**The Effect of Financial Incentives on Physician Performance**

**Evidence from a Pay-For-Performance Program**

BY

JOSEPH ORNELAS
B.S., University of Illinois at Urbana - Champaign, 2001
M.S., Rush University, 2003
D.C., National University of Health Sciences, 2009
A.M., University of Chicago, 2010

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Public Health Sciences
in the School of Public Health of the
University of Illinois at Chicago, 2015

Chicago, Illinois

Defense Committee:

Tony LoSasso, PhD, Chair and Advisor
Benn Greenspan, PhD, Health Policy and Administration
Edward Mensah, PhD, Health Policy and Administration
Surrey Walton, PhD, Pharmacy Systems, Outcomes, and Policy
Larry Wrobel, DBA, Health Policy and Administration

## ACKNOWLEDGEMENTS

I would like to thank my thesis committee (Tony LoSasso, PhD, Benn Greenspan, PhD, Edward Mensah, PhD, Surrey Walton, PhD, Larry Wrobel, DBA) for their support and contributions to the conduct of the study. Also, I want to especially thank my committee chair (Tony LoSasso, PhD) for his guidance and mentorship throughout the dissertation process.

JO

## TABLE OF CONTENTS

**TABLE OF CONTENTS**

CHAPTER PAGE

## LIST OF TABLES

## LIST OF TABLES

## LIST OF FIGURES

FIGURE PAGE

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ACO | Accountable Care Organization |
| CMS | Center for Medicare and Medicaid Services |
| HbA1c | Hemoglobin A1c |
| LDL-C | Low-Density Lipoprotein Type-C |
| P4P | Pay-For-Performance |
| PCP | Primary Care Physician |
| PHO | Physician-Hospital Organization |
| QOF | Quality and Outcomes Framework |

## SUMMARY

Healthcare pay-for-performance programs use financial incentives to influence provider behavior toward improving quality of care and reducing costs. However, little is known about their efficacy. Motivated by theoretical considerations from physician agency, it is predicted larger amounts of financial incentive will be more effective in improving measureable physician performance and improving patient outcomes. Using physician and patient-level panel data, the following study evaluates financial incentive effects in several key performance areas among physicians in a provider-based, pay-for-performance program. There is suggestive evidence that financial incentives modestly improve measureable physician performance in technological infrastructure and clinical effectiveness quality domains. Also, it is suggested financial incentives improve diabetic outcomes in a primary care setting, particularly among physicians with lower baseline performance.

# 1. INTRODUCTION

## 1.1 Background

Pay-for-performance (P4P) in healthcare is a quality improvement framework that presents financial incentives to physicians so desired processes and outcomes occur through changed behavior (1). P4P is designed to improve patient care by encouraging higher quality and cost-effective health care delivery. In most P4P arrangements, physicians are rewarded with various financial bonuses according to performance evaluations. P4P programs collect and rely on individual-specific data to judge and rate physician performance according to pre-established quality indicators. With the proper incentives in place, it is a fundamental aim of P4P to get physicians to achieve as many applicable performance targets as possible.

## 1.2 Statement of the Problem

P4P has shown potential in changing the way physicians practice toward improving patient outcomes, primarily through increased adherence to evidence-based guidelines (2, 3). P4P is also suggested to curb growth in health care costs and address long-term spending growth (4). Yet, it remains uncertain to what degree physician financial incentives in P4P programs produce positive, measureable results. It has also been suggested physician financial incentives that promote higher quality care are inadequate and lacking (5).

## 1.3 Purpose of the Study

To address the problem, the purpose of this study is to investigate the nature of financial incentives in an established P4P program. This study aims to determine to what degree P4P financial incentives are effective in changing physician performance, examined in measureable ways. This study also explores the nature of patient-outcome measures as conceivable indicators of physician performance in a primary care setting and their clinical implications for diabetic patient populations.

## 1.4 Significance of the Problem

A changing healthcare delivery environment brings significance to the problem. As information technology advances, physician-specific data is increasingly collected and used for profiling performance according to predetermined quality indicators and evidence-based care protocols, rather than volume or intensity of services as seen in a fee-for-service environment (6). With the emergence of P4P, physician compensation is increasingly linked to performance and physicians are under mounting pressure to demonstrate value. Currently, most providers participate in some type of P4P arrangement (7). As P4P continues to expand in prevalence and scope, the provision of financial rewards for attainment of a predetermined level of performance is becoming paramount (8). Furthermore, accountable care organizations (ACOs) are becoming a dominant force in healthcare, bringing increased focus on population health and chronic disease management.

The healthcare policy landscape is also changing. Payment reforms presented as value-based healthcare are widely acknowledged in the Patient Protection and Affordable Care Act of 2008. In response, the Center for Medicare and Medicaid Services (CMS) has implemented quality-based payment programs (e.g. Physician Quality Reporting System) and has undergone several demonstration studies to test its potential effectiveness (5).

## 1.5 Significance of the Study

This study has significance for the provision of healthcare services. Implementers of P4P programs need to be aware of the value associated with physician financial incentives, and where the potential lies to augment physician behavior in specific ways. Implications from the following study can influence or be used to modify the design of P4P program, considering the identified potential physician financial incentives have for cost-control, quality improvement and population health management.

## 2. CONCEPTUAL FRAMEWORK

### 2.1 Incentives for Measureable Performance

First, it is predicted greater financial incentive levels will increase physician measureable performance in a P4P program, in a dose response relationship ($H_1$). Theoretical aspects of physician agency can provide motivation to explore the nature of financial incentives for physicians. In a P4P setting, the agent is typically the physician and the principal is typically the payer (whom acts on behalf of the patient). The payer can be a public payer (e.g. CMS), a private, third party payer (insurance company) or the patients themselves (pay-out of pocket). Physician agency describes a desired outcome $Y$, which represents an observable quality indicator to the principal $i$, for all conditions $k$ ($y_l \ldots y_k$). $Y_k$ can be modeled as (9):

$$1.)\quad Y_k = \mu_k(q_i) + \varepsilon_k$$

Where:

- $q_i$ represents the quality level chosen by the agent (physician), unobservable to the principal (payer). Specifically, $q_i$ represents a vector of efforts chosen in a standard multitasking model (10), which captures several dimensions, including physician integrity, environment, and attention. It is assumed the physician sets $q_i$ below the efficient level. Quality indicator $Y_k$ is dependent on $q_i$, but does not reveal the agent's selection of $q_i$.

- $\mu_k$ represents the actual production of care by the physician. It is assumed $\mu_{jk}$ is the marginal increase in quality indicator $Y_k$ from increase in quality dimension $q_j$ ($\frac{dY_k}{dq_j}$).
- $\varepsilon_k$ represents the exogenous patient risk factors, which may affect quality indicator $Y_k$, but is beyond the direct control of the physician. Examples include patient compliance to treatment (e.g. motivation, effort, values), patient health choices (e.g. eating habits, environmental factors, socioeconomic conditions), or patient reaction to treatment (e.g. concurrent conditions, medication interactions, genetic predisposition).

The payer needs to tactically decide to offer incentives to physicians, in the form of a reimbursement *R(Y)*, which can be modeled as (9):

$$2.)\ R(Y) = r_o + \sum_{k=1}^{K} r_k I\ (Y_k \geq T_k)$$

Where:

- $r_o$ is base salary reimbursement and *R(Y)* does not depend on observable quality indicator $Y_k$, for objective $k = 1 \ldots\ldots K$

- $r_k$ is the add-on bonus reimbursement and *R(Y)* does depend on observable quality indicator $Y_{k.}$. The bonus $r_k$ provided to the agent if $Y_k$ reaches or surpasses a threshold level $T_k$, for objective $k = 1.....K$

Theoretically, both the payer and the physician have competing interests and will seek an agreement to maximize each other's utility. The payer seeks to delegate responsibility to the physician for achieving a desired outcome the payer defines. A well-established formulation of the problem involves specifying the utility of the payer ($U_{payer}$), which can be modeled as:

$$3.)\quad U_{payer} = Y_k - r_k$$

Where:

- $Y_k$ represents an outcome the payer desires
- $r_k$ represents financial expense in the form of reimbursement given to the physician

The physician seeks to receive reimbursement from the payer based on performance toward achieving the outcome, as judged by the payer. The utility of the physician ($U_{physician}$) can be modeled as:

$$4.)\ U_{physician} = (r_k - q_i)^{1/2}$$

Where:

- $r_k$ represents financial gain in the form of reimbursement received from the payer
- $q_i$ represents quality expended by the physician toward achieving the outcome
- the fractional exponent represents risk aversion, typically demonstrated by physicians

The physician will behave in a manner consistent with maximizing utility. According to the utility function of the physician, behavior is influenced by several factors:

1. Amount of reimbursement at stake (may fluctuate by sensitivity to the bonus)
2. Amount of effort towards quality needed to achieve the outcome (may fluctuate by outcome threshold levels, amount of training/preparation, physician motivation)
3. Amount of risk of exogenous shocks (may fluctuate by patient acuity, patient motivation; physician sensitivity to risk may fluctuate based on legal, moral, and professional duty considerations)

Theoretically, the physician seeks to increase utility by maximizing reimbursement and minimizing quality and risk. The payer wants the opposite, mainly to maximize physician quality (reflected in achievement of outcome) and minimize the reimbursement paid.

The financial incentive effect may be overstated among physicians who have a higher intrinsic motivation for patient well-being. Although physicians are motivated to maximize profit, they also value the welfare of their patients and thus are prepared to trade off income against patient well-being (11). As payers increasingly have the ability to control reimbursement and outcome, profit-maximization motivations should become more dominant than intrinsic motivations for patient well being as drivers of physician behavior.

If the payer offers less P4P incentives, the payer may favor risk-averse physicians who prefer a lower steady income to possibly higher income prone to variation. In this arrangement, the payer will also favor physicians who will likely shirk on quality, knowing they do not have to put forth much effort because of less accountability for the outcome. On the other hand, if the payer chooses to offer more P4P incentives, the payer may favor risk-seeking physicians who prefer a possibly higher income prone to variation over a lower steady income. The payer may also favor physicians likely to “cherry pick” patients to attain a favorable outcome and a substantial bonus, yet put forth little effort toward quality.

To decrease the chance of adverse selection, the payer may gather signals of the physician's ability and willingness to expend effort from the physician's past observable outcomes. For example, patient satisfaction scores, physician scorecards, references, credentials, and medical malpractice history may all be useful information for the payer to minimize adverse selection. If the payer receives negative signals from the physician, the payer should not offer P4P incentives. On the other hand, if payer receives positive signals from the physician, the payer should offer more P4P incentives.

To maximize efficiency in the production of the outcome, the payer should gradually offer more P4P incentives until just before the point at which there is such a disutility (from increased quality and risk of reimbursement fluctuation) the physician decides to not participate. Assuming the payer wants the physician to participate, the payer would have to increase minimally acceptable outcomes alongside increases in P4P incentives, so the physician would continue to put forth increasing quality and still decide to participate. The payer would seek to satisfy physician utility until it is greater or equal to his next best option ($\bar{U}$). If the payer does not meet the utility threshold, the physician would not participate and likely pursue another opportunity.

Assuming the physician is risk averse, the physician participation constraint can be modeled as:

$$5.)\ \ U_{physician} = (r_k - q_i)^{1/2} \geq \bar{U}$$

Since the payer has the ability to manipulate reimbursement and outcome, the payer has the opportunity to indirectly influence physician behavior. Assuming the physician behavior is based on effort towards $Y_k$, the physician ultimately has a behavioral choice to either participate or not.

The prediction assumes physicians are rational and self-seeking. The prediction also assumes increases in performance indicator achievement is a result of increased physician quality and the physician has complete information (e.g. physician knows the level of financial incentive, score relative to other physicians, etc.). If physicians do not respond to P4P bonus incentives, the traditional principal agent framework may not apply, and this would have critical implications for the design of P4P programs (12).

## 2.2 Incentives for Performance Indicator Achievement

Second, it is predicted more financial incentive is necessary for physicians to achieve patient outcome measures and induce measureable clinical improvement versus process-of-care measures ($H_2$). Using insights from physician agency as a motivational framework (9), a desired outcome $Y$, represents an observable quality indicator to the principal $i$, for all conditions $k$ ($y_l \dots y_k$). $Y_k$ can be modeled as:

$$6.)\quad Y_k = \mu_k (q_i) + \varepsilon_k$$

Where:

- $q_i$ represents the quality level chosen by the agent (physician), unobservable to the principal (payer). Specifically, $q_i$ represents a vector of efforts chosen in a standard multitasking model(10), which captures several dimensions, including physician integrity, environment, and attention. It is assumed the agent sets $q_i$ below the efficient level. Quality indicator $Y_k$ is dependent on $q_i$, but does not reveal the agent's selection of $q_i$.
- $\mu_k$ represents the actual production of care by the agent (physician). It is assumed $\mu_{jk}$ is the marginal increase in quality indicator $Y_k$ from increase in quality dimension $q_j$ ($\frac{dY_k}{dq_j}$).

- $\varepsilon_k$ represents the exogenous patient risk factors, which may affect quality indicator $Y_k$, but is beyond the direct control of the physician. Examples include patient compliance to treatment (e.g. motivation, effort, values), patient health choices (e.g. eating habits, environmental factors, socioeconomic conditions), or patient reaction to treatment (e.g. concurrent conditions, medication interactions, genetic predisposition).

Patient risk factors $\varepsilon_k$ may be exceedingly responsible for quality indicator $Y_k$, rather than physician quality level $q_i$ alone. The occurrence of exogenous patient risk factors may cause physician quality to become under-rewarded. For example, if a payer contracts on patient blood pressure as a quality indicator, physicians may expend a great deal of effort to attain a favorable observable outcome, but may not achieve the outcome because of a patient's family history of high blood pressure or the negligence of the patient to take prescribed medication.

A moral hazard exists when physicians select more desirable patients based on such risk factors. Physicians are inclined to select the healthiest, most compliant patients and avoid the sickest, high-risk patients in hopes of improving outcomes without expending the effort to produce a high level of quality the principal desires (13). Physician quality may be over-rewarded if this situation occurs and physicians choose the easiest path for achieving the outcome.

