Evaluation of a Prescription Drug Pay-for-Performance Program: Impact on Prescribing Trends, Budget, and Patient Outcomes

Elizabeth Baumgardner James

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2015

Reading Committee:
Thomas K. Hazlet, Chair
Louis P. Garrison, Jr.
Ryan N. Hansen

Program Authorized to Offer Degree:
Department of Pharmacy

University of Washington

Abstract

Evaluation of a Prescription Drug Pay-for-Performance Program:
Impact on Prescribing Trends, Budget, and Patient Outcomes

Elizabeth Baumgardner James

Chair of the Supervisory Committee:
Associate Professor Thomas K. Hazlet
Department of Pharmacy

Background

Health plans have incentivized providers for patient care that follows evidence-based prescribing, including through pay-for-performance (P4P) programs. Research has suggested increased cost-sharing for prescription drugs is associated with lower adherence and higher utilization of other health care services. Therefore, one potential way to lower overall health care costs is to optimize prescription drug expenditures by promoting an increased proportion of generic drug use. This study evaluated a state-sponsored, clinic-level prescribing P4P.

Methods

We performed a retrospective cohort study evaluating Uniform Medical Plan's (UMP's) 2007-2008 Generic Incentive Program (GIP). Using administrative claims data, we compared annual generic dispensing rates (GDRs) of participating and non-participating clinics using generalized estimating

equations (GEE). Using interview survey data, we sought to understand program elements that may be important predictors of engagement in and success with this type of program. We calculated budget savings as a result of increased generic dispensing rates (GDRs) by GIP clinics compared to the costs of administering the program. Finally, we compared emergency department visits and hospitalizations for patients who sought care from participating and non-participating GIP clinics, also using GEE.

Results

Program participants had a non-statistically significant difference in GDR during the GIP. Most clinics had clinical infrastructure to support informed prescribing and a contemporaneous incentive program that targeted prescribing. Few clinics informed their providers that their performance was being evaluated, shared incentives with them, or discussed quarterly prescribing report cards. Program administrative costs amounted to $250,795, which was the total budget impact of the GIP. The 2.61% systematically higher GDR by GIP participants year over year amounted to a two-year savings of $1,998,633. We found no association between program participation and increased odds of emergency department visits or hospitalizations.

Conclusion

Clinic-level P4P program participation and increased GDR were not related. UMP's GIP participating clinics demonstrated a savings of over $1.9 million resulting from systematically increased GDRs without compromising patient health-related outcomes. Overall, the results of this study suggest that stronger incentives to increase generic prescribing could produce substantial cost-savings without compromising patient health-related outcomes.

Dedication

For Calleigh (3 years) and Troy (18 months):

May you love to learn all the days of your lives,
courageously and positively impacting the world around you as a result.

I love you.
- Mama

Table of Contents

Chapter 1:

Economic Theory and Pay-for-Performance: A Review of Models and Incentives

Abstract

Introduction

The health care industry's implementation of pay-for-performance (P4P) contracting is adapted from other industries where employees are contractually incentivized to perform such that they would maximize firm profits by delivering valued outcomes.

Approach

We conducted a literature review to identify theory, philosophy, and background related to P4P in general and health care in particular.

Discussion

The fundamental economic paradigm behind incentive contracting is the principal-agent relationship and its inherent information asymmetry. In order to maximize firm profits, employers attempt to motivate behavior change through incentive contracting because they cannot directly observe or control the employee's intrinsic motivation and effort capacity. Incentive contracts require one or more performance measures by which to reward output. Ultimately, the cost of the performance measure(s) should be less than the benefit the employer receives. The Institute of Medicine (IOM) has been a catalyst for movement in the quality of health care, with its work leading to payers focusing on quality of their contracted providers through P4P programs. The goals of a P4P program are to improve and to reduce variation in health care and outcomes. Number and type of measures, incentive size and distribution, risk adjustment, and reporting are all aspects of program framework to be considered during development. In assessing P4P program effectiveness, the most recent systematic review concluded we still await conclusive findings on the long-term impact to payer budgets and health-related patient outcomes.

Conclusions

Multiple players, multiple stakeholders, and a multi-faceted, fragmented health care system make the U.S. health care a complex environment for P4P programs. As newer versions of payment for quality programs are considered, program developers should ensure their programs reflect the IOM's dimensions influencing health care quality, the principal-agent model, and collaborative work with providers. These concepts provide a useful framework for P4P program development and may lead to more successful efforts to implement P4P.

Introduction

The health care industry's implementation of pay-for-performance (P4P) contracts is adapted from other industries, where employees are contractually incentivized to perform such that they would maximize firm profits, without altering employees' intrinsic motivation to do their work. Incentive contracting is intended to be a positive source of motivation, i.e., a "carrot," to encourage productivity in ultimate alignment with firm goals. This overview describes this imperfect, well-intentioned system from its inception within traditional firms to its deployment in health care, including how these principles may be applied to current and future P4P programs.

Approach

We sought to describe the history of performance incentives in health care, including their theoretical underpinnings and application, and the framework for P4P programs.

We conducted a search of the English-language literature in PubMed and Web of Science to find articles published on or before May 19, 2015, to identify theory, philosophy, and background related to P4P in general and health care in particular. We used a combination of MeSH terms and text phrases. We used the MeSH term "reimbursement, incentive" and text phrases "pay for performance," "performance incentive," "quality based purchasing," and "health care purchasing."

Among the articles identified, we reviewed those that focused on economic theory-based incentive contracting structure, behavior of individuals and groups in incentive contracts, physicians as agents in P4P, and the application of P4P to health care delivery systems. These manuscripts described the economics of incentive contracts in firms, including their application, types of physician compensation, the impetus for performance-based pay in the United States (U.S.) health payer community, and health care P4P structure, features and development.

This chapter is divided into the following concept discussions:

1. Theory and Principles behind P4P
2. Physicians as Agents in P4P

3. P4P in the United States Health Care System
4. Health Care P4P Program Framework

We conclude with a summary of the concepts, followed by directions for future research and P4P program development.

Theory and Principles Behind P4P

Studies in and theories about incentive contracts were first published in the business and economics literature, with principal-agent theory at the root of discussions and mathematical proofs. Holmstrom and Milgrom stated this application of principal-agent theory as "allocating risks and rewarding productive behavior" in worker compensation.(1) Compensation can take one of three forms: flat pay, also called salary or fixed wage; piece rate, which is pay solely according to a performance measure, like commission-only pay; or a hybrid of the two. Fixed-wage versus performance-based pay has also been described as paying for inputs (hourly wage or annual salary) versus paying for outputs (number of widgets produced).(2)

The necessary relationship of a principal and an agent is created by information asymmetry. A principal has an objective and enlists the help of the agent to achieve that objective. In the workforce, the principal is represented by the employer, often the owner or chief operating officer of a company, and the agent is represented by the employee. The information asymmetry is a result of the employer having specific objectives he wishes to be met without knowing the exact effort or intrinsic motivation of his employee. Intrinsic motivation can range from the desire to be financially stable to the employee's passion for his work. If an employee's exact effort, skills, and intrinsic motivation were directly measurable, piece rate payment schemes would dominate the market.(3)

The main difficulty in establishing successful performance-based pay is the mismatch between the objective of the owner/officer and accurate measurement of objective achievement using an employee performance metric. For example, for a publicly owned company that makes widgets, the

owner's objective is to maximize profits, and therefore, firm value, to satisfy his shareholders; he does not have an individual employee performance metric to show how satisfied his shareholders are or to reflect the value of his firm. However, he can measure the employee's performance with an imperfect measure, such as his widget output over a given period of time.

