatment group (premier hospitals). The main difference between these two groups is that the premier hospitals opted into a P4P system, whereas the non-premier hospitals did not.

## **Cross-Sectional Studies (Pierce et al., 2007; Bhattacharyya et al., 2009)**

We used two studies that utilized a Cross-Sectional Analysis method. A Cross-Sectional analysis evaluates a specific target at a certain point in time and its progress now. These studies evaluated P4P and its effects on performance outcomes, as well as specific metrics that should be correlated with performance, like quality and process scores.

The first of these studies was the Pierce et al., 2007 paper. Although this research did not compare to groups to provide evidence suggesting pay-for-performance yields different outcomes than other payment models like a pure fee-for-service scheme, the researchers gathered important information about surgeons' opinions on pay-for-performance, in general. In brief, the researchers conducted their study by organizing a survey, which was given to surgeons. The majority of those surgeons were skeptical of the effectiveness of P4P. Their skepticism was the result of disagreement over which metrics should be used for the measurement of health

outcomes. They also disagreed as to whether these incentives promoted cheating the system. However, most of those surveyed believed orthopedic surgeons should have greater involvement in shaping future P4P policies, and most respondents (78%) were aware of P4P as a new reimbursement strategy. There was also disagreement as to whether these incentives promoted cheating the system.

In the Bhattacharyya et al., 2009 paper, researchers conducted the same cross-sectional analysis method for their study. They evaluated hospitals that used P4P and evaluated their quality scores. A result of P4P programs in hospitals is that volume tended to increase, and that increase of volume correlated heavily to high-quality scores for hospitals. However, they determined that the higher quality and process scores given to surgeons did not correlate with the outcomes of patients. They concluded that current P4P metrics need revision. Researchers also mentioned certain outcomes such as “hip dislocation, 30-day mortality, and one-year reoperation as not being included in performance metrics. Based on this, they concluded that there is a fundamental flaw with the current way “performance” in pay-for-performance programs is being measured.

## **Experimental Design Studies (Brosig-Koch et al., 2019)**

The researchers use an experimental design (i.e., laboratory) to explore the effects of a pay-for-performance payment system (e.g., being paid a bonus for improving patient outcomes) on physician behavior and health outcomes compared to a strictly fee-for-service model as well as sole capitation. When a pay-for-performance payment system is introduced, the study found that physicians are more likely to engage in treatments from which the average patient benefits the most (i.e., optimal treatment). However, the motivation driving this decision-making differs between profit-maximization and benefit-maximization (provision of treatment that is thought to improve patient outcomes the most). However, the relationship (i.e., positive association between pay for performance and patient outcomes) could be more or less strong for physicians of different behavioral or psychological profiles. Additionally, although the researchers found the average patient benefit increases with the introduction of pay-for-performance compared to strictly capitation or fee-for-service, there was no substantial evidence to suggest that the benefits justify the increase in cost according to traditional metrics used in cost-benefit analysis.

According to our research, this is the only study that has attempted to understand the relationship between pay-for-performance incentives and patient outcomes under controlled experimental conditions, which greatly improves the internal validity of the study. However, the generalizability to real-world settings is less persuasive given the variables that would be uncontrolled for.

## **Discussion**

Pay-for-performance, at first glance, seems to be a great solution for the compensation of providers. However, when evaluating pay-for-performance, many issues can be found with the compensation models. Much of the difficulty in using P4P compensation models arises from the performance metric. In the studies discussed, researchers used a myriad of tools that attempted to measure performance. However, as the Pierce et al. 2007 paper mentioned, many surgeons felt that the metrics used in P4P programs undermined the overall effectiveness of the compensation model. Another observable trend was the increase in volume as a result of P4P. Many studies echoed this. Looking first at the Petersen et al., 2006 paper, researchers mentioned that the documentation in hospitals using P4P increased drastically. In the Colais et al., 2013 study, researchers observed that 11.7% of patients that sustained a hip fracture received surgery within 48 hours in hospitals that didn't use P4P compensation. When that hospital introduced the model, that proportion increased to 22.2% of hip-fracture patients receiving surgery in the same timeframe. An adequate way of addressing this is to say that the utilization of healthcare services increases as a direct result of P4P compensation models. With some limitations, P4P