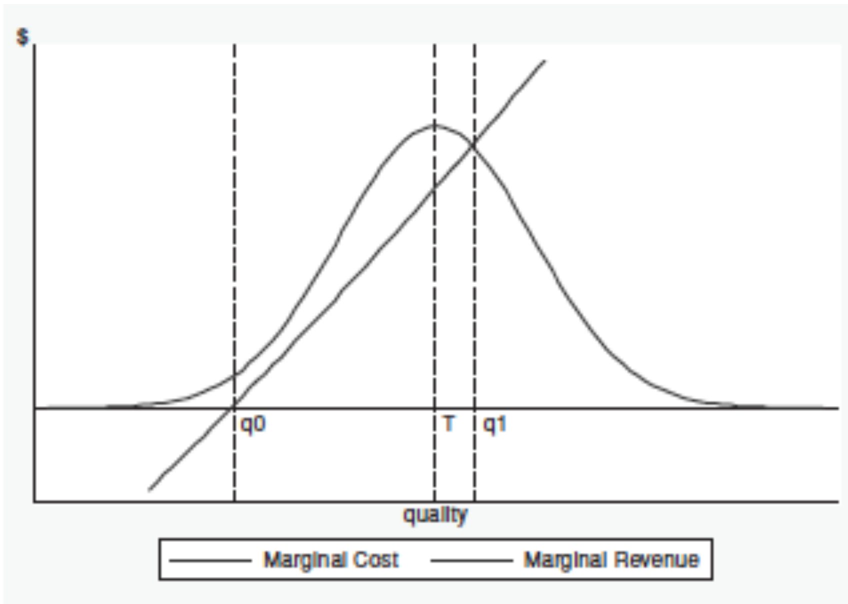
The possibility remains where physicians may treat a disproportionate share of higher risk patients, yet receive lower marks on quality indicators despite making equal efforts to produce high-quality care. It is assumed financial incentives will motivate physicians to choose a higher quality level and it is assumed the principal will not contract with physicians on quality indicators where all patients will most likely respond similarly to treatment and patient outcomes will not vary.

## 2.3 Marginal Incentives for Performance Improvement

Third, it is predicted additional financial incentives are more effective when baseline performance is slightly below quality thresholds versus well below quality thresholds ($H_3$). Target-based P4P programs are criticized for discouraging improvement from very low performers and very high performers (9). It may be argued that paying financial rewards for measureable performance improvement (rather than attainment) would be a more effective method for increasing the quality of clinical care. Yet, a drawback is failing to reward the best providers, where improvement may be more difficult due to ceiling effects (1).

When P4P is introduced, the principal purposely sets target $T_k$ above initial quality $q_0$ to induce quality improvement, marginal revenue greatest at target $T_k$. Under P4P, physician chosen quality increases to chosen quality $q_1$, until marginal cost equals marginal revenue, or where profit is maximized. As the distance between baseline performance and targets increase, there is very little incentive to improve because marginal revenue decreases. On the contrary, as this distance decreases, marginal revenue increases and so does the incentive to improve (9) (Figure 1).

Figure 1: Physician Quality Determination [a]




[a] Used with permission from Mullen et al. (9)

Physicians initially performing furthest below the target may have weak incentives to improve their performance, possibly because the target may seem distant and unreachable. Also, physicians initially performing above the targeted level may have little incentive to improve because they can receive the bonus merely for maintaining performance levels (14).

Since physician utility is maximized when bonus $r_k$ is received and quality $q_i$ expended is minimized, marginal incentive effect is expected to be greatest when $q_0$ is closer to $T_k$, and $q_1$ is just above $T_k$. This prediction assumes physicians are rationally self-seeking and want to maximize profit. The prediction also assumes the incentive is discontinuous around the threshold and the marginal benefit for physicians of improved performance is zero unless the threshold is crossed. This prediction also assumes the bonus is a utility to a physician, and quality is a disutility, according to the physician's utility function:

$$7.)\quad U_{physician} = (r_k - q_i)^{1/2}$$

Overall, this paper offers a greater understanding of physician behavior in an economic context, which is important in relation to pending healthcare policy changes. By determining performance differences among physicians when varying financial incentive amounts and other influences are introduced, this paper has the potential to provide insight and guide payers and providers through the planning and implementation of a successful P4P program as a payment model for physician incentive alignment.

# 3. LITERATURE REVIEW

The effect of P4P incentives on improvement in quality of care has not been convincingly confirmed in the scientific literature. Early studies on P4P comprise of well-conducted trials focusing on the physician financial incentives effect on immunizations. Kouides et al. (15) was one of the first to propose that financial incentives directed at physicians translated into desired practice behavior. This study suggested an increase in financial bonuses improved immunization rates among patients. Despite significant improvements in immunization rates, the incentives were relatively small and the change in immunization rate was relatively modest. Hillman et al. (16, 17) and Fairbrother et al. (18, 19) also considered the effects of financial incentives on immunization rates. Although Hillman et al. (16, 17) did not find any significant improvements in quality, Fairbrother et al. (18, 19) suggested bonuses were slightly more effective in influencing physician behavior than none. Both studies found increases in immunizations, yet little noticeable translation to the actual provision of care. These classic trials did not detect any substantial effect of P4P incentives on improvement of immunization rates. Although convincing, small sample sizes make meeting statistical significance thresholds difficult and their samples cover highly particular settings. It is also unsettling that positive outcomes were linked only to improved documentation rather than actual treatment given.

## 3.1 Reviews Express Limited Evidence on Payment Level for Quality

The current evidence of P4P incentive effectiveness is mixed with varying generalizability. (20, 21) Two more recent reviews address the significance of payment level.

Scott et al. (20) reviewed the literature to examine the effect of changes in the level of payment on the quality of care provided by physicians. Most studies showed positive but very modest effects on measures of quality of care (1, 9, 22-25). Also, Emmert et al. (21) assessed economic evaluations of P4P programs (15, 22, 26-32) and determined economic efficiency for P4P incentives could not be demonstrated.

Overall, these reviews had mixed results, and the narrow range of costs and consequences limited significance and generalizability. Although the results remain mixed, common themes among these studies suggest insufficient incentive amount were responsible for the small improvements gained in quality measures. Larger financial incentives for physicians may be necessary to make significant improvements in quality of care.

The state of the literature implies physician responsiveness to financial incentives is uncertain, and a variety of factors could explain such behavior, including the size and nature of the financial incentive.  More specifically, a dose-response effect for P4P incentives is not evident. It is unclear if a dose-response relationship exists, how efficient P4P incentives translate into improved performance, and whether or not the incentive structure could be altered to achieve similar quality levels with lower cost. A common thread of limitation across recent studies was the size of financial incentive available as inadequate for physicians to put forth the effort required to change practice behavior. Other than the size of the financial incentive, existing evidence implies the true effect of the financial incentive itself may be overstated because of unaccounted for confounding factors (e.g. peer comparison) and potential patient selection by physicians. Also, multifaceted interventions made it difficult to disentangle the pure effects of P4P financial incentives (33). The size of the financial incentive as an exclusive intervention needs to be particularly addressed by controlling for identified observable and unobservable confounding factors. To make a contribution to the literature, this study examined whether a dose-response relationship was present between financial incentives levels and physician performance levels ($Q_1$).

## 3.2 Modest Incentive Effects in Only Some Processes-of-Care

A string of more recent evidence on physicians' responsiveness to financial incentives addressed a spectrum of common disease states and provided more generalizable conclusions to today's healthcare environment. Rosenthal et. al. (1) looked at administrative reports to evaluate P4P initiatives introduced in the California PacifiCare Health Plan and suggested incentives modestly increased cervical cancer screening, but did not effect mammography or Hemoglobin A1c (HbA1c) testing. Similarly, Mullen et al. (9) used performance report data to suggest a small increase in cervical cancer screening, yet no effect on mammography or HbA1c testing.

Some studies were more conclusive. Fagan et al. (34) examined claims data and did not find any significant effects of P4P on quality and resource use measures for patients with diabetes mellitus. Also, Campbell et al. (35) looked at patients with asthma, diabetes, or coronary heart disease before and after incentives were introduced. There was no improvement in quality for asthma or diabetes and quality was reduced for heart disease. Li et al. (33) used administrative data from Ontario to analyze the financial incentive effect on the provision of primary care services. Using a difference-in-differences approach, the results suggested performance improvement with pap smears, mammograms, flu shots, and colorectal cancer screenings. Yet, the effect sizes were small.

There is low evidence either supporting or not supporting the use of financial incentives to improve the processes of care. Despite its growing use, P4P incentives remain controversial due to insufficient and inconclusive evidence on effectiveness and varying generalizability.

## 3.3 Incentive Effects Comparatively Lacking in Patient Outcomes

Most P4P literature documenting the effects of financial incentives on influencing physician behavior is mostly about processes-of-care. There is still comparatively little evidence of P4P successes in improving outcomes or care. Yet, a public health perspective considers how financial incentives influence outcomes of care among a patient population.

There is a string of studies addressing the impact of financial incentives in improving both physician processes and patient outcomes within the same program. Roski et al. (23) tested the effects of the provision of financial incentives on provider adherence to evidence-based practice guidelines. The study suggested physicians introduced to financial incentives significantly improved implementation of the process of smoking cessation counseling. Yet, there was no effect on the outcomes of quitting rates. Similarly, Beaulieu et al. (36) investigated the effects of physician incentives on diabetic quality indicators. Physicians with incentives improved on several process and outcome quality measures for diabetic patients. However, it was concluded that physician selection and low sample size limited the generalizability of the findings. Also, Bardach et al. (37) assessed financial incentive effects on quality improvement among small practices. The P4P program resulted in small improvements in cardiovascular care processes and outcomes compared with usual care. Clinics under P4P had greater improvement in rates of appropriate processes such as antithrombotic prescription and smoking cessation counseling and outcomes like blood pressure control. However, the small clinical settings limit the generalizability of the results.

A comparable group of studies from the Quality and Outcomes Framework (QOF) in the United Kingdom reveal similar results. Serumaga et al. (38) assessed financial incentive impact on hypertension. The study suggested modest, positive effects on testing. Yet, P4P had no apparent effects on hypertension related clinical outcomes. The study concluded financial incentive amount may not be sufficient to effect clinical outcomes. Also, Doran et al. (39) investigated if P4P physicians neglected quality areas with less incentives. The study observed mixed results across services and only some processes measures slightly improved, while most outcome measures did not improve. Substantial improvements from financial incentives were achieved at the expense of other less incentivized aspects of care.

Physician P4P schemes have produced mixed results across studies, which may indicate physicians are more responsive to financial incentives in some areas but not others. This thread of studies suggests the size of financial incentive available may be inadequate for physicians to put forth the effort required to change patient behavior and may imply patient outcomes are more difficult for a physician to achieve than process of care in a P4P program. Other than the size of the financial incentive, existing evidence may indicate the true effect of the financial incentive itself may be overstated because of unaccounted for patient confounding factors.

Gaps in the literature concerning how financial incentives influence physicians to achieve patient outcome measures, besides measures for processes of care, need to be addressed. There is reason to believe physicians with significant incentives would be more likely to respond with higher quality. This study examined how physician financial incentives influence achievement of patient-outcome quality indicators versus process-of-care quality indicators ($Q_2$).

## 3.4 Incentive Effects Vary by Baseline Performance from Target Threshold

There are just a few key studies that serve as the primary foundation of evidence for physician performance relative to thresholds. Most significant is a retrospective study by Rosenthal et al. (1) observed administrative data to assess P4P programs introduced in the California PacifiCare Health Plan. For measures of cervical cancer, mammography or HbA1c testing, physicians initially performing at or above performance targets improved the least but received the largest share of bonus payments. Physicians who were the worst performers at baseline improved the most, despite having almost no financial incentive available. This study demonstrates incentives facing physicians at different levels of baseline performance affect their improvement in performance differently.

Also, a prominent interrupted time-series study by Campbell et al. (35) analyzed incentive data from the QOF among patients with asthma, diabetes, or coronary heart disease. The P4P program improved quality in most chronic conditions. Yet, improvements in quality of care slowed down and the rate declined when the targets were reached. This study demonstrates incentives for physicians performing above targeted levels may have less affect in improving quality than physician performing below quality thresholds.

Finally, a more recent retrospective analysis (Richardson, 2014, unpublished data) analyzed medical record data from the QOF to analyze primary care physician response to financial incentives, focusing on discrete thresholds around which marginal incentives are discontinuous. The results of the study suggest significant behavioral differences when physicians were performing above versus below the diabetic performance thresholds.

The literature suggests that physicians respond differently to financial incentives when baseline performance is presented relative to a threshold and the effect of financial incentives may vary when physicians perform at different levels about quality indicator thresholds. There may be several reasons for this. For example, physicians who exceed a threshold do not receive an additional bonus payment for further improvement beyond the threshold, thus there may exist little motivation to drive performance improvement even further. Still, physicians just above yet close to a threshold may run the risk of losing their payment the next year if their performance slightly slacks and drops below the given threshold, which may creep up over time. Similarly, physicians performing just below a threshold may value a financial reward more than physicians performing well below a threshold, considering the marginal increase in performance is relatively small for receipt of the bonus payment. Those physicians furthest below the threshold may put forth little effort to improve, even though the largest improvements in quality can be gained here. Such observations raise inquiry regarding whether physicians choose performance behavior based on baseline performance in addition to the financial incentive offered. Among patient outcomes, this study examined how financial incentive effects vary by baseline performance from target threshold ($Q_3$).

# 4. METHODS

## 4.1 Setting

Administrative data was used from a P4P program established in 2004, within a large physician-hospital organization (PHO) of a multi-hospital health system in a metropolitan-suburban area. The program was originally developed from a Health Maintenance Organization that used capitation as the model of payment and paid physicians on a per-member, per-month basis. In return, physicians received a steady flow of patients. In this model, the PHO was negotiating with third party payers on behalf of their physician groups (e.g. 150% of Medicare reimbursement).