The employer's objective, maximizing firm value, V, and the employee's performance, P, are both characterized by e, the employee's actions, and ε, the state of the world. The employer's function V(e, ε) is not the same as the employee's function P(e, ε) because of the information asymmetry between the parties indicating the employee's level of effort. The higher the correlation between the firm's objective and the performance measure, the more efficient and effective a pay incentive is in maximizing firm value and the higher the incentive should optimally be.(2, 4) The better the information the agent has on the state of the world, the more likely the compensation will include a piece rate or commission.(4)

In sharp contrast to public companies, nonprofit organizations and government agencies do not have a particular firm value to maximize and may have even more difficulty determining a performance metric, however imperfect, to measure achievement of their primary objective. When the cost of developing an acceptable performance metric exceeds its benefit, organizations may not develop performance incentives and instead rely on the employees' intrinsic motivation, i.e., their desire to be productive, in meeting the employer's primary objectives. Therefore, those employees receive their rewards in the form of promotions and peer recognition.(2)

Employee response to performance-based pay depends on the employee's level of risk aversion, the incentive intensity, and employee anticipation of future performance-based pay schemes. An employee's level of risk aversion impacts how he responds to performance incentives. The more pay is attributed to any performance measure, the more risk the employee must bear. That said, the less risk averse the employee, the better the correlation between firm value and the performance

measure. Holmstrom and Milgrom suggest that job restructuring and relative performance evaluation are mechanisms to balance an employee's risk aversion leading to forsaking certain tasks in favor of those being measured as part of a payment scheme.(1) Generally, the employee has more than one responsibility or one multi-faceted responsibility, such as "producing a high volume of good quality output."(1) The more risk averse an employee, the more performance-based pay will lead him to minimize the risk involved, which often takes the form of forsaking quality for quantity, or whatever method decreases his risk of losing performance-based pay.(2) Holmstrom and Milgrom reference the now age-old example of teachers' performance-based pay for student test scores. (1) Depending on the payment scheme and the risk aversion of the individual teacher, he or she may decide to focus teaching solely those skills more likely to influence his students' test scores at the expense of teaching other important skills i.e., "teaching to the test." Further, this performance measure is highly unlikely to reflect the primary objective of the school system. Employees who already have high intrinsic motivation and/or are productive are more likely to self-select into employment that offers incentive pay. The benefit of the incentive outweighs the cost of being observed.(3)

Incentive intensity also impacts employee behavior, particularly the risk-averse employee. Incentive intensity is the ratio of the piece rate to the wage and can be represented graphically as the slope of their relationship.(5) Zenger and Marshall performed a regression analysis to determine predictors of incentive intensity. They believed that performance-based pay increases employee effort, attracts talent to the organization, and establishes a high-performing organization. In this analysis, they found that smaller numbers of evaluated employees, longer history of an incentive pay scheme in the organization, and schemes with relatively fewer quality measures were associated with higher incentive intensity. The authors' conclusions to explain these quantitative findings were that smaller groups allow employees more control over performance measures, employees are more

trusting of employers' incentive plans the longer the plans have been in effect, and quality measures are less precise than quantity measures.(5)

Agents, whose performance is being measured, may anticipate that if they perform at a high level, the goal threshold will be increased or their incentive payout will decrease, an experience of the "rachet effect."(5, 6) Therefore, agents may decide not to perform initially at their highest potential to keep from providing too much information on their performance capacity to the principal.

Physicians as Agents in P4P

When P4P schemes are deployed by health plans and insurers two principal-agent pairs exist: the health plan-physician relationship and the physician-patient relationship. This complication creates a chain of behavior influence from payers to medical groups to individual providers to patients. The physician has been characterized as a "double agent," because of his agency in providing health care services to an individual patient and his agency to the payer as a steward of resources.(7) Because of the barriers to entry into the medical profession, physicians are their patients' agents. Information asymmetry exists in both directions in this relationship: patients have more knowledge of their behaviors in response to their physicians' treatment plans and physicians have more knowledge of the effectiveness of potential treatments.(8) In some cases, a physician may or may not be rewarded by a health plan based on the ultimate actions of his patients.(9) Additionally, providers may feel that P4P programs create a tension between patient-centered care and population-level health outcomes, which may present a quandary to providers, depending on their practice patterns.(10)

Physicians often perform their profession as members of a group, such as in a multi-physician group practice setting, which leads to a discussion of individual versus group performance. While some of an individual's preferences dictate his performance as an individual regardless of the environment, group sociology literature contributes valuable information to understand how an

individual's beliefs and values lead him to perform in a certain way within a group. In addition to risk aversion, income norms, effort norms, and mutual help activities have been studied and found to contribute to physicians' behavior in response to incentive pay.(11) Whether or not the incentives are paid to the group or to the individual physician determine how these informal sociological interactions influence the individual physician's performance, either increasing or decreasing the incentive's marginal cost to the individual physician. Encinosa et al. describe these interactions as activities and psychological experiences. In a group physician practice setting, doctors share certain expenses, such as the cost of equipment, space, and administration. Total clinic net income is the combined revenue of individual doctors minus these expenses. When incentive pay is part of an individual doctor's revenue, his utility is defined as:

Utility = fraction of revenue kept –
(private cost of effort x effort) – expected cost of outcomes

An individual physician's expected cost of incentives is composed of his risk aversion and his response to income norms, effort norms, and mutual help activities. Income norms refers to how much better or worse off an individual is when comparing his income to that of other group members. The greater the range in pay within a group, the greater role income norms plays in that group, as well as the in determination of the cost and benefit of incentive pay. Encinosa et al. define effort norm as "the informal interactions that make it costly for individuals to perform below the level of others in their work group." This norm is driven by peer pressure, real or perceived, not by dollars. Mutual help activities are interactions of members working together in a group not motivated by dollars. The greater the individual incentive intensity, the less likely an individual is to spend time helping other members in the group.(11)

The following trends have been noted as to how these interactions play out in group physician practices:

1. As group size increases, income inequality increases or equal sharing likelihood decreases.
2. With increasing group size and incentive intensity, risk aversion and mutual help activities decrease, while effort norms and income norms increase.(11)

The economics of payment incentives is applied to the physician practice setting in two ways: (1) choosing from among fee-for-service (FFS), capitation, and salary for general compensation and (2) incorporating P4P measures into compensation. FFS, a retrospective form of physician payment, is an example of piece-rate payment, while capitation, a prospective form of physician payment, is an example of flat pay.(12) Each type of compensation has a direct and linear relationship encouraging some aspect of the physician's performance, as listed in Table 1.

FFS payment arrangements have been argued to promote productivity, patient service, and risk acceptance while risking overtreatment and overbilling. Capitation, on the other hand, has been argued to promote efficiency, cooperation, and evidence-based medicine while risking adverse selection, avoidance of sicker patients, and discharging care prematurely, such as from a hospital stay. (12) For these reasons, capitation-type payment systems can also be referred to as "supply-side cost sharing," because of the supplier's (i.e., the provider's) financial risk-bearing.(13) A hybrid mechanism, such as flat pay plus some form of performance-based pay may keep the best of both FFS and capitation payment schemes but risks physicians practicing the measured practice patterns at the expense of others. Performance-based pay in health care has been termed "supply-side cost sharing."(13)

The question of the carrot versus the stick to change behavior is a timeless one. In her Health Economics editorial, Jody Sindelar describes the move from negative incentives of "correcting externalities, providing information and protecting youths" to positive incentives to change patient behavior. (14) Her examples of the sin tax for alcohol and tobacco products and legal system

penalties for drunk driving and drug trafficking do not retain freedom of personal choice and are too stringent, which she claims makes these ineffective methods to change behavior. However, giving positive incentives for ceasing bad behavior implicitly compensates those who choose these bad behaviors while doing nothing to reward those who already practice good behaviors. This type of reward system could initiate a landslide of unintended consequences.