was shown to have effects on quality outcomes. For example, in the Shih et al., 2014 paper, hospitals using P4P were shown to have lower mortality and complication rates for cardiac and orthopedic patients. Lastly, an important trend that we observed in multiple studies – predominantly in Brosig-Koch et al., 2013 – is that, on average P4P models are not cost-effective.

## Conclusion

Overall, while pay-for-performance models are a promising avenue for compensation in the future, some fundamental issues, such as measurement and cost-effectiveness, should be addressed before their widespread use in the healthcare industry. As Figure 1 shows, providers must evaluate and adjust their compensation models before redeploying them in healthcare.



**Figure 1.** Pay for Performance Conceptual Model and Deployment (NEJM, 2018)

## Limitations

Although we worked diligently to compile this literary review, there are still limitations to this analysis. First, we did not exhaust all studies. There are plenty of studies that we haven't discussed or observed. Due to this, there may be more recent studies that potentially have different conclusions than what we determined. Additionally, there could be studies not written in English that we failed to capture, though it is hard to determine whether this would bias our conclusions.

## Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

## References

Bhattacharyya, T., Freiberg, A. A., Mehta, P., Katz, J. N., & Ferris, T. (2009). Measuring the report card: the validity of pay-for-performance metrics in orthopedic surgery. *Health Affairs*, 28(2), 526-532.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Kokot, J., & Wiesen, D. (2019). Physician performance pay: Experimental evidence. Available at SSRN 3467583.  
<https://dx.doi.org/10.2139/ssrn.3467583>

Catalyst, N. E. J. M. (2018). What is pay for performance in healthcare?. *NEJM Catalyst*, 4(2).

Colais, P., Pinnarelli, L., Fusco, D., Davoli, M., Braga, M., & Perucci, C. A. (2013). The impact of a pay-for-performance system on timing to hip fracture surgery: experience from the Lazio Region (Italy). *BMC health services research*, 13, 1-7.  
<https://doi.org/10.1186/1472-6963-13-393>

Long III, R. D., Wilder, D. A., Betz, A., & Dutta, A. (2012). Effects of and preference for pay for performance: An analogue analysis. *Journal of applied behavior analysis*, 45(4), 821-826.  
<https://doi.org/10.1901/jaba.2012.45-821>

Pierce, R. G., Bozic, K. J., & Bradford, D. S. (2007). Pay for performance in orthopaedic surgery. *Clinical Orthopaedics and Related Research*®, 457, 87-95.

Ryan, A. M., & Werner, R. M. (2013). Doubts About Pay-for-Performance in health care. *Harvard Business Review*.

Rosenthal, M. B., & Dudley, R. A. (2007). Pay-for-performance: will the latest payment trend improve care?. *Jama*, 297(7), 740-744.  
doi:10.1001/jama.297.7.740

Shih, T., Nicholas, L. H., Thumma, J. R., Birkmeyer, J. D., & Dimick, J. B. (2014). Does pay-for-performance improve surgical outcomes? An evaluation of phase 2 of the Premier Hospital Quality Incentive Demonstration. *Annals of surgery*, 259(4), 677.  
doi: 10.1097/SLA.0000000000000425

Van Herck, Pieter, et al. "Systematic review: effects, design choices, and context of pay-for-performance in health care." *BMC health services research* 10.1 (2010): 1-13.  
<https://doi.org/10.1186/1472-6963-10-247>

Petersen, L. A., Woodard, L. D., Urech, T., Daw, C., & Sookanan, S. (2006). Does pay-for-performance improve the quality of health care?. *Annals of internal medicine*, 145(4), 265-272.  
<https://doi.org/10.7326/0003-4819-145-4-200608150-00006>