The P4P program eventually built on the capitation model by increasing financial rewards for physicians who provide higher quality care as measured by process and outcome-based performance indicators. The new arrangement structured the PHO more like an ACO, where population health management was the main focus.

The PHO negotiated a lump sum of money from a third party payer for administering the P4P program. The lump sum was divided on a pro-rated amount, based on how much that insurance company billed out to that physician the previous year. Specifically, the insurance company provided detail to the PHO showing how much they paid to each physician across all payments to the PHO in the previous year. The PHO then used this approach to determine the incentive amount to each physician across the total pool of money. The PHO dispersed the incentive to physicians based on adherence to a broad set of measures of physician performance derived from leading industry groups, including the Joint Commission, the National Committee for Quality Assurance, the Agency for Health Care Research and Quality and the American Medical Association.

## 4.2 Design

The following study of physician performance was conducted using a retrospective, cohort design. Measurements of physician performance, financial incentives, and other descriptive variables were taken annually at the end of each year. The incentive amount served as the intervention and there was no parallel group of physicians to serve as a control. Because incentives varied in size, the same cohort of physicians also served as their own control.

All physicians in the PHO were automatically included in the P4P program (and thus the dataset). Therefore, PHO membership requirements served as the inclusion criteria for the sample:

- High speed internet access capable
- Intranet portal account usage
- Active participation in the PHO
- Physician medical staff membership at a partner hospital
- Cooperation with providing claims information

- Email address usage for physician and office manager
- Completion of core patient safety continuing medical education
- Attendance at new physician orientation
- Satisfaction of board certification criteria
- Participation agreement signature
- One observation during the previous calendar year on at least one performance measure

Each year, physicians could join and leave the program, providing an unbalanced panel dataset. Physicians were excluded from the program every year either because they electively opted out of the program if they had a conflict of interest, or they joined the program after the cutoff date of July $1^{st}$. If physicians joined the program before July $1^{st}$, their data was included for that calendar year. If they joined after July $1^{st}$, their data was not included. Physicians who left the program were removed from the dataset.

## 4.3 Samples

### 4.3.1 Physicians of All Specialties

To address the first research question ($Q_1$), a sample of physicians was evaluated. The physicians were incentivized with P4P financial bonuses during the first five years of the program's existence, 2004 - 2008. Physicians varied in practicing specialties, professional designations and practicing hospital location. The sample consisted of roughly 2,500 physicians in 2004, which grew to nearly 2,700 physicians by the end of 2008. A unique physician identifier allowed physician tracking over time. The physicians practiced 77 different specialties (top 10 listed); internal medicine and family medicine was the most common, respectively. The physicians practiced at 7 different hospitals; Hospitals *A*, *E* and *D* were the most common, respectively. An overwhelming majority of physicians had a Doctor of Medicine degree compared to a Doctor of Osteopathy degree (Table I).

**TABLE I**

DESCRIPTION OF PHYSICIAN SAMPLE BY TOP SPECIALTIES, DEGREE, AND HOSPITAL, 2004 – 2008

| | **2004** *n=2,437* | | **2005** *n=2,487* | | **2006** *n=2,558* | | **2007** *n=2,590* | | **2008** *n=2,661* | | **All Years** *n=12,773* | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Count** | **Percent** | **Count** | **Percent** | **Count** | **Percent** | **Count** | **Percent** | **Count** | **Percent** | **Count** | **Percent** |
| Internal Medicine | 292 | 12% | 298 | 12% | 307 | 12% | 285 | 11% | 293 | 11% | 1,405 | 11% |
| Family Medicine | 195 | 8% | 199 | 8% | 205 | 8% | 207 | 8% | 213 | 8% | 1,022 | 8% |
| Pediatrics | 146 | 6% | 174 | 7% | 205 | 8% | 207 | 8% | 186 | 7% | 894 | 7% |
| Cardiology | 0 | 0% | 149 | 6% | 153 | 6% | 130 | 5% | 186 | 7% | 639 | 5% |
| Obstetrics and Gynecology | 195 | 8% | 174 | 7% | 179 | 7% | 181 | 7% | 186 | 7% | 894 | 7% |
| Emergency Medicine | 97 | 4% | 99 | 4% | 128 | 5% | 104 | 4% | 133 | 5% | 511 | 4% |
| Anesthesiology | 97 | 4% | 99 | 4% | 51 | 2% | 52 | 2% | 53 | 2% | 383 | 3% |
| Orthopedic Surgery | 97 | 4% | 99 | 4% | 102 | 4% | 104 | 4% | 106 | 4% | 511 | 4% |
| Ophthalmology | 97 | 4% | 99 | 4% | 102 | 4% | 104 | 4% | 106 | 4% | 511 | 4% |
| Diagnostic Radiology | 73 | 3% | 75 | 3% | 77 | 3% | 78 | 3% | 80 | 3% | 383 | 3% |
| Doctor of Osteopathy | 97 | 4% | 149 | 6% | 153 | 6% | 130 | 5% | 160 | 6% | 639 | 5% |
| Doctor of Podiatric Medicine | 49 | 2% | 50 | 2% | 51 | 2% | 52 | 2% | 53 | 2% | 255 | 2% |
| Doctor of Medicine | 2,266 | 93% | 2,288 | 92% | 2,353 | 92% | 2,383 | 92% | 2,448 | 92% | 11,751 | 92% |
| Hospital A | 609 | 25% | 597 | 24% | 614 | 24% | 622 | 24% | 639 | 24% | 3,066 | 24% |
| Hospital B | 268 | 11% | 274 | 11% | 281 | 11% | 285 | 11% | 319 | 12% | 1,405 | 11% |
| Hospital C | 268 | 11% | 274 | 11% | 307 | 12% | 311 | 12% | 319 | 12% | 1,533 | 12% |
| Hospital D | 366 | 15% | 373 | 15% | 409 | 16% | 414 | 16% | 426 | 16% | 2,044 | 16% |
| Hospital E | 609 | 25% | 597 | 24% | 588 | 23% | 622 | 24% | 612 | 23% | 3,066 | 24% |
| Hospital F | 195 | 8% | 199 | 8% | 205 | 8% | 181 | 7% | 213 | 8% | 1,022 | 8% |
| Hospital G | 122 | 5% | 149 | 6% | 128 | 5% | 130 | 5% | 160 | 6% | 639 | 5% |

Physician performance was categorized into five quality domains:

1. A technological infrastructure domain included process-based measures related to participation and adoption of technological resources available from the health system (e.g. usage of an electronic medical record, populating registry databases).
2. An efficiency domain included process-based measures related to cost effective practices and resourcefulness (e.g. generic drug usage, electronic claims submission).
3. A patient safety domain included process-based measures related to physician practices proven to decrease risk to the patient (e.g. admission protocol used, patient communication tactics, patient safety checklist).
4. A patient experience domain included process-based measures related to reinforcement of best practices aimed to give the patient a positive experience in the healthcare environment (e.g. usage of pre-determined patient handoff procedures, patient satisfaction scores)
5. A clinical effectiveness domain included both process and outcome-based measures gathered to address a variety of common patient scenarios, which specifically include smoking cessation, coronary artery disease, diabetes care, asthma, congestive heart failure, childhood immunization activity, and depression screening.

Each domain contained several performance categories (Table II). Each performance category contained several individual performance measures, and different sets of measures were provided to different specialties of physicians. If a physician was provided with a measure, they were compelled to achieve the best outcome with their applicable patients. If the physician did not have any applicable patients for that measure, they were automatically awarded the available points for that measure.

**TABLE II**
PERFORMANCE CATEGORIES BY DOMAIN

| **Domain** | **Performance Categories** |
|---|---|
| Infrastructure | Electronic Intensive Care Unit Participation |
| | Electronic Medical Record Usage |
| | Email Access |
| | High Speed Internet Access |
| | Quality Improvement Registries Access |
| | Quality Improvement Registries Populated |
| | Medicare "Quality" Coding Usage |
| Effectiveness | Smoking Cessation Advice Offered |
| | Coronary Artery Disease Testing |
| | Diabetes Care Testing |
| | Asthma Testing |
| | Congestive Heart Failure Testing |
| | Childhood Immunization Activity |
| | Depression Screening |
| Efficiency | Effective Use of Resources |
| | Average Length of Stay (Days) |
| | Generic Drug Usage |
| | Hospitalist Attestation |
| | Lab Outreach |
| | Quality Improvement Activity |
| | Physician Roundtable Participation |
| | Patient Safety Continuing Medical Education |
| | Computerized Physician Order Entry Usage |
| | Peer Satisfaction Survey |
| | Electronic Claims Submission |
| Safety | Average Re-Admissions |
| | Competency Evaluation |
| | Physician Safety Questionnaire |
| | Communication Efficiency |
| | Patient Safety Rating |
| Experience | Efficient Handoffs |
| | Satisfaction Specialty Care |
| | Satisfaction in Emergency Department |

The amount of points available to each physician varied by performance domain and by year. Over time, the program added new domains and gradually increased the amount of points available for each domain. The most available points on average were among the efficiency and infrastructure domains respectively. Among all domains, the amount of available points generally increased each year from 2005-2008, with a slight decrease beginning in 2004-2005 (Table III).

**TABLE III**
POINTS AVAILABLE PER DOMAIN, 2004 – 2008

| Year | | Infrastructure | Effectiveness | Efficiency | Safety | Experience | All Initiatives |
|---|---|---|---|---|---|---|---|
| **2004** | **Mean** | **3.98** | **2.28** | **4.70** | **.** | **.** | **9.57** |
| *n=2,437* | *SD* | *1.04* | *0.92* | *1.31* | *.* | *.* | *1.76* |
| **2005** | **Mean** | **3.45** | **2.81** | **3.01** | **.** | **.** | **7.53** |
| *n=2,487* | *SD* | *0.53* | *1.14* | *0.47* | *.* | *.* | *2.19* |
| **2006** | **Mean** | **4.90** | **3.67** | **3.09** | **1.05** | **.** | **9.96** |
| *n=2,558* | *SD* | *0.43* | *1.66* | *1.19* | *0.36* | *.* | *3.01* |
| **2007** | **Mean** | **4.37** | **4.02** | **4.59** | **1.58** | **1.00** | **13.48** |
| *n=2,590* | *SD* | *0.66* | *2.09* | *1.43* | *0.91* | *0.00* | *4.42* |
| **2008** | **Mean** | **4.43** | **4.62** | **6.06** | **3.33** | **1.03** | **17.21** |
| *n=2,661* | *SD* | *0.76* | *3.05* | *2.12* | *1.69* | *0.17* | *6.05* |
| **All Years** | **Mean** | **4.24** | **3.72** | **4.30** | **2.21** | **1.01** | **11.64** |
| *n=12,773* | *SD* | *0.86* | *2.31* | *1.82* | *1.57* | *0.11* | *5.18* |

Similarly, each domain had available financial incentive, which grew over time and varied by physician. The most available incentive on average was among the infrastructure and efficiency domains, respectively. Among all domains, the amount of available incentive has increased each year from 2004 - 2007, with the largest increase being between 2004 - 2005 and 2005 - 2006 respectively (Table IV).

**TABLE IV**
INCENTIVE AVAILABLE PER DOMAIN, 2004 – 2008

| Year | | Infrastructure | Effectiveness | Efficiency | Safety | Experience | All Initiatives |
|---|---|---|---|---|---|---|---|
| **2004** | **Mean** | **$401.31** | **$241.15** | **$494.22** | | | **$1,136.67** |
| *n=2,437* | *SD* | *$562.94* | *$363.31* | *$683.68* | | | *$597.77* |
| **2005** | **Mean** | **$1,668.35** | **$831.80** | **$1,525.14** | | | **$4,025.29** |
| *n=2,487* | *SD* | *$1,368.23* | *$609.62* | *$1,230.87* | | | *$1,250.61* |
| **2006** | **Mean** | **$3,198.08** | **$1,418.46** | **$2,040.92** | **$552.01** | | **$7,209.47** |
| *n=2,558* | *SD* | *$2,773.31* | *$1,052.02* | *$1,673.52* | *$661.45* | | *$2,184.75* |
| **2007** | **Mean** | **$2,666.34** | **$1,668.54** | **$2,727.35** | **$961.67** | **$752.41** | **$8,776.31** |
| *n=2,590* | *SD* | *$2,383.36* | *$1,325.21* | *$2,394.49* | *$1,025.26* | *$623.61* | *$2,028.55* |
| **2008** | **Mean** | **$2,323.97** | **$1,664.61** | **$2,994.84** | **$1,797.64** | **$624.13** | **$9,405.19** |
| *n=2,661* | *SD* | *$2,167.41* | *$1,344.33* | *$2,854.97* | *$1,847.32* | *$506.03* | *$2,195.32* |
| **All Years** | **Mean** | **$2,073.18** | **$1,303.47** | **$1,976.09** | **$1,236.78** | **$698.46** | **$7,287.98** |
| *n=12,773* | *SD* | *$2,239.05* | *$1,226.75* | *$2,139.45* | *$1,471.37* | *$580.36* | *$1,958.52* |

The incentive amount was eventually dispersed to the physicians as a bonus, which can be modeled as:

$$8.)\ Bonus_{it} = Incentive_{it} * \%PtsEarned_{it}$$

Where:

- $\%PtsEarned_{it}$ is the percentage of available points earned

Physician performance, in terms of the percentage of available points earned, was measured by calendar year. Based on the performance recorded as of December 31st, the bonuses were distributed to physicians in April of the following year.

### 4.3.2 Primary Care Physicians with Diabetic Patients

In certain conditions, the P4P program provided rewards to physicians when they had zero applicable patients for a group of measures. Thus, the design of the program created an incentive to not treat patients with certain disease states measured in the program (e.g. diabetes, coronary artery disease) It is assumed each physician specialty would need to draw on particular groups of patients by the nature of their practice. Primary care was a broad specialty that served several types of disease states within the clinical effectiveness domain. Therefore, the performance of only primary care physicians (PCPs) with patients was evaluated. Since this approach focused on PCPs matched with patients, the potential for selecting or "dumping" patients was not an issue with this approach.