P4P in the United States Health Care System

In his 1963 landmark manuscript describing the political economy of the U.S. health care system, Kenneth Arrow defined ideal insurance. His example was "a system in which the payment to the physician is made in accordance with the degree of benefit."(8) This was an early description of P4P in health care. Maybe Dr. Dennis O'Leary, President Emeritus for 21 years of the Joint Commission on Accreditation of Healthcare Organizations had a premonition of things to come when he said, "When I came out of school [in the mid-1960s] I thought that someone would be measuring my performance, and that I would be paid accordingly. That seemed logical to me. But it turned out that I could do as well or poorly as I wanted and get paid anyway." (15)

The Institute of Medicine (IOM) has been a catalyst for movement in the quality of health care. The IOM's Committee on Quality of Health Care in America produced its first report "To Err is Human" in November 1999, highlighting the prevalence of patient safety issues in medicine. In 2001, this report was followed by "Crossing the Quality Chasm," the Committee's second and more complete description of quality issues in medicine. The Committee identified six dimensions that influenced health care quality: safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity (listed in Table 2). Donald Berwick's "A User's Manual for the IOM's 'Quality Chasm' Report" described how these dimensions are influenced at four levels: "the experience of patients; the functioning of small units of care delivery ("microsystems"); the functioning of the organizations that house or otherwise support microsystems; and the environment of policy, payment, regulation,

accreditation, and other such factors, which shape the behavior, interests, and opportunities of the [microsystems' support organizations]."(16)

Given the incentives under insurance for "overutilization" as a result of moral hazard, health plans and institutions employ a variety of mechanisms to try to encourage efficient use and limit unsafe use of medical technologies and prescription drugs, as an aspect of quality oversight. Kenneth Arrow identified the need for third-party control of health care utilization as a result of moral hazard. He states, "Insurance removes the incentive on the part of individuals, patients, and physicians to shop around for better prices for hospitalization and surgical care. The market forces, therefore, tend to be replaced by direct institutional control."(8)

The IOM's report created a market paradigm shift in the health care market by the payer community, invoking a form of this "direct institutional control," where quality became the focus. Quality would be assessed and, potentially, monetarily rewarded, i.e., paid according to performance. Quality was defined loosely as "administering evidence-based medical care." (17) This compensation arrangement was to be contrasted with the older compensation mechanisms such as capitation and FFS. However, some have argued that P4P is simply a variation of FFS.(17, 18) Nicholson and colleagues suggest P4P could become an "optimal" FFS compensation arrangement if payment was directly tied to improvement in health outcomes.(18)

With the application of P4P to health care came the translation of principal-agent theory to principles governing incentive contracts in health care delivery systems. Town and colleagues have summarized the applications of the principal-agent model to health care P4P into five issues, reinforcing the principles of P4P described in the principles section of this chapter:

1. "Financial incentives, if large enough, change behavior."
2. "Individuals respond to what is measured and rewarded."
3. "If individuals can game the reward system to their advantage, anticipate that they will do so."

4. "In the face of asymmetric information, the optimal incentive structure often (but not necessarily so) has the following structure: a fixed payment component (e.g., salary) and a component that is tied to the outcome that is measurable and desired."
5. "The institution of pay-for-performance incentives likely has two effects: behavior modification and the selection of more productive physicians into the practice."(19)

Providers and patients may both benefit when a provider's market includes multiple payers with P4P programs in place, when all programs are evidence-based and include similar measures. Providers deliver the same evidence-based services to all patients, earning incentives from multiple sources, and patients not covered by P4P payers receive evidence-based services.(20) Practice patterns and behaviors established by P4P programs may improve provider care in non-measured areas; this has been termed the "halo effect."(21) However, the converse may also be true: if providers are responding to multiple performance programs that are not similar or have too many measures, programs' intended effects may not come to fruition. As mentioned earlier with the example of teachers "teaching to the test," the possibility exists that providers will focus solely on rewarded measures or those with higher rewards when among many.(22, 23) Another potentially negative effect of performance-based incentives is a concern adopted from the economics and business literature: decreased intrinsic motivation of physicians, as the external incentives replace or "crowd out" intrinsic motivation, creating instead a marketplace transaction.(24, 25) However, improving certain internal processes, especially electronic record-keeping or other technical improvements, may lead to "spillover" effects in a practice; unmeasured metrics improve with the global investment.(23)

Unobservable patient variables in health plan P4P may influence outcomes and include physical activity, diet, hygiene and honesty as well as the relationship between the physician and patient, e.g., trust, respect, and motivation.(26) This dynamic may also include self-selection of both physicians and patients to a practice and to each other.(12, 19) Another difference with P4P in health care versus other industries is the relatively "public" knowledge of health care P4P results.(12) Rosenthal

and colleagues suggest P4P is an alternative to market forces alone, which have been unable to "enforce quality competition."(27)

Providers as Stakeholders

While payers and patients are stakeholders in P4P programs, the double-agent provider community in the middle must attempt to meet the expectations and requirements in relationship with both parties. The American College of Physicians (ACP) and the American Medical Association (AMA) have been vocal regarding P4P. ACP's position on P4P programs includes the need for transparency before patients of providers' participation and rankings in P4P programs, use of measurements that are important to patients, and monitoring providers for unethical behavior.(17) In an effort to standardize measures across P4P programs, the AMA formed the Physician Consortium for Performance Improvement (PCPI) in the early 2000s. At that time, PCPI developed a list of 184 measures with the hope that program sponsors would adopt them, invoking standardization.(15)

Providers must also consider the relationship between financial incentives and the health of individual patients. A provider's opportunity cost, or the "shadow price of patient health" will also determine how he responds to performance-based payments.(22)

In 2005, in an effort to communicate their position as stakeholders, Forrest and colleagues held a summit of 250 physicians convened by the disease management firm, Healthways, from which a list of physician preferences for P4P programs was developed (28). The following list includes the key preferences:(29)

1. Groups rather than individual providers should incentivized.
2. Local markets (i.e., providers, payers, and employers) should coordinate P4P program development.
3. Improvement targets should be included in addition to threshold targets.
4. Assessments should include risk-adjustment for practice case mix.
5. Providers should be permitted to validate the results presented to them.
6. Results should be made publicly available.
7. Programs should be periodically evaluated to insure the intended results are those detected in the data.

These preferences hold true today, and have been included in manuscripts outlining suggested best practices for P4P program framework, which is discussed in the next section. However, there is some discussion over the difference in program success in group versus individual provider incentives. While group incentives require alignment within the group, particularly among leadership and may allow for reinvestment in infrastructure and less patient shifting between practice providers, individual incentives may improve individual outcome measures.(21, 30, 31)

Health Care P4P Program Framework

The goal of a P4P program is to reduce variation in health care and outcomes, as well as increase safety, quality, transparency and productivity of systems and processes.(32) Nicholson and colleagues describe parameters for a successful P4P program in terms of the "health production function," which they define as "information about specific treatments and clinical management that can transform people with an initially poor health status into those with better outcomes."(18) The conditions that must be satisfied for a P4P program to be effective are: (1) providers having more complete and/or accurate understanding of the health production function and (2) payers' ability to measure outcomes as they compare to baseline measurements. If payers knew as much or more of the health production function than providers, reimbursement could simply be structured around services known to improve the health production function and P4P would be irrelevant.(18) The list of preferences resulting from the Healthways' summit is an appropriate starting point when developing a P4P program.