To address the second research question ($Q_2$), diabetic care was chosen as a "drilled-down" research area. Besides having significant public health implications (40), diabetic care included the greatest number of physicians participating in the program over time: PCPs. Additionally, the data available for diabetes was succinct and clearly expressed with a consistent and even combination of process and outcome measures.

A subset of the original dataset was used, consisting of roughly 400 PCPs specializing in family medicine or internal medicine serving adult diabetic patients from years 2005-2008. Pediatrics and geriatrics were also considered primary care, but not included in the sample because their treated diabetic patients were below the age of 18 or above the age of 65. The performance measures used in the dataset were only applicable to adult diabetic patients between the ages of 18 and 65.

From 2005 to 2008, the number of PCPs in the sample gradually increased. There were slightly more internal medicine physicians compared to family medicine physicians, and substantially more physicians with an medical doctor degree compared to a doctor of osteopathy degree. The PCPs practiced at 7 different hospitals. The largest groups of physicians were from Hospitals A, E and D, respectively (Table V).

**TABLE V**
PRIMARY CARE PHYSICIAN SAMPLE BY SPECIALTY, DEGREE, AND HOSPITAL, 2005 - 2008

| | **2005** *n= 342* | | **2006** *n= 394* | | **2007** *n= 412* | | **2008** *n= 428* | | **All Years** *n= 1576* | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Count** | **Percent** | **Count** | **Percent** | **Count** | **Percent** | **Count** | **Percent** | **Count** | **Percent** |
| Family Medicine | 137 | 40% | 154 | 39% | 169 | 41% | 176 | 41% | 636 | 40% |
| Internal Medicine | 205 | 60% | 240 | 61% | 243 | 59% | 252 | 59% | 940 | 60% |
| Doctor of Medicine | 316 | 92% | 362 | 92% | 377 | 92% | 391 | 91% | 1446 | 92% |
| Doctor of Osteopathy | 26 | 8% | 32 | 8% | 35 | 8% | 37 | 9% | 130 | 8% |
| Hospital A | 91 | 27% | 108 | 27% | 108 | 26% | 110 | 26% | 417 | 26% |
| Hospital B | 31 | 9% | 36 | 9% | 37 | 9% | 37 | 9% | 141 | 9% |
| Hospital C | 22 | 6% | 27 | 7% | 32 | 8% | 37 | 9% | 119 | 8% |
| Hospital D | 49 | 14% | 64 | 16% | 76 | 18% | 78 | 18% | 267 | 17% |
| Hospital E | 83 | 24% | 86 | 22% | 86 | 21% | 90 | 21% | 345 | 22% |
| Hospital F | 38 | 11% | 44 | 11% | 47 | 11% | 49 | 11% | 178 | 11% |
| Hospital G | 27 | 8% | 29 | 7% | 26 | 6% | 27 | 6% | 109 | 7% |

PCPs in the sample have gradually acquired more diabetic patients over the same time period. By 2008, over half of the PCPs had greater than 10 diabetic patients under their care. Also, and increasing number of PCPs had acquired more than 40 diabetic patients (Table VI).

**TABLE VI**
COUNT OF DIABETIC PATIENTS PER PRIMARY CARE PHYSICIAN, 2005 – 2008

| | **2005** *n= 342* | | **2006** *n= 394* | | **2007** *n= 412* | | **2008** *n= 428* | |
|---|---|---|---|---|---|---|---|---|
| **Number of Patients** | **Number of Physicians** | **Percent** | **Number of Physicians** | **Percent** | **Number of Physicians** | **Percent** | **Number of Physicians** | **Percent** |
| 1 to 10 | 291 | 85% | 298 | 76% | 236 | 57% | 211 | 49% |
| 11 to 20 | 43 | 13% | 83 | 21% | 113 | 27% | 116 | 27% |
| 21 to 30 | 6 | 2% | 8 | 2% | 44 | 11% | 55 | 13% |
| 31 to 40 | 2 | 1% | 3 | 1% | 12 | 3% | 30 | 7% |
| Greater than 40 | 0 | 0% | 2 | 1% | 7 | 2% | 16 | 4% |

As an in-depth analysis, the physician-level data subset was expanded to the patient-level (be means of matched registry data) to observe clinical outcome levels of each diabetic patient. From 2005 to 2008, more diabetic patients were gradually seen by PCPs in the program. By 2008, PCPs were seeing over 6,000 adult diabetic patients, mostly aged 45-64 years (Table VII).

**TABLE VII**
DESCRIPTION OF DIABETIC PATIENT SAMPLE BY AGE, 2005 – 2008

| | **2005** *n=1,987* | | **2006** *n=2,813* | | **2007** *n=4,681* | | **2008** *n=6,084* | | **All Years** *n=15,565* | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Age** | **Number of Patients** | **Percent** | **Number of Patients** | **Percent** | **Number of Patients** | **Percent** | **Number of Patients** | **Percent** | **Number of Patients** | **Percent** |
| 18 to 24 | 20 | 1% | 19 | 1% | 28 | 1% | 29 | 0% | 96 | 1% |
| 25 to 34 | 83 | 4% | 100 | 4% | 142 | 3% | 171 | 3% | 496 | 3% |
| 35 to 44 | 225 | 11% | 287 | 10% | 486 | 10% | 589 | 10% | 1587 | 10% |
| 45 to 54 | 527 | 27% | 727 | 26% | 1233 | 26% | 1577 | 26% | 4064 | 26% |
| 55 to 64 | 740 | 37% | 1105 | 39% | 1866 | 40% | 2426 | 40% | 6137 | 39% |
| 65 to 74 | 355 | 18% | 522 | 19% | 837 | 18% | 1141 | 19% | 2855 | 18% |
| 75 to 84 | 37 | 2% | 53 | 2% | 89 | 2% | 147 | 2% | 326 | 2% |
| > 85 | 0 | 0% | 0 | 0% | 0 | 0% | 4 | 0% | 4 | 0% |

Over the four years studied, there were a total of 54 performance measures available to PCPs, 12 of which were specific to diabetic care. The series of diabetic performance measures presented to PCPs were categorized under the clinical effectiveness initiative of the program and consisted of both process and outcome-based measures. All the measures addressed adult patients (between 18 and 65 years old) diagnosed with either Type 1 or Type 2 diabetes. Also, the measures were produced in agreement with evidence-based clinical practice guidelines from a variety of leading clinical associations (Table VIII).

**TABLE VIII**

DIABETIC MEASURES BY CLINICAL GUIDELINE RECOMMENDATION AND DEVELOPER [a]

| Diabetic Measure | Clinical Guideline Recommendation | Developer |
|---|---|---|
| HbA1c Testing | HbA1c test is recommended during an initial assessment and during follow-up assessments. | American Association of Clinical Endocrinologists, American College of Endocrinology, American Diabetes Association |
| LDL-C Screening | Fasting lipid profile is recommended during an initial assessment and follow-up assessments. | |
| Eye Examination | Dilated eye exam is recommended during an initial assessment and at least annually thereafter. | American Academy of Ophthalmology, American College of Endocrinology, American Diabetes Association, American Optometric Association |
| Monitoring Nephropathy | Urinalysis including micro albuminuria and creatinine clearance, is recommended as part of an initial assessment and annually thereafter. | American Association of Clinical Endocrinologists, American College of Endocrinology, American Diabetes Association, National Kidney Foundation |
| Smoking Cessation Counseling | Smoking cessation counseling is recommended as part of optimal care of the patient with diabetes. | American Diabetes Association |
| Foot Monitoring | Complete foot exam is recommended during an initial assessment and follow-up assessments. | |
| Good HbA1c Control | HbA1c is ≤ 7% is recommended. | |
| Good LDL-C Control | LDL cholesterol is <100 mg/dl is recommended. | National Cholesterol Education Program |
| Hypertension Control | Blood pressure determination is recommended during an initial assessment and follow-up assessments. | American Association of Clinical Endocrinologists, American Diabetes Association |
| Pre-Hypertension Control | Blood pressure is <130/80 mm Hg is recommended. | |

[a] accessed from www.guideline.gov on January 15th, 2015

The diabetic performance measures were based on percentages. For process-based measures, the numerator consisted of the physician performing a specific action during a patient encounter (e.g. prescribed a medication, checked blood pressure). For outcome-based measures, the numerator consisted of the physician's patient meeting or exceeding a minimum clinical outcome threshold (e.g. $HbA1c < 7$). For both types of measures, the denominator in the performance measures was the total number of patients with the relevant condition seen by the physician (Table IX).

**TABLE IX**
DEFINITION OF DIABETIC MEASURES

| Measure Type | Diabetic Measure | Definition [a] |
|---|---|---|
| Process | HbA1c Testing | Received at least one HbA1c test during the measurement year |
| | LDL-C Screening | Received at least one lipid profile test during the measurement year |
| | Eye Examination | Received at least one dilated retinal exam with interpretation during the measurement year |
| | Monitoring Nephropathy | Received at least one test for micro albumin during the measurement year |
| | Smoking Cessation Counseling | Offered an intervention for smoking cessation at least once during the measurement year |
| | Foot Monitoring | Received at least one complete foot exam in the measurement year |
| Outcome | Good HbA1c Control (<7) | Most recent hemoglobin HbA1c test during the measurement year was < 7.0% |
| | Poor HbA1c Control (>9) | Most recent hemoglobin HbA1c test during the measurement year was > 9.0% |
| | Good LDL-C Control (<100) | Most recent low density lipoprotein cholesterol test during the measurement year was < 100 mg/dL |
| | Poor LDL-C Control (>130) | Most recent low density lipoprotein cholesterol test during the measurement year was >130 mg/dL |
| | Hypertension BP Control (<140/90) | Most recent blood pressure reading during the measurement year was < 140 / 90 mm Hg |
| | Pre-Hypertension BP Control (<130/80) | Most recent blood pressure reading during the measurement year was < 130 / 80 mm Hg |

[a] accessed from www.qualityforum.org on March 1st, 2015

The calculated percentage of each performance measure was compared against a threshold predetermined by the P4P program as minimally acceptable performance. PCPs were awarded the available incentive dollars in the form of a bonus (unique for each measure and each year) if their performance was above the threshold. PCPs received no payment for performance below the threshold.  Yet, PCPs did not receive additional payment for performance exceeding the threshold. As a means of quality improvement, the thresholds of most diabetic performance measures gradually increased over time and new performance measures were added each year in effort to “raise the quality bar” and gradually make achievement more difficult for PCPs over time (Table X).

**TABLE X**

DEFINITION OF DIABETIC MEASURES BY PATIENT THRESHOLDS, 2005 - 2008

| Measure Type | Diabetic Measure | 2005 | 2006 | 2007 | 2008 |
|---|---|---|---|---|---|
| Process | HbA1c Testing | >=75% | >=78% | >=81% | >=81% |
| | LDL-C Screening | >=73% | >=76% | >=79% | >=79% |
| | Eye Examination | >=43% | >=46% | >=49% | >=50% |
| | Monitoring Nephropathy | . | >=70% | >=60% | >=60% |
| | Smoking Cessation Counseling | . | . | >=73% | >=76% |
| | Foot Monitoring | . | . | . | >=50% |
| Outcome | Good HbA1c Control (<7) | >=23% | >=26% | >=29% | >=32% |
| | Poor HbA1c Control (>9) | <=46% | <=43% | <=40% | <=40% |
| | Good LDL-C Control (<100) | >=40% | >=43% | >=46% | >=46% |
| | Poor LDL-C Control (>130) | <=42% | <=39% | <=36% | <=36% |
| | Hypertension BP Control (<140/90) | . | . | >=40% | >=46% |
| | Pre-Hypertension BP Control (<130/80) | . | . | >=20% | >=23% |

Over time, there was more financial incentive available for process measures compared to outcome measures. From 2005 - 2007, the amount of incentive available for all measures decreased. From 2007 to 2008, the amount of incentive available increased, with the most increase being among process measures (Table XI).

**TABLE XI**
INCENTIVE AVAILABLE PER DIABETIC MEASURE TYPE, 2005 - 2008

| Year | | Process Measures | Outcome Measures | All Measures |
|---|---|---|---|---|
| **2005** | **Mean** | **$72.94** | **$37.15** | **$52.49** |
| *n= 342* | *SD* | *$51.59* | *$28.20* | *$43.68* |
| **2006** | **Mean** | **$60.60** | **$40.50** | **$50.55** |
| *n=394* | *SD* | *$48.44* | *$27.80* | *$40.75* |
| **2007** | **Mean** | **$45.38** | **$44.75** | **$45.04** |
| *n=412* | *SD* | *$36.51* | *$35.72* | *$36.08* |
| **2008** | **Mean** | **$85.71** | **$51.54** | **$68.63** |
| *n=428* | *SD* | *$66.32* | *$35.64* | *$55.91* |
| **All Years** | **Mean** | **$66.93** | **$44.80** | **$55.32** |
| *n=1576* | *SD* | *$55.69* | *$33.47* | *$46.73* |

Specific to outcome measures, the incentive available to PCPs was slightly higher among the blood pressure measures compared to the HbA1c and low-density lipoprotein-C (LDL-C) measures. Also, the incentive amounts gradually increased over time among all outcome measures (Table XII).

**TABLE XII**

INCENTIVE AVAILABLE PER DIABETIC OUTCOME MEASURES, 2005 - 2008

| **Year** | | **Good HbA1c Control (<7)** | **Poor HbA1c Control (>9)** | **Good LDL-C Control (<100)** | **Poor LDL-C Control (<130)** | **Hypertension BP Control (<140/90)** | **Pre-Hypertension BP Control (<130/80)** |
|---|---|---|---|---|---|---|---|
| **2005** | **Mean** | **$36.60** | **$37.66** | **$36.71** | **$37.64** | | |
| *n= 342* | *SD* | *$26.90* | *$29.54* | *$26.80* | *$29.53* | | |
| **2006** | **Mean** | **$40.49** | **$40.50** | **$40.44** | **$40.57** | | |
| *n=394* | *SD* | *$27.90* | *$27.82* | *$27.82* | *$27.76* | | |
| **2007** | **Mean** | **$44.35** | **$44.27** | **$44.27** | **$44.38** | **$46.98** | **$44.27** |
| *n=412* | *SD* | *$35.47* | *$35.55* | *$35.55* | *$35.49* | *$36.82* | *$35.55* |
| **2008** | **Mean** | **$51.10** | **$51.44** | **$51.44** | **$51.44** | **$52.42** | **$51.44** |
| *n=428* | *SD* | *$35.43* | *$35.66* | *$35.66* | *$35.66* | *$35.95* | *$35.66* |
| **All Years** | **Mean** | **$43.54** | **$43.84** | **$43.62** | **$43.88** | **$49.75** | **$47.92** |
| *n=1576* | *SD* | *$32.37* | *$32.92* | *$32.45* | *$32.88* | *$36.46* | *$35.77* |

To address the third research question ($Q_3$), PCPs from the data subset were placed into percentiles and ranked based on where they performed relative to each other. For all diabetic outcome measures, the distance from the threshold based upon baseline percent achievement of a physician's patients the previous year is represented as:

$$9.)\ f\,(Percentile\_Rank_{kt}) = Y_{ikt-1} - T_{kt}$$

Between the individual years from 2005 to 2008, PCPs with a lower percentile rank the previous year had more marginal incentive available to them the following year for all outcome measures and the measure specific to "Good HbA1c Control (<7)" (Tables XIII and XIV).

**TABLE XIII**
MARGINAL INCENTIVE AVAILABLE BY INITIAL PERFORMANCE QUARTILE, ALL OUTCOME MEASURES, 2005 - 2008

| Year | | <0.25 | >=0.25 to <0.50 | >=0.50 to <0.75 | >=0.75 |
|---|---|---|---|---|---|
| **2005 - 2006** | **Mean** | **$5.98** | **$5.31** | **$6.18** | **$6.44** |
| | *SD* | *$11.22* | *$8.88* | *$9.19* | *$9.03* |
| **2006 - 2007** | **Mean** | **$5.10** | **$3.52** | ***$3.45*** | ***$4.58*** |
| | *SD* | *$13.57* | *$13.49* | *$15.90* | *$13.54* |
| **2007 - 2008** | **Mean** | **$10.54** | **$6.60** | ***$6.88*** | ***$5.79*** |
| | *SD* | *$20.50* | *$20.08* | *$22.04* | *$33.10* |
| **All Years** | **Mean** | **$7.32** | **$5.45** | ***$5.76*** | ***$5.62*** |
| | *SD* | *$15.96* | *$15.95* | *$18.34* | *$24.98* |

**TABLE XIV**
MARGINAL INCENTIVE AVAILABLE FOR PERCENTILE RANK PREVIOUS YEAR, GOOD HBA1C CONTROL (<7) MEASURE, 2005 - 2008

| Year | | <0.25 | >=0.25 to <0.50 | >=0.50 to <0.75 | >=0.75 |
|---|---|---|---|---|---|
| **2005 - 2006** | **Mean** | **$6.58** | **$4.84** | **$7.23** | **$4.49** |
| | *SD* | *$11.56* | *$6.96* | *$9.84* | *$7.96* |
| **2006 - 2007** | **Mean** | **$5.20** | **$2.96** | ***$2.14*** | ***$6.88*** |
| | *SD* | *$13.37* | *$12.39* | *$15.30* | *$15.60* |
| **2007 - 2008** | **Mean** | **$12.79** | **$7.75** | ***$5.86*** | ***$3.74*** |
| | *SD* | *$21.68* | *$18.01* | *$17.82* | *$34.88* |
| **All Years** | **Mean** | **$7.27** | **$5.22** | ***$4.90*** | ***$4.66*** |
| | *SD* | *$15.00* | *$13.43* | *$15.52* | *$26.91* |

## 4.4 Estimation Models

### 4.4.1 Percentage of Quality Domain Points Earned

To estimate physician performance on different levels, several models were considered. First, a model of the degree of physician achievement of quality domains over time was estimated. A sequential regression technique was applied to demonstrate the value of controlling for physician descriptive factors and utilizing physician specific fixed effects. The first model included incentive amount with no physician descriptive variables, the second model included physician descriptive variables, and the third model included physician fixed effects. The estimation of three different models tested the sensitivity of the main coefficient of interest (on the incentive amount variable) to the inclusion of relevant omitted variables. To the extent that the coefficient changed suggests there might be important unobservable variables that are correlated with both the outcome and the incentive amount. The sequential estimation models of percentage of domain points earned were as follows:

$$10.)\quad \%PtsEarned_{it} = \beta_0 + \beta_1 Incentive_{it} + \beta_2 Year_t + \varepsilon_{it}$$

$$11.)\quad \%PtsEarned_{it} = \beta_0 + \beta_1 Incentive_{it} + \beta_2 Year_t + \beta_3 Hospital_i + \beta_4 Degree_i + \beta_5 Specialty_i + \varepsilon_{it}$$

$$12.)\quad \%PtsEarned_{it} = \beta_0 + \beta_1 Incentive_{it} + \beta_2 Year_t + \alpha_i + \varepsilon_{it}$$

### 4.4.2 Diabetic Measure Type Achievement

In the "drilled-down" analysis specific to diabetic care, the following models were used using a sequential regression technique to determine if a PCP having any diabetic patients was sensitive to the total diabetic incentive available.

$$13.)\ Any\_Diabetic_{it} = \beta_0 + \beta_1 Incentive_{it} + \beta_2 Year_t + \varepsilon_{it}$$

$$14.)\ Any\_Diabetic_{it} = \beta_0 + \beta_1 Incentive_{it} + \beta_2 Year_t + \beta_3 Hospital_i + \beta_4 Degree_i + \beta_5 Specialty_i + \varepsilon_{it}$$

$$15.)\ Any\_Diabetic_{it} = \beta_0 + \beta_1 Incentive_{it} + \beta_2 Year_t + \alpha_i + \varepsilon_{it}$$

Among PCPs with diabetic patients, models of diabetic measure type achievement were estimated utilizing the same sequential regression technique. Since this level of analysis was only pertaining to diabetic care, the number of diabetic patients was controlled for in the models. The sequential estimation models for diabetic measure type achievement was as follows:

$$16.)\ Achievement_{ikt} = \beta_0 + \beta_1 Incentive_{ikt} + \beta_2 DiabPts_{it} + \beta_3 Year_t + \varepsilon_{it}$$

17.) $Achievement_{ikt} = \beta_0 + \beta_1 Incentive_{ikt} + + \beta_2 DiabPts_{it} + \beta_3 Year_t + \beta_4 Hospital_i + \beta_5 Degree_i + \beta_6 Specialty_i + \varepsilon_{it}$

18.) $Achievement_{ikt} = \beta_0 + \beta_1 Incentive_{ikt} + \beta_2 DiabPts_{it} + \beta_3 Year_t + \alpha_i + \varepsilon_{it}$

### 4.4.3 Diabetic Outcome Measure Achievement

For further analysis among PCPs with diabetic patients, the model utilizing fixed effects was specifically utilized to estimate achievement of individual diabetic outcome measures. An additional variable was added to control for patient clinical outcome levels, as shown in the following model:

19.) $Achievement_{ikt} = \beta_0 + \beta_1 Incentive_{ikt} + \beta_2 DiabPts_{it} + \beta_3 Clinical_{ikt} + \beta_4 Year_t + \alpha_i + \varepsilon_{it}$

The same model was slightly modified to applied towards diabetic outcome measures across baseline performances the previous year, categorized by quartile ranking relative to achievement thresholds. Because this approach looked at dynamic changes between years, the marginal incentive amount was used instead of the total incentive amount. The model was as follows:

20.) $Achievement_{ikt} = \beta_0 + \beta_1 MargIncentive_{ikt} + \beta_2 DiabPts_{it} + \beta_3 Clinical_{ikt} + \beta_4 Year_t + \alpha_i + \varepsilon_{it}$

#### 4.4.4 Clinical Outcome Levels of Diabetic Patients

The use of a physician-specific fixed effect accounted for unobservable PCP characteristics invariant of time, but not unobservable PCP characteristics that changed with time. In the case of the P4P program, the possibility existed of the PCPs composition of diabetic patients changing over time. The diabetic outcome measures within the patient-level data provided a unique opportunity for testing compositional effects among the PCPs diabetic patients over time.

To analyze the diabetic patients of the PCPs, the fixed effect model was applied to the patient-level data to estimate individual patient clinical outcome levels. Thus, the patient clinical outcome was placed as the dependent variable, instead of a measure of PCP performance. The model is as follows:

21.) $Clinical_{ikt} = \beta_0 + \beta_1 Incentive_{ikt} + \beta_2 Year_t + \alpha_i + \varepsilon_{it}$

Across all models:

- *%PtsEarned* was the percent of available points earned in a quality domain
- *Achievement* was the indicator of diabetic performance measure achievement
- *Incentive* was the amount of dollars available
- *MargIncentive* was the amount of change in available dollars from the previous year
- *DiabPts* was the number of diabetic patient seen
- *Hospital* was the indicator for the location of practice
- *Degree* was the indicator for the professional designation
- *Specialty* was the indicator for the type of medicine practiced
- *Clinical* was the patient clinical outcome level
- *Year* was the indicator of the calendar year of practice

The error term ($\varepsilon_{it}$) captured the unobserved factors influencing the outcome but not included as variables in the regression model. Individual fixed effects ($\alpha_i$) allowed control for unobserved, time-invariant physician-specific heterogeneity. Some examples of unobservable factors, different by individual but static over time, included motivational traits (including peer-effects), intrinsic ability or propensity to excel.

The parameter of interest ($\beta_1$) represented the expected change in *%PtsEarned* for one unit change in *Incentive*, when all the other covariates were held constant. The hypothesis implied $\beta_1$ was positive, statistically significant, and clinically meaningful. Also, elasticity coefficients were calculated to determine physician responsiveness to incentives. Robust standard errors were used in all regressions to account for heteroskedasticity.

# 5. RESULTS

## 5.1 Incentives for Performance Domain Scores

From 2004 to 2008, physicians of all specialties in the P4P program have gradually achieved a greater percentage of available points (or a greater score) among the infrastructure, effectiveness and safety domains, respectively. Physician scores of the other efficiency and experience domains have gradually decreased over the same time period. Among all domains, physicians were achieving roughly 4 out of every 5 points on average, with the exception of the experience domain where scores were much lower. The experience domain was added later in 2007, and contains comparatively much fewer measures than the other domains, which were added earlier. Physician performance was at its best among the efficiency domain in 2004, although these scores slightly dropped in subsequent years (Table XV).

**TABLE XV**

ACHIEVEMENT OF AVAILABLE POINTS (SCORE) PER DOMAIN, 2004 – 2008

| Year | | Infrastructure | Effectiveness | Efficiency | Safety | Experience | All Initiatives |
|---|---|---|---|---|---|---|---|
| **2004** | **Mean** | **71%** | **84%** | **92%** | **.** | **.** | **82%** |
| *n=2,437* | *SD* | *31%* | *29%* | *9%* | . | . | *26%* |
| **2005** | **Mean** | **72%** | **81%** | **90%** | **.** | **.** | **81%** |
| *n=2,487* | *SD* | *29%* | *29%* | *10%* | . | . | *24%* |
| **2006** | **Mean** | **73%** | **80%** | **82%** | **58%** | **.** | **75%** |
| *n=2,558* | *SD* | *25%* | *32%* | *19%* | *49%* | . | *30%* |
| **2007** | **Mean** | **82%** | **82%** | **77%** | **81%** | **60%** | **79%** |
| *n=2,590* | *SD* | *20%* | *24%* | *23%* | *37%* | *49%* | *30%* |
| **2008** | **Mean** | **89%** | **83%** | **79%** | **90%** | **51%** | **83%** |
| *n=2,661* | *SD* | *16%* | *26%* | *25%* | *17%* | *50%* | *25%* |
| **All Years** | **Mean** | **78%** | **82%** | **84%** | **81%** | **56%** | **80%** |
| *n=12,773* | *SD* | *26%* | *27%* | *20%* | *35%* | *50%* | *28%* |

After accounting for unobservable, individual-specific differences between physicians, incentive effects were found to be very small. From 2004 to 2008, for every $1000 of incentive available, physicians earned about 1.5 more infrastructure points ($p<0.01$) and 1.5 more effectiveness points ($p<0.01$). Physicians were relatively more responsive (elastic) to infrastructure incentives versus effectiveness incentives (Table XVI). Yet, the available incentive amount was highest for the infrastructure domain compared to the effectiveness domain.