The number of measures is important to consider. An optimal number of measures exists where the marginal benefit of adopting a measure is no longer positive but zero. Beyond that point, performance may decline.(19) Figure 1 displays this relationship.

Practice size may be a consideration for structuring a program. Robinson and colleagues studied medical groups receiving incentive payments and defined a large medical group as one with 20 or

more providers. Their hypothesis was that groups of this size may be more successful because they have larger numbers of patients, more opportunity for collaboration among providers, and more resources available to invest in infrastructure improvements that may be required for program reporting.(33)

Despite the popularity and prevalence of incentive programs, individual providers may be unaware they are participating in any program. One Medicaid program analysis showed an association between improvement in quality measures and provider knowledge of program participation. Another incentive program evaluation showed no improvement in quality measures; however, half of the providers were unaware they were participating.(34) A third study confirmed provider awareness with again only half of the providers aware of their participation.(35) These results suggest communication of incentive programs to participating providers may influence the degree of improvement in quality measures.

Part of P4P program development includes determining the target of the incentive: individual providers or groups of providers. One advantage for targeting individual providers is that individual providers will be more motivated than if they were evaluated as a group.(9) Targeting groups of providers promotes cooperation among group members and may enable infrastructure improvements that lead to improved quality overall, but some individuals may become "free riders."(9, 29) Providers have a higher chance of meeting set target measures when their costs and reporting burden for participation are low.(36) Two examples would be improved office technology, such as computerized clinical reminders, or academic detailing to support participation.(37)

Program developers may consider allowing providers to exclude higher risk patients, thus creating a sort of "risk adjustment" to their metrics. An alternative is to perform post-risk adjustment according to a physician's or practice's calculated risk, which was shown to dramatically increased payments. (38) As stated earlier, this permission resulted as the best predictor of how

much providers were paid. Furthermore, providers may find that complex patients with comorbidities that preclude them from being subject to first-line or the most cost-effective care, which may be detrimental to providers' overall metrics.(17) As Forrest and colleagues state, "Patients do not randomly distribute themselves across providers."(29) Certain providers may see sicker patients, on average, and thus some type of risk adjustment may be warranted. Additionally, progressive and/or chronic disease may cause a decline in outcomes, regardless of care provided.

Reward payments should be made as close to the performance timeframe as possible for providers to make the connection between their actions and reward outcomes.(36, 37) Recent research also advocates for a hybrid approach to payments, distributing some at the clinic level and others at the individual provider level to optimize motivation and behavior.(31)

A P4P program framework is not without the possibility of unintended or negative consequences. Provision of cost-effective care and patient health are difficult to directly and easily measure in a manner that health plans are able to incentivize. As suggested by Baker and Holmstrom and Milgrom, these imperfect measures may lead to weak incentives or a substitution of effort from one or more duties to others.(1, 2) In health care, this may lead to patient dumping, adverse patient selection, or focus on the measures incentivized at the expense of other patient care.(39) In addition to adverse selection, reduced access to care may result when providers care disproportionately for underserved populations. These practices and hospitals may not be able to financially sustain themselves if payments are at risk.(40) Both payers and providers can become "free riders" when P4P programs are in place. If only one regional payer has a program, other payers benefit from the intended effects of the program if providers service other payers.(9) If clinics receive rewards in aggregate, providers who do not fully participate will benefit from colleagues' participation.(37) Finally, because providers may be tempted to game the system, such as upcoding services, program developers may consider some type of audit requirement for participants.(36)

P4P Measure Selection and Setting

Measures may be best designed when subjected to the project management requirement "SMART," or specific, measurable, attainable, relevant and time-bound.(36) The types of measures include many from inpatient, outpatient, pharmacy, and patient perspectives. Metrics can be classified as clinical, efficiency, technology, and patient satisfaction.(41) Table 3 displays a sample list of measures.(41-43) Metrics have also been categorized into domains of quality, access, prevention and coding quality and timeliness.(36) These domains and categories reflect the IOM's dimensions influencing health care quality.

Examples of disease states with metrics derived from evidence-based treatment guidelines include angina, asthma, atrial fibrillation, chronic obstructive pulmonary disease, depression, diabetes, heart failure, hypercholesterolemia, hypertension, osteoarthritis, osteoporosis, otitis media, smoking cessation, and substance abuse.(17, 41) P4P programs with prescription benefit measures are less common, perhaps because prescriptions benefits usually have built-in restrictive mechanisms intended to contain cost, such as prior authorization and step therapy requirements, as well as preferred drug lists or formularies. However, prescription benefit P4P measures do not directly restrict prescription drug use at the point of service like the more common cost-containment measures do. Increased rates of generic prescribing are more cost-effective for both patients—because of their lower cost-share, and payers because of the difference in drug cost between brand and generic drugs.

Ideal or "true" performance indicators are process or outcome measures known to decrease morbidity, mortality and/or disability.(44, 45) Indicators available may be a reflection of process, i.e., did the provider perform the service, or a surrogate, or proxy, measure of morbidity, mortality, and/or disability. If the process is a perfect determinant of outcome, then the process is appropriate as a measure.(18, 46) In other words, if the process of measuring HbA_{1c} in diabetic patients (i.e., did

the provider measure HbA_{1c}?) ensured low rates of morbidity, mortality and/or disability, then reporting the rate of measuring HbA_{1c} would be ideal. Another step between process and outcome that is important is whether the measure is known to predict morbidity, mortality or disability. For example, evidence must demonstrate that HbA1c below 7.4% reduces the risk of morbidity, mortality and/or disability.

Measures can be set as benchmark targets or percent improvement from baseline or previous time period. One hazard of using benchmark targets only is the prominence of rewarding providers and clinics that are already high-performing, essentially rewarding their past achievement. (34) Likewise, targets must be above a theoretical change point where poor performers are no longer motivated.(45) As an example of setting relative targets, a large PacifiCare P4P program based its targets on the 75th percentile of performance the year prior to program implementation.(47)

Payment and Reporting

Incentives can be categorized as financial or nonfinancial. Financial incentives are monetary. Nonfinancial incentives may be a provider's or clinic's ranking or reputation.

Incentive money comes only from two places: withhold of a percentage of compensation or "new" money—money paid as a bonus on top of usual compensation. The Centers for Medicare and Medicaid Services (CMS) has established expectations and payment strategies related to performance. For hospitals, 0.4% of payments were withheld unless they reported on 10 quality indicators. This small percentage was increased to 2% with the Deficit Reduction Act.(48) Christianson and colleagues' review of the evidence on P4P programs, published in 2008, evaluated programs initiated between 1987 and 2004, including Medicaid managed care and commercial plans. Their research noted payments constituted anywhere from 0.5-12% of total compensation, and up to $270,000 annually.(27, 34) A systematic review of programs with preventive measures looked at six studies; those that provided payment information had a range of $50-8,682.(26) A study of

hospital P4P showed a 5% bonus was enough to motivate providers.(40) A study of physicians showed a 10% bonus was enough to motivate and sustain their practices in P4P.(49) Because research indicates that higher payments in general are more motivating to providers and that can be achieved through local health plan collaboration.(50) However, no payment range has been identified for incentive size that incurs diminishing returns.(32)

The IOM recommends performance data be made available to payers and patients; this induces a "reputational incentive," whereby providers respond to P4P programs based on the value they place on their public recognition.(51) The positive impact to a provider's practice or ego may be a stronger incentive than a monetary incentive.(29) Simply achieving a bonus or failing to meet the target in a competitive market may be incentive enough "where providers place a premium on reputation."(36) Finally, reporting alone may be enough to change behavior. An individual physician's awareness of his performance, compared to peers or top-ranking institutions, combined with intrinsic motivation to perform according to the best evidence, may induce behavior change that benefits payers, institutions and patients.(52)