**TABLE XVI**
INCENTIVE COEFFICIENT ESTIMATES FOR DOMAIN SCORE, 2004-2008 [a b c]

| Performance Domain | (10) | | | (11) | | | (12) | | | Mean Incentive | Mean Achievement | Elasticity of Incentive [d] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Infrastructure** | **0.0028** | 0.0010 | *** | **0.0068** | 0.0011 | *** | **0.0147** | 0.0017 | *** | **$2,073.18** | **78%** | 0.04 |
| *Adjusted $R^2$* | *0.0730* | | | *0.1477* | | | *0.5194* | | | | | |
| **Effectiveness** | **0.0255** | 0.0028 | *** | **0.0161** | 0.0029 | *** | **0.0151** | 0.0045 | *** | **$1,303.47** | **82%** | 0.02 |
| *Adjusted $R^2$* | *0.0134* | | | *0.0888* | | | *0.3523* | | | | | |
| **Efficiency** | **0.0040** | 0.0011 | *** | **0.0030** | 0.0012 | *** | **0.0024** | 0.0015 | | **$1,976.09** | **84%** | 0.01 |
| *Adjusted $R^2$* | *0.0956* | | | *0.1437* | | | *0.3137* | | | | | |
| **Safety** | **0.0351** | 0.0023 | *** | **0.0301** | 0.0026 | *** | **0.0032** | 0.0053 | | **$1,236.78** | **81%** | 0.00 |
| *Adjusted $R^2$* | *0.1174* | | | *0.1937* | | | *0.2771* | | | | | |
| **Experience** | **0.0203** | 0.0214 | | **0.1232** | 0.0208 | *** | **0.2055** | 0.0624 | *** | **$698.46** | **56%** | 0.25 |
| *Adjusted $R^2$* | *0.0085* | | | *0.1851* | | | *0.5257* | | | | | |
| **Overall** | **0.0104** | 0.0007 | *** | **0.0128** | 0.0008 | *** | **0.0249** | 0.0012 | *** | **$1,732.84** | **80%** | 0.05 |
| *Adjusted $R^2$* | *0.0166* | | | *0.0433* | | | *0.1367* | | | | | |
| *Year* Dummies | Yes | | | Yes | | | Yes | | | | | |
| *Hospital* Dummies | No | | | Yes | | | No | | | | | |
| *Degree* Dummies | No | | | Yes | | | No | | | | | |
| *Specialty* Dummies | No | | | Yes | | | No | | | | | |
| Physician F.E. | No | | | No | | | Yes | | | | | |

[a] n= 12,733
[b] Statistical Significance Levels: * = p<0.10; ** = p<0.05; *** = p<0.01
[c] Coefficient (Standard Error) estimates for incentive are multiplied by a factor of $10^{-3}$
[d] Elasticity based on model with physician-specific fixed effects (12)

## 5.2 Incentives for Diabetic Measure Type Achievement

During a preliminary analysis of any PCP in the P4P program, it was shown that PCPs conditional on having any diabetic patients was positively correlated with increases in the available diabetic incentive amount (Table XVII). This relationship suggested PCPs were not avoiding diabetic patients as incentives associated with their care increased. Additionally, this relationship was consistent with the P4P program increasing the number of diabetic patients in its registry at the same time they were increasing incentive amounts.

**TABLE XVII**
INCENTIVE COEFFICIENT ESTIMATES FOR DIABETIC PATIENT INDICATOR, 2005-2008 [a, b]

| | **(13)** | | | **(14)** | | | **(15)** | | |
|---|---|---|---|---|---|---|---|---|---|
| **Diabetic Patient** | **0.0002** | 0.0000 | *** | **0.0002** | 0.0000 | *** | **0.0001** | 0.0000 | *** |
| *Adjusted $R^2$* | *0.0776* | | | *0.1138* | | | *0.5334* | | |

[a] n= 1,917
[b] Statistical Significance Levels: * = p<0.10; ** = p<0.05; *** = p<0.01

Specific to PCPs with diabetic patients, it was shown PCPs have gradually achieved a greater percentage of diabetic measures each year, even as the "bar" for each measure has gradually been raised and more measures have been added from 2005 to 2008. Diabetic measure achievement was generally equal for process and outcome measures during this time period, and PCPs were achieving roughly 7 out of every 10 diabetic measures presented on average (Table XVIII).

**TABLE XVIII**

ACHIEVEMENT OF DIABETIC MEASURE TYPE, 2005 - 2008

| Year | | Process Measures | Outcome Measures | All Measures |
|---|---|---|---|---|
| **2005** | **Mean** | **66%** | **60%** | **62%** |
| *n= 342* | *SD* | *48%* | *49%* | *48%* |
| **2006** | **Mean** | **61%** | **64%** | **62%** |
| *n=394* | *SD* | *49%* | *48%* | *48%* |
| **2007** | **Mean** | **74%** | **78%** | **76%** |
| *n=412* | *SD* | *44%* | *42%* | *43%* |
| **2008** | **Mean** | **77%** | **77%** | **77%** |
| *n=428* | *SD* | *42%* | *42%* | *42%* |
| **All Years** | **Mean** | **71%** | **72%** | **71%** |
| *n=1576* | *SD* | *46%* | *45%* | *45%* |

After accounting for unobservable, individual-specific differences between PCPs, incentive effects were found to be small and modest. From 2005 to 2008, for every $100 of incentive available, PCPs were 2% points more likely to achieve a process measure ($p<0.10$) and 7% points more likely to achieve an outcome measure ($p<0.10$) on average. Although incentives were slightly more effective toward outcome measures, PCPs were more responsive (elastic) to such incentives compared to process measures (Table XIX). Yet, the available incentive amount was higher for process measures.

**TABLE XIX**

INCENTIVE COEFFICIENT ESTIMATES FOR MEASURE TYPE ACHIEVEMENT, 2005 – 2008 [a b]

| Diabetic Measure Type | (16) | | | (17) | | | (18) | | | Mean Incentive | Mean Achievement | Elasticity of Incentive [c] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Process** | **0.0006** | 0.0001 | *** | **0.0004** | 0.0001 | *** | **0.0002** | 0.0001 | * | **$66.93** | **71%** | 0.02 |
| Adjusted $R^2$ | *0.0267* | | | *0.0532* | | | *0.3262* | | | | | |
| **Outcome** | **0.0013** | 0.0002 | *** | **0.0008** | 0.0002 | *** | **0.0007** | 0.0003 | * | **$44.80** | **72%** | 0.04 |
| Adjusted $R^2$ | *0.0482* | | | *0.0993* | | | *0.3375* | | | | | |
| **Overall** | **0.0007** | 0.0001 | *** | **0.0005** | 0.0001 | *** | **0.0002** | 0.0001 | * | **$55.32** | **71%** | 0.01 |
| Adjusted $R^2$ | *0.0352* | | | *0.0741* | | | *0.3177* | | | | | |
| *Year* Dummies | Yes | | | Yes | | | Yes | | | | | |
| *Hospital* Dummies | No | | | Yes | | | No | | | | | |
| *Degree* Dummies | No | | | Yes | | | No | | | | | |
| *Specialty* Dummies | No | | | Yes | | | No | | | | | |
| Physician F.E. | No | | | No | | | Yes | | | | | |

[a] n= 1,576

[b] Statistical Significance Levels: * = p<0.10; ** = p<0.05; *** = p<0.01

[c] Elasticity based on model with physician-specific fixed effects (15)

## 5.3 Incentives for Diabetic Outcome Measure Achievement

PCPs have progressively achieved more outcome measures each year specific to HbA1c and LDL-C from 2005 to 2008, even the threshold for achievement has become more difficult to achieve. Outcome measures specific to blood pressure control were added in 2007. On average, the most achievement was among the outcome measures of blood pressure, HbA1c and LCL-C, respectively. Specifically, PCPs were more easily achieving hypertension control (<140/90) measures than pre-hypertension control (<130/80) measures as expected. Interestingly, there was a slight drop in PCPs achievement of outcome measures related to hypertension and HbA1c control from 2007 to 2008 (Table XX).

**TABLE XX**

ACHIEVEMENT OF DIABETIC OUTCOME MEASURES, 2005 - 2008

| Year | | Good HbA1c Control (<7) | Poor HbA1c Control (>9) | Good LDL-C Control (<100) | Poor LDL-C Control (<130) | Hypertension BP Control (<140/90) | Pre-Hypertension BP Control (<130/80) |
|---|---|---|---|---|---|---|---|
| **2005** | **Mean** | **0.59** | **0.58** | **0.60** | **0.63** | | |
| *n= 342* | *SD* | *0.49* | *0.49* | *0.49* | *0.48* | | |
| **2006** | **Mean** | **0.63** | **0.68** | **0.61** | **0.64** | | |
| *n=394* | *SD* | *0.48* | *0.47* | *0.49* | *0.48* | | |
| **2007** | **Mean** | **0.75** | **0.80** | **0.67** | **0.72** | **0.90** | **0.86** |
| *n=412* | *SD* | *0.43* | *0.40* | *0.47* | *0.45* | *0.30* | *0.35* |
| **2008** | **Mean** | **0.70** | **0.81** | **0.71** | **0.80** | **0.86** | **0.73** |
| *n=428* | *SD* | *0.46* | *0.39* | *0.45* | *0.40* | *0.35* | *0.45* |
| **All Years** | **Mean** | **0.67** | **0.72** | **0.65** | **0.70** | **0.88** | **0.78** |
| *n=1576* | *SD* | *0.47* | *0.45* | *0.48* | *0.46* | *0.33* | *0.41* |

From 2005 to 2008, PCPs were 1% point more likely to achieve the "Good HbA1c Control (<7)" outcome measure for every $10 of incentive available to them (p<0.01). PCPs were also relatively more responsive (elastic) to these incentives. Although not statistically significant, outcome measures attempting to decrease poor performance (“Poor HbA1c Control (>9)” and “Poor LDL-C Control (>130)”) were negatively affected by incentive amount, as expected. So, the more available incentive would decrease the chances of PCPs having poor HbA1c or LDL-C control among their patients (Table XXI).

**TABLE XXI**

INCENTIVE COEFFICIENT ESTIMATES FOR OUTCOME MEASURE ACHIEVEMENT, 2005 – 2008 [a b]

| **Diabetic Outcome Measures** | **(19)** | | | **Mean Incentive** | **Mean Achievement** | **Elasticity of Incentive** [c] |
|---|---|---|---|---|---|---|
| **Good HbA1c Control (<7)** | **0.0013** | 0.0007 | *** | **$43.54** | **0.67** | 0.09 |
| Adjusted $R^2$ | *0.4217* | | | | | |
| **Poor HbA1c Control (>9)** | **-0.0003** | 0.0008 | | **$43.84** | **0.72** | -0.02 |
| Adjusted $R^2$ | *0.3965* | | | | | |
| **Good LDL-C Control (<100)** | **0.0000** | 0.0006 | | **$43.62** | **0.65** | 0.00 |
| Adjusted $R^2$ | *0.4647* | | | | | |
| **Poor LDL-C Control (>130)** | **-0.0002** | 0.0006 | | **$43.88** | **0.70** | -0.01 |
| Adjusted $R^2$ | *0.4475* | | | | | |
| **Hypertension BP Control (<140/90)** | **0.0003** | 0.0006 | | **$49.75** | **0.88** | 0.02 |
| Adjusted $R^2$ | *0.3231* | | | | | |
| **Pre-Hypertension BP Control (<130/80)** | **0.0005** | 0.0008 | | **$47.92** | **0.78** | 0.03 |
| Adjusted $R^2$ | *0.2996* | | | | | |
| *Year* Dummies | Yes | | | | | |
| *Hospital* Dummies | No | | | | | |
| *Degree* Dummies | No | | | | | |
| *Specialty* Dummies | No | | | | | |
| Physician-Specific Fixed Effects | Yes | | | | | |

[a] n= 1,576

[b] Statistical Significance Levels: * = p<0.10; ** = p<0.05; *** = p<0.01

[c] Elasticity based on model with physician-specific fixed effects (16)

## 5.4 Marginal Incentives for Diabetic Outcome Measure Achievement

From 2005 to 2008, PCPs initially performing in a higher quartile (3rd or 4th respectively) the previous year were more likely to achieve an outcome measure. PCPs initially performing the 1st quartile were the least likely to achieve an outcome measure on average, although having the most marginal incentive available to them. From 2005 to 2008, PCPs initially performing the 3rd quartile performed worse over time, although having the most overall achievement on average. In all other quartiles, PCPs generally performed better over the same time period (Table XXII).

**TABLE XXII**
ACHIEVEMENT FOR PERCENTILE RANK PREVIOUS YEAR, ALL OUTCOME MEASURES, 2005 - 2008

| Year | | <0.25 | >=0.25 to <0.50 | >=0.50 to <0.75 | >=0.75 |
|---|---|---|---|---|---|
| **2005 - 2006** | **Mean** | **0.42** | **0.76** | **0.92** | **0.77** |
| | *SD* | *0.49* | *0.43* | *0.27* | *0.42* |
| **2006 - 2007** | **Mean** | **0.51** | **0.82** | **0.91** | **0.86** |
| | *SD* | *0.50* | *0.39* | *0.29* | *0.35* |
| **2007 - 2008** | **Mean** | **0.59** | **0.72** | **0.84** | **0.89** |
| | *SD* | *0.49* | *0.45* | *0.37* | *0.32* |
| **All Years** | **Mean** | **0.51** | **0.76** | **0.88** | **0.85** |
| | *SD* | *0.50* | *0.43* | *0.33* | *0.35* |

PCPs initially performing the 1st quartile were 2% points more likely to achieve an outcome measure for every $10 of marginal incentive available to them ($p<0.01$). Also, these PCPs were relatively more responsive (elastic) to these marginal incentives. Interestingly, PCPs initially performing in the 3rd or 4th percentiles were less likely to achieve an outcome measure, although not statistically significant (Table XXIII).

**TABLE XXIII**
MARGINAL INCENTIVE COEFFICIENT ESTIMATES FOR PERCENTILE RANK ACHIEVEMENT, ALL OUTCOME MEASURES, 2005 – 2008 [a b]

| **Percentile Rank** | **(20a)** | | | **Mean Marginal Incentive** | **Mean Achievement** | **Elasticity of Marginal Incentive** [c] |
|---|---|---|---|---|---|---|
| **<0.25** | **0.0022** | 0.0009 | *** | **$7.32** | **0.51** | **0.03** |
| Adjusted $R^2$ | | *0.3151* | | | | |
| **>=0.25 to <0.50** | **0.0000** | 0.0009 | | **$5.45** | **0.76** | **0.00** |
| Adjusted $R^2$ | | *0.3035* | | | | |
| **>=0.50 to <0.75** | **-0.0006** | 0.0005 | | ***$5.76*** | **0.88** | **0.00** |
| Adjusted $R^2$ | | *0.3705* | | | | |
| **>=0.75** | **-0.0001** | 0.0003 | | ***$5.62*** | **0.85** | **0.00** |
| Adjusted $R^2$ | | *0.3880* | | | | |
| *Year* Dummies | | Yes | | | | |
| *Hospital* Dummies | | No | | | | |
| *Degree* Dummies | | No | | | | |
| *Specialty* Dummies | | No | | | | |
| Physician-Specific Fixed Effects | | Yes | | | | |

[a] n= 1,576
[b] Statistical Significance Levels: * = p<0.10; ** = p<0.05; *** = p<0.01
[c] Elasticity based on model with physician-specific fixed effects (17)

Because incentives were significantly effective for the "Good HbA1c Control (<7)" outcome measure, an additional analysis was conducted to explore how PCP achievement varied by baseline performance. PCPs initially performing in a higher quartile (4th or 3rd respectively) the previous year were more likely to achieve this specific outcome measure. PCPs initially performing in the 1st quartile were the least likely to achieve this outcome measure, on average. From 2005 to 2008, PCPs were mostly performing better over time in all quartiles. However, PCP performance abruptly dropped in the last year of observation (2007 to 2008) among those initially performing in the 1st, 2nd or 3rd quartiles, while PCP performance spiked among those initially performing in the 4th quartile (Table XXIV).