Established P4P Programs and Outcomes Evidence

A multitude of quality programs, organizations and initiatives emerged in the U.S. in the 1990s and 2000s. Some of the first major P4P programs in the U.S. were identified in Rosenthal and colleagues' 2004 publication titled, "Paying for quality: providers' incentives for quality improvement." Among those were: (1) physician reward programs administered by such groups as Aetna, the Blues, Bridges to Excellence, CIGNA, Harvard Pilgrim Health Care, HealthPartners, Integrated Healthcare Association, and PacifiCare; and (2) hospital reward programs such as Premier and those administered by such groups as Anthem, the Blues, CIGNA, CMS, and Leapfrog. A 2005 study conducted by MedVantage identified 107 active P4P programs, 95% of which targeted primary care physicians, and 52% of which included specialists.(49) Cromwell and colleagues, in their book, also

include 15 early, large P4P programs along with each program's performance measures and financial incentive size.(53)

Outside of the U.S., a noteworthy program because of its duration is the Quality and Outcomes Framework (QOF) deployed by the United Kingdom (U.K.) Government. A goal of the QOF was to increase recruitment of primary care providers, so the government was investing a substantial amount of money in increased pay to these providers.(54) The British Medical Association had rejected the government's mid-1980s attempt to establish a national P4P program, denying the existence of inappropriate or problematic variation in patient care, or that quality could be measured. However, in 2003, 79% of family practitioners (among 70% of U.K. family practitioners) voted to adopt the QOF.(55) The QOF began in 2004, and it continues today. Although a voluntary program, 99.6% of the U.K.'s family practitioners chose to participate. At the outset, the practitioners were measured using 136 indicators of quality, and their payments increased their income by approximately 25%, up to $40,000 in some cases. Indicators included clinical, organizational, and patient experience indicators, one of which was continuity of care, assessed by asking how often patients saw their usual doctor.(56) Published national guidelines, including those of the National Institutes of Clinical Excellence, were used to set clinical indicators.(55) Physicians were permitted to exclude certain patients who could significantly negatively influence measured outcomes but were required to submit an "exception report" to do so. In early evaluation of the QOF, exclusion of these higher risk patients was the best predictor of incentive dollars paid to providers.(34) The first couple years of the program demonstrated positive outcomes for chronic diseases, but once the incentives were removed, outcomes returned to baseline.(56, 57) When providers were surveyed in 2004 and 2005, their response to the program included "greater job satisfaction, decreases hours worked, … increased incomes, … a decrease in autonomy and an increase in administrative and clinical workload."(58) As late as 2011, hospital admissions for

ambulatory-sensitive conditions continued to decline.(54) Systematic reviews specific to the QOF suggest that although some disease-specific morbidity and mortality indicators and both national and local processes and process measures have improved, additional, more robust research is needed to provide conclusive evidence of its overall and sustained impact.(59, 60)

Systematic Review Evidence Base

The most recent P4P systematic review (an international systematic review of systematic reviews) concluded we still await conclusive findings for the long-term impact to payer budgets and health-related patient outcomes.(61) Eijkenaar and colleagues sought to comprehensively examine P4P program outcomes, including domains of effectiveness, cost-effectiveness, unintended consequences, patient inequalities, non-financial incentives and program design. Regarding effectiveness and cost-effectiveness, systematic reviews conducted by different research groups had some discordant conclusions for specific programs, but the overall conclusion of insufficient evidence for P4P program effectiveness or cost-effectiveness remained. Within a given review, different measures, such as cancer screenings versus immunizations, and different disease states, such as asthma versus heart disease, may have had different effectiveness or cost-effectiveness findings. It is possible that the difference in patient burden to complete his or her responsibility in the measure (cancer screening versus immunization) could lead to discordant findings and conclusions. Likewise, it is possible that proxy outcome measures for one disease state (e.g., emergency department visits for asthma) versus another (e.g., death incidence for patients with heart disease) have very different time horizons for ascertainment. Finally, regarding effectiveness and cost-effectiveness, without a proper control group, program evaluators may be required to use projected trends, which have their own inherent biases, as the comparator. The projected trend as comparator proxy is heavily biased at best and completely irrelevant at worst. Regarding unintended consequences, Eijkenaar et al. found a potential reduced continuity of care, provider gaming, and

both positive and negative spillover effects. Regarding patient inequalities, P4P programs reduce them at best and maintain them at least; both geography and specific performance measure or disease state impacted results. Regarding non-financial incentives, Eijkenaar et al. found that results of direct provider feedback versus public reporting of outcomes had different impacts; public reporting may be more incentivizing. Regarding program design, a specific finding of this systematic review was that provider collaboration and communication led to more effective programs. Table 4 is the list of program design elements Eijkenaar et al. found to be consistent among effective programs.

Finally, Eijkenaar and colleagues concluded it is difficult to disentangle effects of P4P programs from effects of other, concurrent infrastructure or quality improvements within health care delivery systems.

Conclusion: State of the Art and Implications

State of the Art

Health care P4P programs take cues from the basic principles governing incentive contract development, complete with information asymmetry within the principal-agent pair. In the payer-physician pair, there is information asymmetry in both directions. The physician is unaware of the full picture of the payer's goal in maximizing dollars spend, and the payer is unaware of the physician's intrinsic motivation in meeting the payer's quality benchmarks. In the physician-patient pair, the physician is unaware of the patient's intrinsic motivation or intent to follow through with a care plan that will ultimately influence a performance indicator. Within the payer-physician pair also exists the persistent tension between population-level health care and patient-centered care, which influences the difference between their respective performance functions. Eijkenaar and colleagues' international systematic review of systematic reviews found that the more effective P4P programs were developed collaboratively between payers and providers.(61) This tactic could serve to reduce

the information asymmetry between payers and providers and encourages better alignment between the payer and provider(s) performance functions.

The framework of a P4P program must consider that a provider's performance is based on risk aversion, incentive intensity, and anticipation of future performance-based pay.(1) The extent of risk aversion may lead to a provider's focus on measured practice patterns at the expense of others (i.e., teaching to the test) or adverse selection of patients. Maintaining a smaller list of the most appropriate outcome measures—or well-accepted proxy outcome measures—and either allowing patient exclusion in measurements or invoking risk-adjustment in payments could alleviate these risk aversion behaviors. Higher incentive intensity has been found to be associated with smaller groups.(5) Similarly, Eijkenaar and colleagues found that the more effective programs had been directed to smaller groups and even individual physicians.(61) For larger groups, perhaps a hybrid of incentive pay given to the group as well to as individual contributors could be considered. Individual incentive pay could be distributed by the health plan or by the practice as a contractual requirement of participation in the P4P program. An individual contributor aspect to incentive pay would also help to reduce the free rider problem in group practices. While significant financial incentives have been found to improve performance in P4P programs,(61) payers must also consider that increased payment potential may result in additional unintended or negative consequences.(54) Payer-provider collaboration on incentive size is another opportunity to reduce information asymmetry that may be contributing to P4P programs' lack of success. Regular communication between payers and providers, including payer distribution of high quality feedback reports could promote greater success. Because of the importance of provider reputation, feedback to providers elucidating their performance compared to peers, best practices, or high-performing institutions may encourage continued provider engagement toward meeting performance targets.(52) Eijkenaar and colleagues' review confirms that "public reporting can be more effective when used together with P4P."(61)

In addition to measures being as close to accurate representations of true morbidity, mortality and/or disability, targets should be set so that poor performers are engaged to participate and higher performers do not quickly reach a plateau.(45, 54) Measures should be continually evaluated for their appropriateness in invoking performance that improves overall care. Program advisors to the QOF are aware of the need to continually evaluate metrics and their targets to ensure their appropriateness and meaningfulness to practice, including whether or not metrics should be removed altogether.(21, 54, 62, 63) This led to the 2012 removal of all organizational indicators that were not able to invoke continued improvement in performance.(54)

Implications for Future P4P Development

The prescription drug P4P evaluation described in the following two chapters highlights the impact of some of the economic principles and P4P design elements that shape its outcomes and success. While many reviews within the recent international systematic review of systematic reviews did not include effectiveness, budget impact and patient outcomes for one program, this research includes all.

Multiple players, multiple stakeholders, and a multi-faceted health care system makes U.S. health care a complex environment for successful P4P program development and implementation. Eijkenaar and colleagues' international systematic review of systematic reviews considers the limitation of the U.S. fragmented health care delivery system as a difficulty in P4P program success.(61) An understanding of theoretical underpinnings of P4P in the health care delivery system sheds lights on research that highlights successes or failures of P4P programs, including hypothesized reasons for failure or attributes that contribute to success.

Today, payer to health system payment for quality is seen in platforms such as patient-centered medical homes, accountable care organizations, and coordinated care organizations, as well as in hospital contracting that audits for the presence of never-events and 30-day readmissions. For the

best chance of successful program development and effectiveness, program developers should refer to the IOM's dimensions influencing health care quality, Town and colleagues' principal-agent model health care applications, and Forrest and colleagues' provider preference lists during conceptual program development.(16, 19, 29) These concepts reflect the basic principles behind incentive contracts, provide a useful framework for P4P program, and may lead to more successful efforts to implement P4P.

Table 1: Linear Relationships between Physician Compensation Mechanism and Performance Component

Compensation Mechanism	Linear Relationship
Fee-for-service	Relative value units billed
Capitation	Enrollment
Salary	Hours worked

Table 2: The IOM's Dimensions Influencing Health Care Quality

1. Safety
2. Effectiveness
3. Patient-Centeredness
4. Timeliness
5. Efficiency
6. Equity

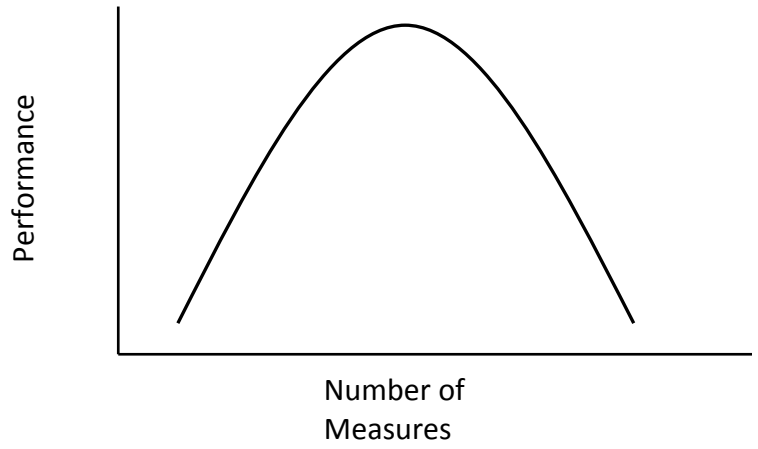

Figure 1: Relationship between Number of Measures and Physician Performance

Table 3: P4P Domains and Examples of Each

Clinical	Efficiency
▪ Cancer screenings ▪ Chronic disease state care ▪ Immunizations ▪ Prenatal care ▪ Preventable/unnecessary hospital admissions ▪ Preventable/unnecessary imaging ▪ Rate of generic drug prescribing ▪ Substance abuse treatment ▪ Obesity reduction	▪ Hospital length of stay ▪ Rates of generic and/or formulary prescription drug use
Technology	**Patient Satisfaction**
▪ Electronic claims submission ▪ Electronic medical records ▪ Electronic prescribing ▪ Patient disease registries	▪ Access to providers and services ▪ Care experience ▪ Care coordination ▪ Communication with physicians

Table 4: Eijkenaar and Colleagues' Design Elements of More Effective Programs[1]

1. Measures are used that have more room for improvement and are easy to track.
2. [Incentives are] directed at individual physicians or small groups.
3. Rewards are based on providers' absolute performance.
4. The program is designed collaboratively with providers.
5. Larger payments are used.

[1]*Text reprinted from Health Policy, 119, Eijkenaar F, Emmert M, Scheppach M, Schoffski O, Effects of pay for performance in health care: a systematic review of systematic reviews, Page 127, Copyright 2013, with permission from Elsevier.*

Chapter 1 References

1. Holmstrom B, Milgrom P. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. Journal of Law, Economics, & Organization. 1991;7:24-52.
2. Baker G. Distortion and Risk in Optimal Incentive Contracts. Journal of Human Resources. 2002;37(4):728-51.
3. Daniel P. The effect of pay-for-performance contracts on wages. Empirical Economics. 2009;36(2):269-95.
4. Baker GP. Incentive Contracts and Performance Measurement. The Journal of Political Economy. 1992;100(3):598-614.
5. Zenger TR, Marshall CR. Determinants of incentive intensity in group-based rewards. Academy of Management Journal. 2000;43(2):149-63.
6. Laffont J-J, Tirole J. The Dynamics of Incentive Contracts. Econometrica. 1988;56(5):1153-75.
7. Blomqvist A. The doctor as double agent: information asymmetry, health insurance, and medical care. J Health Econ. 1991;10(4):411-32.
8. Arrow KJ. Uncertainty and the Welfare Economics of Medical Care. The American Economic Review. 1963;53(5):33.
9. Dudley RA, Rosenthal MB. Pay for Performance: A Decision Guide for Purchasers. Rockville, MD: Agency for Healthcare Research and Quality; 2006.
10. Wharam JF, Paasche-Orlow MK, Farber NJ, Sinsky C, Rucker L, Rask KJ, et al. High quality care and ethical pay-for-performance: a Society of General Internal Medicine policy analysis. J Gen Intern Med. 2009;24(7):854-9.
11. Encinosa III WE, Gaynor M, Rebitzer JB. The sociology of groups and the economics of incentives: Theory and evidence on compensation systems. Journal of Economic Behavior & Organization. 2007;62(2):187-214.
12. Robinson JC. Theory and practice in the design of physician payment incentives. Milbank Q. 2001;79(2):149-77, III.
13. Ellis RP, McGuire TG. Supply-side and demand-side cost sharing in health care. J Econ Perspect. 1993;7(4):135-51.
14. Sindelar JL. Paying for performance: the power of incentives over habits. Health Econ. 2008;17(4):449-51.
15. Darves B. Physician Pay-for-Performance Programs Taking Hold. NEJM Online Career Center. 2007; December.
16. Berwick DM. A user's manual for the IOM's 'Quality Chasm' report. Health Aff (Millwood). 2002;21(3):80-90.
17. Greene SE, Nash DB. Pay for performance: an overview of the literature. Am J Med Qual. 2009;24(2):140-63.
18. Nicholson S, Pauly MV, Wu AY, Murray JF, Teutsch SM, Berger ML. Getting real performance out of pay-for-performance. Milbank Q. 2008;86(3):435-57.
19. Town R, Wholey DR, Kralewski J, Dowd B. Assessing the influence of incentives on physicians and medical groups. Med Care Res Rev. 2004;61(3 Suppl):80S-118S.
20. Hellinger FJ. The impact of financial incentives on physician behavior in managed care plans: a review of the evidence. Med Care Res Rev. 1996;53(3):294-314.