**TABLE XXIV**
ACHIEVEMENT FOR PERCENTILE RANK PREVIOUS YEAR, GOOD HBA1C CONTROL (<7) MEASURE, 2005 - 2008

| Year | | <0.25 | >=0.25 to <0.50 | >=0.50 to <0.75 | >=0.75 |
|---|---|---|---|---|---|
| **2005 - 2006** | **Mean** | **0.43** | **0.81** | **0.85** | **0.92** |
| | *SD* | *0.50* | *0.40* | *0.36* | *0.28* |
| **2006 - 2007** | **Mean** | **0.57** | **0.89** | **0.88** | **0.89** |
| | *SD* | *0.50* | *0.32* | *0.33* | *0.31* |
| **2007 - 2008** | **Mean** | **0.49** | **0.51** | **0.71** | **0.93** |
| | *SD* | *0.50* | *0.50* | *0.46* | *0.25* |
| **All Years** | **Mean** | **0.50** | **0.73** | **0.80** | **0.92** |
| | *SD* | *0.50* | *0.44* | *0.40* | *0.27* |

PCPs initially performing in the $1^{st}$ quartile were 5% points more likely to achieve the "Good HbA1c Control (<7)" outcome measure for every $10 of marginal incentive available to them ($p<0.05$), although measure achievement among PCPs in this quartile were the least on average. Also, PCPs were relatively more responsive (elastic) to these particular marginal incentives. Although not statistically significant, incentive effects were much lower if PCPs were initially performing higher (Table XXV).

**TABLE XXV**
MARGINAL INCENTIVE COEFFICIENT ESTIMATES FOR PERCENTILE RANK ACHIEVEMENT, GOOD HBA1C CONTROL (<7) MEASURE, 2005 – 2008 [a b]

| **Percentile Rank** | **(20b)** | | | **Mean Marginal Incentive** | **Mean Achievement** | **Elasticity of Marginal Incentive** [c] |
|---|---|---|---|---|---|---|
| **<0.25** | **0.0050** | 0.0021 | ** | **$7.27** | **0.50** | **0.07** |
| Adjusted $R^2$ | *0.3969* | | | | | |
| **>=0.25 to <0.50** | **0.0043** | 0.0056 | | **$5.22** | **0.73** | **0.03** |
| Adjusted $R^2$ | *0.4051* | | | | | |
| **>=0.50 to <0.75** | **0.0007** | 0.0031 | | ***$4.90*** | **0.80** | **0.00** |
| Adjusted $R^2$ | *0.4971* | | | | | |
| **>=0.75** | **0.0007** | 0.0008 | | ***$4.66*** | **0.92** | **0.00** |
| Adjusted $R^2$ | *0.6265* | | | | | |
| *Year* Dummies | Yes | | | | | |
| *Hospital* Dummies | No | | | | | |
| *Degree* Dummies | No | | | | | |
| *Specialty* Dummies | No | | | | | |
| Physician-Specific Fixed Effects | Yes | | | | | |

[a] n= 1,576
[b] Statistical Significance Levels: * = p<0.10; ** = p<0.05; *** = p<0.01
[c] Elasticity based on model with physician-specific fixed effects (17)

## 5.5 Incentives for Diabetic Clinical Outcomes Levels

When observing the patients of these same PCPs, the clinical outcome levels for HbA1c, LDL-C and blood pressure have generally decreased (patients were becoming healthier) from 2005-2008. However, there was a very small increase in blood pressure among the patients from 2007 to 2008. Yet, PCPs treated significantly more patients in 2007 and 2008. On average, PCPs treated nearly three times the amount of patients in 2008 than in 2005 (Table XXVI).

**TABLE XXVI**
CLINICAL DIABETIC OUTCOME LEVELS, 2005 - 2008

| Year | | HbA1c | LDL-C | Systolic BP | Diastolic BP |
|---|---|---|---|---|---|
| **2005** | **Mean** | **7.58** | **99.13** | **.** | **.** |
| *n=1,987* | *SD* | *1.64* | *52.54* | . | . |
| **2006** | **Mean** | **7.33** | **92.78** | **.** | **.** |
| *n=2,813* | *SD* | *1.66* | *43.46* | . | . |
| **2007** | **Mean** | **7.24** | **90.89** | **125.77** | **74.69** |
| *n=4,681* | *SD* | *1.61* | *38.57* | *14.31* | *9.76* |
| **2008** | **Mean** | **7.37** | **90.19** | **125.85** | **75.25** |
| *n=6,084* | *SD* | *1.58* | *41.98* | *13.61* | *9.13* |
| **All Years** | **Mean** | **7.35** | **91.99** | **125.82** | **75.01** |
| *n=15,565* | *SD* | *1.61* | *42.86* | *13.91* | *9.41* |

When controlling for unobservable, individual-specific differences between patients, incentive effects were generally small and modest. From 2005 to 2008, for every $1000 of incentive available to PCPs, there was a decrease by a clinical value of 1.0 in the HbA1c outcome levels of their patients ($p<0.10$). Incentive effects were larger for changes in blood pressure levels, although not statistically significant (Table XXVII).

**TABLE XXVII**
INCENTIVE COEFFICIENT ESTIMATES FOR OUTCOME LEVEL, 2005 – 2008 [a b]

| Diabetic Measure Type | (21) | | | Mean Physician Incentive | Mean Patient Clinical Level | Elasticity of Incentive [c] |
|---|---|---|---|---|---|---|
| **HbA1c** | **-0.0010** | 0.0006 | * | **$43.54** | **7.35** | -0.01 |
| Adjusted $R^2$ | *0.6668* | | | | | |
| **LDL-C** | **-0.0005** | 0.0205 | | **$43.62** | **91.99** | 0.00 |
| Adjusted $R^2$ | *0.5049* | | | | | |
| **Systolic BP** | **-0.0047** | 0.0077 | | **$47.92** | **125.82** | 0.00 |
| Adjusted $R^2$ | *0.4595* | | | | | |
| **Diastolic BP** | **-0.0069** | 0.0070 | | **$47.92** | **75.01** | 0.00 |
| Adjusted $R^2$ | *0.4205* | | | | | |
| *Year* Dummies | Yes | | | | | |
| *Hospital* Dummies | No | | | | | |
| *Degree* Dummies | No | | | | | |
| *Specialty* Dummies | No | | | | | |
| Physician-Specific Fixed Effects | Yes | | | | | |

[a] n= 15,565
[b] Statistical Significance Levels: * = p<0.10; ** = p<0.05; *** = p<0.01
[c] Elasticity based on model with physician-specific fixed effects (18)

# 6. DISCUSSION AND CONCLUSION

## 6.1 Suggestions from Evidence

The evidence presented in this study suggests P4P financial incentives can be effective when applied appropriately to targeted areas, resulting in small and limited dose-response relationships. As predicted, a greater financial incentive level has modestly increased physician measureable performance, yet specific to the clinical effectiveness quality domain. As physician agency theory suggests, both the physician and the payer acted in a manner consistent with maximizing individual utility. Balancing between the competing interests of reimbursement paid and quality expended toward outcome achievement, this study demonstrated a productive agreement is possible between both parties within a utility framework under constraints.

To maximize efficiency in the production of the outcome, the PHO has gradually offered increases in financial incentive alongside gradual increases in desired quality outcomes. The PHO also ensured physicians would continue to participate within their respective participation constraints while producing increasing quality. With such considerations, this study demonstrates such a relationship could successfully exist, stemming from reimbursement level and outcome amount as influencers of physician behavior.

A common criticism of P4P programs is that they present an uneven playing field, where some physicians may have “different” or “sicker” patients. Exogenous patient risk factors may be exceedingly responsible for the quality indicator, rather than physician quality level alone. Also, a moral hazard may exist on behalf of physicians selecting and treating a disproportionate share of higher risk patients.

Among PCPs, this study suggests financial incentive level increased achievement of both process-based and outcome-based diabetic measures specific to HbA1c, to roughly equal levels. Also, PCPs were more responsive to these incentives compared to those of other measures. This is contrary to what was predicted, where more financial incentive was necessary for outcome-based measure achievement compared to process-of-care measures, primarily because the PCP would take on more exogenous risk of uncontrollable patient factors with outcome-based measures. Such conclusions suggest exogenous patient factors may not present as much risk to diabetic outcomes as theoretically expected.

Among all diabetic outcome measures, only those for HbA1c were significant. HbA1c measurement is often the standard by which PCPs determine diabetic acuity and is largely exclusive to diabetic care. The LDL or blood pressure outcomes are also common measures for diabetic care, but may crossover into other disease states (e.g. cardiovascular disease, atherosclerosis), potentially affecting any direct relationship a diabetic incentive has for a diabetic outcome.

This study presented suggestive evidence that financial incentives aimed to keep patient HbA1c levels below 7, were effective while directly controlling for patient composition changing over time. Also discovered at the individual patient level, clinical HbA1c levels decreased as a result of targeted PCP financial incentives. With this unique “drilled down” approach, more unobservable patient characteristics were controlled, providing additional assurance PCPs were not under-rewarded and over-rewarded for their efforts toward high-quality care because of differences in patient composition.

Among diabetic outcome-based measures, the results of this study suggest additional incentives were most effective when PCP baseline performance was well below the threshold (within the first quartile). Yet, it was predicted additional financial incentives would be more effective when baseline performance was just slightly below (not well below) quality thresholds (9). This prediction assumed the bonus was discontinuous around the threshold and PCPs were incentivized with the same marginal bonus amount. However, PCPs originally performing in the $1^{st}$ quartile (the lowest performing physicians at baseline) were marginally incentivized the most on average, potentially shifting PCPs behavior disproportionally. Regardless, PCPs with the lowest baseline performance were influenced the most from financial incentives, even though significant quality effort (a disutility for the PCP) was required to achieve a threshold. These conclusions suggest either PCPs value the reimbursement received more than quality expended (compared to higher performing PCPs) or additional present factors influenced PCP behavior beyond the financial bonus itself.

The introduction of a threshold (along with indication of where PCPs perform relative to a threshold) may create a peer effect from competition, which may intensify when PCPs perform well below their colleagues. To capture such effects, this study examined marginal physician performance from baseline by grouping physicians together with other similar performing physicians into quartiles. Incentive effects were estimated with similar performing physician groups, rather than as a whole. Thus, it was determined how physician-performing groups performed relative to each other. In addition, the reaction to peer effects (in the form of motivation or propensity to excel when faced with competition) was unobservable and varied by individual physician. This was largely accounted for statistically in the estimation model with the physician-specific fixed effect.

## 6.2 Limitations

A key limitation of this study was the narrow scope of the data. It mostly covers just a small subset of physicians and patients. Also, there were just a few years of data from which to draw conclusions from, and physicians were moving in and out of the program during this time. Furthermore, the data was limited to one private health system and general geographic area, which limit the generalizability of the findings.

In the domain-level analysis (physicians of all specialties), an incentive appeared to exist to drop patients because points were rewarded if a physician had zero patients. Among all specialties, this was the most probable with primary care (internal medicine, family medicine) because PCPs most likely saw the widest variety of patients among several disease states by the nature of their practice. Also, PCPs were the most numerous and the most performance measures available to them. Across the performance domains, patient dropping may be a plausible problem in the clinical effectiveness domain because most of these measures were outcome related and disease specific. Measures in the other domains could generally be applied to any patient. Besides PCPs, other physicians in the domain-level analysis were more specialized and mainly treated patients of certain disease states with pre-specified measures. Such specialized physicians may be less likely to dump their patients considering they may have no other remaining ones to choose from. For this reason, a further "drilled down" analysis was performed among PCPs with diabetic patients, exclusively. In essence, the number of patients was known, along with the patient-level clinical data on these specific patients. If a PCP had no applicable patients, they were excluded from the dataset. Yet, the results of the domain-level analysis (physicians of all specialties) still contained this limitation

In 2007, the patient experience domain was added, which had the least amount of incentive and measures associated with it, mostly related to patient satisfaction scores and handoff procedures. Achievement of the measures in the patient experience domain was significantly lower than other domains and decreased over time. Such trends suggest the program placed less importance on the patient experience domain relative to the other domains. In result, physician's efforts may be focused toward the other domains with more measures that increased in number and incentive amount during 2007 and 2008 (10). In addition, incorporation of more measures and incentive in the overall program over time may be distracting physician efforts away from the undocumented and unrewarded aspects of the physician-patient relationship. Arguably, patient satisfaction may be consequently suffering as indicated in the lowering scores in the patient experience domain.

Considering the substantial prevalence of adult diabetes in the general population (40), diabetic patient numbers were relatively lower on average, particularly in the earlier years addressed in the data subset. During the later years (mainly 2007 and 2008), diabetic patient numbers were noticeably higher. Such trends suggest the program may be actively recruiting diabetic patients or have expanded their definition for a patient to be considered diabetic.

In patient populations like diabetics, practice policies in the form of protocols-of-care or patient algorithms may create a strong incentive itself and help further explain PCPs practice patterns. In these situations, PCPs may strictly perform procedures indicated in diabetic performance measures, in addition to what financial incentive effects suggest. If such practice-level policies shift PCP behavior toward clinical standards of care, PCPs may be performing better than they would otherwise.