21. Lester H, Schmittdiel J, Selby J, Fireman B, Campbell S, Lee J, et al. The impact of removing financial incentives from clinical quality indicators: longitudinal analysis of four Kaiser Permanente indicators. BMJ. 2010;340:c1898.
22. Conrad DA, Perry L. Quality-based financial incentives in health care: can we improve quality by paying for it? Annu Rev Public Health. 2009;30:357-71.
23. Mullen KJ, Frank RG, Rosenthal MB. Can you get what you pay for? Pay-for-performance and the quality of healthcare providers. Rand J Econ. 2010;41(1):64-91.
24. Conrad DA, Christianson JB. Penetrating the "black box": financial incentives for enhancing the quality of physician services. Med Care Res Rev. 2004;61(3 Suppl):37S-68S.
25. Himmelstein DU, Ariely D, Woolhandler S. Pay-for-performance: toxic to quality? Insights from behavioral economics. Int J Health Serv. 2014;44(2):203-14.
26. Town R, Kane R, Johnson P, Butler M. Economic incentives and physicians' delivery of preventive care: a systematic review. Am J Prev Med. 2005;28(2):234-40.
27. Rosenthal MB, Fernandopulle R, Song HR, Landon B. Paying for quality: providers' incentives for quality improvement. Health Aff (Millwood). 2004;23(2):127-41.
28. Christopher B. Forrest MD, PhD. 2015.
29. Forrest CB, Villagra VV, Pope JE. Managing the metric vs managing the patient: the physician's view of pay for performance. Am J Manag Care. 2006;12(2):83-5.
30. Chung S, Palaniappan LP, Trujillo LM, Rubin HR, Luft HS. Effect of physician-specific pay-for-performance incentives in a large group practice. Am J Manag Care. 2010;16(2):e35-42.
31. Greene J, Kurtzman ET, Hibbard JH, Overton V. Working under a clinic-level quality incentive: primary care clinicians' perceptions. Ann Fam Med. 2015;13(3):235-41.
32. Maynard A, Street A, Hunter R. Using 'payment by results' to fund the treatment of dependent drug users--proceed with care! Addiction. 2011;106(10):1725-9.
33. Robinson JC, Shortell SM, Rittenhouse DR, Fernandes-Taylor S, Gillies RR, Casalino LP. Quality-based payment for medical groups and individual physicians. Inquiry. 2009;46(2):172-81.
34. Christianson JB, Leatherman S, Sutherland K. Lessons from evaluations of purchaser pay-for-performance programs: a review of the evidence. Med Care Res Rev. 2008;65(6 Suppl):5S-35S.
35. Teleki SS, Damberg CL, Pham C, Berry SH. Will financial incentives stimulate quality improvement? Reactions from frontline physicians. Am J Med Qual. 2006;21(6):367-74.
36. Duckett SJ. Design of price incentives for adjunct policy goals in formula funding for hospitals and health services. BMC Health Serv Res. 2008;8:72.
37. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? Ann Intern Med. 2006;145(4):265-72.
38. Damberg CL, Elliott MN, Ewing BA. Pay-for-performance schemes that use patient and provider categories would reduce payment disparities. Health Aff (Millwood). 2015;34(1):134-42.
39. Sherry TB. A Note on the Comparative Statics of Pay-for-Performance in Health Care. Health Econ. 2015.
40. Werner RM, Dudley RA. Making the 'pay' matter in pay-for-performance: implications for payment strategies. Health Aff (Millwood). 2009;28(5):1498-508.
41. Dunn JD. Pharmacy management approach: how do we align all the incentives? J Manag Care Pharm. 2007;13(2 Suppl B):S16-9.

42. Meddings J, McMahon L, Jr. Measuring Quality in Pay-for-Performance Programs. Disease Management & Health Outcomes. 2008;16(4):205-16.
43. Smith AL. Merging P4P and disease management: how do you know which one is working? J Manag Care Pharm. 2007;13(2 Suppl B):S7-10.
44. Gellad WF, Detsky AS, Choudhry NK. Implications of recent clinical trials on pay-for-performance. Am J Health Syst Pharm. 2009;66(9):864-7.
45. Dowd B, Feldman R, Nersesian W. Setting pay for performance targets: do poor performers give up? Health Econ. 2013;22(2):168-79.
46. Doran T. Lessons from early experience with pay for performance. Disease Management & Health Outcomes. 2008;16(2):69-77.
47. Mullen KJ, Frank, Richard G., Rosenthal, Meredith B. Can You Get What You Pay For? Pay-for-Performance and the Quality of Healthcare Providers. Rand J Econ. 2009;41(1):28.
48. Epstein AM. Paying for performance in the United States and abroad. N Engl J Med. 2006;355(4):406-8.
49. Endsley S, Baker G, Kershner BA, Curtin K. What family physicians need to know about pay for performance. Fam Pract Manag. 2006;13(7):69-74.
50. de Brantes FS, D'Andrea BG. Physicians respond to pay-for-performance incentives: larger incentives yield greater participation. Am J Manag Care. 2009;15(5):305-10.
51. Dudley RA, Frolich A, Robinowitz DL, Talavera JA, Broadhead P, Luft HS. Strategies To Support Quality-based Purchasing: A Review of the Evidence. Rockville (MD); 2004.
52. O'Neal BC, Couldry RJ, Wilkinson ST, Cannella CA, Williams CB, Scott LA, et al. Leveraging drug-utilization and external benchmarking data to drive change in prescribing behaviors. Am J Health Syst Pharm. 2012;69(21):1916-22.
53. Cromwell J, Research Triangle Institute. Pay for performance in health care : methods and approaches. Research Triangle Park, NC: RTI Press; 2011.
54. Harrison MJ, Dusheiko M, Sutton M, Gravelle H, Doran T, Roland M. Effect of a national primary care pay for performance scheme on emergency hospital admissions for ambulatory care sensitive conditions: controlled longitudinal study. BMJ. 2014;349:g6423.
55. Roland M. Linking physicians' pay to the quality of care--a major experiment in the United kingdom. N Engl J Med. 2004;351(14):1448-54.
56. Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of pay for performance on the quality of primary care in England. N Engl J Med. 2009;361(4):368-78.
57. Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, et al. Pay-for-performance programs in family practices in the United Kingdom. N Engl J Med. 2006;355(4):375-84.
58. Whalley D, Gravelle H, Sibbald B. Effect of the new contract on GPs' working lives and perceptions of quality of care: a longitudinal survey. British Journal of General Practice. 2008;58(546):8-14.
59. Gillam SJ, Siriwardena AN, Steel N. Pay-for-performance in the United Kingdom: impact of the quality and outcomes framework: a systematic review. Ann Fam Med. 2012;10(5):461-8.
60. Langdown C, Peckham S. The use of financial incentives to help improve health outcomes: is the quality and outcomes framework fit for purpose? A systematic review. J Public Health (Oxf). 2014;36(2):251-8.

61. Eijkenaar F, Emmert M, Scheppach M, Schoffski O**.** Effects of pay for performance in health care: a systematic review of systematic reviews. Health Policy. 2013;110(2-3):115-30.
62. Doran T, Kontopantelis E, Reeves D, Sutton M, Ryan AM**.** Setting performance targets in pay for performance programmes: what can we learn from QOF? BMJ. 2014;348:g1595.
63. Caley M, Burn S, Marshall T, Rouse A**.** Increasing the QOF upper payment threshold in general practices in England: impact of implementing government proposals. Br J Gen Pract. 2014;64(618):e54-9.