## 6.3 Implications for Practice

This paper suggests specific performance differences exist among physicians when varying financial incentive amounts are introduced, which present significant implications for healthcare payers and providers considering P4P approaches for physician alignment. Implementers of P4P programs stand to be more informed about where the value lies with physician financial incentives, and where it is suggested to modify physician behavior, from an economic perspective. Thus, the design of P4P programs will be influenced for physicians treating targeted patient populations, especially for chronic care management of diabetics. Also, as target-based quality indicators may discourage high performing PCPs from improving, results from this study suggest providing financial incentives for measureable performance improvement (rather than attainment) may be a more effective method for increasing clinical quality of diabetic care.

The ACO framework is structured for providers to take on more risk and become largely responsible for the outcomes of patient populations. Results from this study may imply when physicians become more responsible for the outcomes of diabetic patients, small (yet statistically significant) incentive effects result from the incentive offered. Thus, the provision of physician financial incentives has potential, yet a higher incentive amount may be necessary to see more substantial and meaningful changes in populations of diabetic patients. Although the application of this framework to diabetic patient populations show promise, such incentive for outcomes arrangement should be further explored in other primary care settings (e.g. CAD, Asthma) in interest of advancing population health management.

## 6.4 Implications for Research

This paper is one of few studies suggesting targeted physician financial incentives can make a difference in patient HbA1c outcomes. The significance of the conclusions can be partially attributable to the use of patient-level, registry data to verify the effects diabetic incentives at the physician-level. This approach has not been commonly used in the related literature, possibly because patient level data is difficult to obtain for confidentiality purposes. Understandably, most healthcare providers are reluctant to openly share their data, unless required (e.g. public reporting for CMS). The physician-level and patient-level data used in this study was private and used for internal purposes. Considering the potential research implications, private healthcare providers should strongly consider data sharing arrangements for the purposes of research applications. Patient registries are a rich data source. When matched with physician-level performance data, it has significant research implications, as this study demonstrates.

# CITED LITERATURE

1. Rosenthal, M. B., R. G. Frank, Z. Li, and A. M. Epstein: Early experience with pay-for-performance: from concept to practice. *JAMA* 294: 1788-93, 2005.

2. Antos, J., J. Bertko, M. Chernew, D. Cutler, D. Goldman, M. McClellan, E. McGlynn, M. Pauly, L. Schaeffer, and S. Shortell: Bending the curve: effective steps to address long-term healthcare spending growth. *Am J Manag Care* 15: 676-80, 2009.

3. Giuffrida, A., T. Gosden, F. Forland, I. S. Kristiansen, M. Sergison, B. Leese, L. Pedersen, and M. Sutton: Target payments in primary care: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*: CD000531, 2000.

4. Wheeler, J. R., B. White, S. Rauscher, T. A. Nahra, K. L. Reiter, K. M. Curtin, and C. L. Damberg: Pay-for-performance as a method to establish the business case for quality. *J Health Care Finance* 33: 17-30, 2007.

5. Rosenthal, M. B.: Beyond pay for performance--emerging models of provider-payment reform. *N Engl J Med* 359: 1197-200, 2008.

6. Baker, D. W., A. Qaseem, P. P. Reynolds, L. A. Gardner, E. C. Schneider, and Committee American College of Physicians Performance Measurement: Design and use of performance measures to decrease low-value services and achieve cost-conscious care. *Ann Intern Med* 158: 55-9, 2013.

7. Steiger, B.: Poll finds physicians very wary of pay-for-performance programs. *Physician executive* 31: 6-11, 2005.

8. Greene, S. E., and D. B. Nash: Pay for performance: an overview of the literature. *Am J Med Qual* 24: 140-63, 2009.

9. Mullen, K. J., R. G. Frank, and M. B. Rosenthal: Can you get what you pay for? Pay-for-performance and the quality of healthcare providers. *Rand J Econ* 41: 64-91, 2010.

10. Holmstrom, Milgrom: Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization* 7: 24-52, 1991.

11. McGuire, T. G. 2000. "Physician Agency." In *Handbook of Health Economics*, edited by Culyer A. J. and Newhous J.P., 462-527. Elsevier Science B.V.

12. Liebman JB, and Zeckhauser RJ.: Schmeduling. 60, 2004.

13. Shen, J., R. Andersen, R. Brook, G. Kominski, P. S. Albert, and N. Wenger: The effects of payment method on clinical decision-making: physician responses to clinical scenarios. *Med Care* 42: 297-302, 2004.

14. Rosenthal, M., and A. Milstein: Awakening consumer stewardship of health benefits: prevalence and differentiation of new health plan models. *Health Serv Res* 39: 1055-70, 2004.

15. Kouides, R. W., N. M. Bennett, B. Lewis, J. D. Cappuccio, W. H. Barker, and F. M. LaForce: Performance-based physician reimbursement and influenza immunization rates in the elderly. The Primary-Care Physicians of Monroe County. *Am J Prev Med* 14: 89-95, 1998.

16. Hillman, A. L., K. Ripley, N. Goldfarb, I. Nuamah, J. Weiner, and E. Lusk: Physician financial incentives and feedback: failure to increase cancer screening in Medicaid managed care. *Am J Public Health* 88: 1699-701, 1998.

17. Hillman, A. L., K. Ripley, et al. : The Use of Physician Financial Incentives and Feedback to Improve Pediatric Preventive Care in Medicaid Managed Care . *Pediatrics* 104: 931-935, 1999.

18. Fairbrother, G., K. L. Hanson, S. Friedman, and G. C. Butts: The impact of physician bonuses, enhanced fees, and feedback on childhood immunization coverage rates. *Am J Public Health* 89: 171-5, 1999.

19. Fairbrother, G., M. J. Siegel, S. Friedman, P. D. Kory, and G. C. Butts: Impact of financial incentives on documented immunization rates in the inner city: results of a randomized controlled trial. *Ambul Pediatr* 1: 206-12, 2001.

20. Scott, A., P. Sivey, D. Ait Ouakrim, L. Willenberg, L. Naccarella, J. Furler, and D. Young: The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrane Database Syst Rev*: CD008451, 2011.

21. Emmert, M., F. Eijkenaar, H. Kemter, A. S. Esslinger, and O. Schoffski: Economic evaluation of pay-for-performance in health care: a systematic review. *Eur J Health Econ* 13: 755-67, 2012.

22. An, L. C., J. H. Bluhm, S. S. Foldes, N. L. Alesci, C. M. Klatt, B. A. Center, W. S. Nersesian, M. E. Larson, J. S. Ahluwalia, and M. W. Manley: A randomized trial of a pay-for-performance program targeting clinician referral to a state tobacco quitline. *Arch Intern Med* 168: 1993-9, 2008.

23. Roski, J., R. Jeddeloh, L. An, H. Lando, P. Hannan, C. Hall, and S. H. Zhu: The impact of financial incentives and a patient registry on preventive care quality: increasing provider adherence to evidence-based smoking cessation practice guidelines. *Prev Med* 36: 291-9, 2003.

24. Twardella, D., and H. Brenner: Effects of practitioner education, practitioner payment and reimbursement of patients' drug costs on smoking cessation in primary care: a cluster randomised trial. *Tob Control* 16: 15-21, 2007.

25. Young, G. J., M. Meterko, H. Beckman, E. Baker, B. White, K. M. Sautter, R. Greene, K. Curtin, B. G. Bokhour, D. Berlowitz, and J. F. Burgess, Jr.: Effects of paying physicians based on their relative performance for quality. *J Gen Intern Med* 22: 872-6, 2007.

26. Nahra, T. A., K. L. Reiter, R. A. Hirth, J. E. Shermer, and J. R. Wheeler: Cost-effectiveness of hospital pay-for-performance incentives. *Med Care Res Rev* 63: 49S-72S, 2006.

27. Lee, T. T., S. H. Cheng, C. C. Chen, and M. S. Lai: A pay-for-performance program for diabetes care in Taiwan: a preliminary assessment. *Am J Manag Care* 16: 65-9, 2010.

28. Norton, E. C.: Incentive regulation of nursing homes. *J Health Econ* 11: 105-28, 1992.

29. Rosenthal, M. B., Z. Li, A. D. Robertson, and A. Milstein: Impact of financial incentives for prenatal care on birth outcomes and spending. *Health Serv Res* 44: 1465-79, 2009.

30. Ryan, A. M.: Effects of the Premier Hospital Quality Incentive Demonstration on Medicare patient mortality and cost. *Health Serv Res* 44: 821-42, 2009.

31. Curtin, K., H. Beckman, G. Pankow, Y. Milillo, and R. A. Green: Return on investment in pay for performance: a diabetes case study. *J Healthc Manag* 51: 365-74; discussion 375-6, 2006.

32. Parke, D. W., 2nd: Impact of a pay-for-performance intervention: financial analysis of a pilot program implementation and implications for ophthalmology (an American Ophthalmological Society thesis). *Trans Am Ophthalmol Soc* 105: 448-60, 2007.

33. Li, J., J. Hurley, P. DeCicca, and G. Buckley: Physician response to pay-for-performance: evidence from a natural experiment. *Health Econ* 23: 962-78, 2014.

34. Fagan, P. J., A. B. Schuster, C. Boyd, J. A. Marsteller, M. Griswold, S. M. Murphy, L. Dunbar, and C. B. Forrest: Chronic care improvement in primary care: evaluation of an integrated pay-for-performance and practice-based care coordination program among elderly patients with diabetes. *Health Serv Res* 45: 1763-82, 2010.

35. Campbell, S. M., D. Reeves, E. Kontopantelis, B. Sibbald, and M. Roland: Effects of pay for performance on the quality of primary care in England. *N Engl J Med* 361: 368-78, 2009.

36. Beaulieu, N. D., and D. R. Horrigan: Putting smart money to work for quality improvement. *Health Serv Res* 40: 1318-34, 2005.

37. Bardach, N. S., J. J. Wang, S. F. De Leon, S. C. Shih, W. J. Boscardin, L. E. Goldman, and R. A. Dudley: Effect of pay-for-performance incentives on quality of care in small practices with electronic health records: a randomized trial. *JAMA* 310: 1051-9, 2013.

38. Serumaga, B., D. Ross-Degnan, A. J. Avery, R. A. Elliott, S. R. Majumdar, F. Zhang, and S. B. Soumerai: Effect of pay for performance on the management and outcomes of hypertension in the United Kingdom: interrupted time series study. *BMJ* 342: d108, 2011
.

39. Doran, T., E. Kontopantelis, J. M. Valderas, S. Campbell, M. Roland, C. Salisbury, and D. Reeves: Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. *BMJ* 342: d3590, 2011.

40. : National Diabetes Statistics Report. National Center for Chronic Disease Prevention and Health Promotion - Centers for Disease Control and Prevention, 2014.

## APPENDIX

**Signed Letter with Permission to Use Copyright Material**

May 8th, 2015

Dear Dr. Kathleen Mullen,

I am writing to request permission to use the following material from your publication in my doctoral thesis:

*Figure 1: Quality Determination Under P4P*
*"Can you get what you pay for? Pay-for-performance and the quality of healthcare providers", 2010, RAND Journal of Economics*

This material will appear as originally published. Unless you request otherwise, I will use the conventional style of the Graduate College of the University of Illinois at Chicago as acknowledgment.

Thank you for your kind consideration of this request.

Sincerely,

Joe Ornelas
PhD Candidate
Health Policy and Administration
School of Public Health
University of Illinois at Chicago
2035 W Taylor St, Chicago, IL 60612

Author Approval

Signature: ______________________

Print: ______________________

Date: ______________________

# VITA

NAME: Joseph Ornelas III

EDUCATION: B.S., Community Health – Planning and Administration, University of Illinois, Urbana-Champaign, Illinois, 2001

M.S., Health Systems Management, Rush University, Chicago, Illinois, 2003

D.C., Chiropractic, National University of Health Sciences, Lombard, Illinois, 2009

A.M., Public Policy, University of Chicago, Chicago, Illinois, 2010

TEACHING: Division of Education, American College of Chest Physicians, Methodology for Developing Evidence-Based Clinical Practice Guidelines, 2012

Department of Health Systems Management, Rush University, Healthcare Quality, 2013 – Present

HONORS: Administrative Fellowship, Centegra Health System, Crystal Lake, Illinois, 2003-2004

PROFESSIONAL MEMBERSHIP: American Society of Healthcare Economists
American College of Healthcare Executives
Academy Health
Academy of Management

PUBLICATIONS:

1. Greeley, S. A., P. M. John, A. N. Winn, J. Ornelas, R. B. Lipton, L. H. Philipson, G. I. Bell, and E. S. Huang: The cost-effectiveness of personalized genetic medicine: the case of genetic testing in neonatal diabetes. Diabetes Care 34: 622-7, 2011.

2. Lewis, S. Z., R. Diekemper, J. Ornelas, and K. R. Casey: Methodologies for the development of CHEST guidelines and expert panel reports. Chest 146: 182-92, 2014.

3. Ornelas, J., J. R. Dichter, A. V. Devereaux, N. Kissoon, A. Livinski, M. D. Christian, Care Task Force for Mass Critical, and Care Task Force for Mass Critical: Methodology: care of the critically ill and injured during pandemics and disasters: CHEST consensus statement. Chest 146: 35S-41S, 2014.

4. Taichman, D. B., J. Ornelas, L. Chung, J. R. Klinger, S. Lewis, J. Mandel, H. I. Palevsky, S. Rich, N. Sood, E. B. Rosenzweig, T. K. Trow, R. Yung, C. G. Elliott, and D. B. Badesch: Pharmacologic therapy for pulmonary arterial hypertension in adults: CHEST guideline and expert panel report. Chest 146: 449-75, 2014.

5. Criner, G. J., J. Bourbeau, R. L. Diekemper, D. R. Ouellette, D. Goodridge, P. Hernandez, K. Curren, M. S. Balter, M. Bhutani, P. G. Camp, B. R. Celli, G. Dechman, M. T. Dransfield, S. B. Fiel, M. G. Foreman, N. A. Hanania, B. K. Ireland, N. Marchetti, D. D. Marciniuk, R. A. Mularski, J. Ornelas, J. D. Road, and M. K. Stickland: Prevention of Acute Exacerbations of COPD: American College of Chest Physicians and Canadian Thoracic Society Guideline. Chest 147: 894-942, 2015.