Chapter 2:

Impact of a Clinic-Level Pay-for-Performance Program on Generic Drug Prescribing

Abstract

Background

Promotion of higher rates of generic drug use is one important cost control tool for health care expenditures. In addition to restrictive measures before or at the time the prescription is dispensed and sold, health plans have incentivized providers for patient care that follows evidence-based prescribing, including the use of pay-for-performance programs. This study evaluates the response to a clinic-level prescribing pay-for-performance (P4P) program.

Methods

We performed a retrospective cohort study evaluating Uniform Medical Plan's (UMP) Generic Incentive Program (GIP). Using administrative claims data, we compared annual generic dispensing rates (GDRs) of participating and non-participating clinics among continuously enrolled members aged 18-89 years. The impact of the GIP (2007 to 2008) was evaluated using prescription claims data from 2006 to 2009 in the generalized estimating equations (GEE) models. We adjusted for baseline demographics, secular trends, and year-by-participant interaction terms in the model. Additionally, using interview survey data, we sought to identify and understand program elements that may be important predictors of both engagement in and success with this type of program.

Results

Of 25 invited clinics, we evaluated annual GDRs for 16 participating and 8 non-participating clinics. Program participants had an unadjusted 3.09% (p=0.02) absolute difference in annual GDR, compared to program non-participants, taking secular trends into account only. After adjusting for available confounders, there was no significant difference in annual GDR increase during the years of the GIP between program participants and non-participants. Seventeen of the 25 invited clinics responded to the survey. Only two of the respondents were GIP non-participants. Most clinics had clinical infrastructure that would support informed prescribing by clinic providers. Most also had a

contemporaneous payment incentive program that targeted prescribing. Few clinics informed their providers that their performance was being evaluated, shared incentives with them, or discussed quarterly report cards by UMP. Only one respondent believed the GIP impacted prescribing within the clinic.

Discussion

Clinic-level P4P program participation was not associated with an increase in annual GDR. Larger and more diverse incentives and greater collaboration and coordination among providers and local health plans should be further investigated to better understand this health plan cost containment strategy.

Background

U.S. health care expenditures continue to increase, and prescription drugs, as a portion of that expenditure, account for 9.1% of the total national health care expenditures.(1) One important cost control tool is to promote a greater proportion of generic drug use, encouraging providers toward more efficient substitution toward generic drugs within drug classes. The U.S. Food and Drug Administration states that generic drugs are "comparable to a [brand] drug product in dosage form, strength, route of administration, quality, performance characteristics and intended use."(2) Indeed, one key aim of the Drug Price Competition and Patent Term Restoration Act of 1984, (also known as the "Hatch-Waxman Act") was to promote access to lower cost (but equivalent) generic medicines after the patent expiration of branded medicines.(3)

Health plans make many key decisions at a covered-population level, employing a variety of mechanisms to try to encourage efficient use and limit unnecessary and unsafe use of prescription drugs, in the process of optimizing their expenditures. In addition to restrictive measures, such as prior authorization and step therapy algorithms, health plans have incentivized providers for patient care that follows evidence-based guidelines. These restrictive mechanisms can, however, create a temporary barrier between a patient and a medication at the point of service. But other mechanisms, such as provider incentive programs, may avoid this barrier. Provider incentive programs, including pay-for-performance (P4P) programs, are often established by health plans and other payers. Providers are rewarded based upon their ability to meet quality benchmarks, as defined by the governing health plan. Incentive programs set target metrics that may be clinical, efficiency, technology, or patient satisfaction measures, from domains of quality, access, prevention, coding quality, and timeliness.(4, 5) Quality management of certain disease states may be evaluated by metrics that represent evidence-based prevention or treatment. Prescription benefit P4P programs are less common than those developed for medical benefits. These programs encourage the use of

higher value medications, realizing the difference in value between brand-name and generic drugs.(6) Compared to medical benefit incentive programs, which require not only provider involvement but patient follow-through with laboratory evaluations or diagnostic procedures, prescription benefit incentive programs only require patient follow-through by filling the prescriptions written for them.

The body of scientific literature on P4P for prescribing is primarily rooted in incentive schemes intended to manage prescription drug budgetary constraints in countries outside the U.S. These reviews and systematic reviews point to a lack of high quality research or inconclusive research on prescribing P4P programs.(7-9) To date, only one research publication has evaluated a U.S. health plan prescribing incentive program.(10) That study evaluated four interventions to increase the generic dispensing rate within Blue Cross Blue Shield of Michigan: member mailings, statewide major media advertising, incentives to six physician groups based on dollars saved with increased generic prescribing, and generic sampling. No significant impact on prescribing was found, and the incentive paid to prescribers was based on dollars saved by the health plan as a result of prescribing pattern changes, and was not based on the prescribing pattern change itself.(10) No study to date has evaluated the impact of a U.S. P4P program on clinic-level generic prescribing.

As a health care insurer for 45% of the U.S. population,(11) the public sector has a limited budget with which to procure services for its covered population, aiming to meet its members' health needs. Public sector payers [64% of whom are state and local entities(11)] are continually evaluating how best to use these scarce financial resources to improve population health. A 2005 Washington State Senate bill required a state purchasing project that coupled health care system performance with provider payments.(12) In response to this bill, Uniform Medical Plan (UMP), a self-insured preferred provider organization (PPO) health plan covering almost half of the State of Washington employees, retirees and dependents, developed a prescription benefit P4P program. The

primary focus of the UMP program was increasing generic prescribing in the two-year period between January 1, 2007 and December 31, 2008.

We sought to understand participating clinics' response to this program by way of their generic prescribing rates. Additionally, we sought to identify and understand program elements that may be important predictors of both engagement in and success with this type of program, as part of the payer-provider partnership. Answers to these questions are important to the payer community and the existing body of literature on P4P in developing future prescription P4P models. Additionally, and understanding of prescriber response to incentivized behavior may impact newer models of P4P, such as Accountable Care Organizations, Patient-Centered Medical Homes, and Coordinated Care Organizations.

Methods

We used a mixed methods approach to evaluate UMP's Generic Incentive Program (GIP) on generic prescribing. We tested the hypothesis that clinics participating in the GIP had higher annual generic dispensing rates (GDR) during the course of the program than did non-participating clinics. This constituted a longitudinal study of a natural experiment.

This study was reviewed and approved by the University of Washington Human Subjects Division (#35845). All statistical analyses were completed in Stata/IC 13.

Setting

UMP is one of health plans offered to Washington State public employees and administered by the Washington State Health Care Authority (HCA). At the time of UMP's P4P program, UMP covered the lives of nearly 180,000 state employees, retirees, their dependents, and other Washington State employer groups. UMP had approximately 24,000 unique providers whose patients filled at least one prescription using their UMP prescription coverage.

Twenty-four of UMP's PPO network clinics within Washington State were invited to participate in a one-year pilot GIP, which rewarded clinics quarterly based on their ability to meet certain prescribing targets for their patients covered by UMP. Prior to the GIP implementation in 2006, these clinics constituted "UMP Neighborhood," a clinic-based, HMO-like health plan offering for state employees, and all were invited to participate in the GIP. Clinics or clinic systems were part of UMP Neighborhood based on the following criteria: the clinic had at least 20 primary care providers, at least 300 UMP members receiving care, a member to provider ratio of at least 8, and were determined by an external entity to be the most cost-effective UMP network clinics among their peers. For those clinics that elected to participate, the GIP incentive included both a flat dollar award as well as a per-patient award, to account for diversity in clinic size. Those clinics that agreed to participate in the GIP signed a network contract addendum and were deemed participants for the purpose of this study. Those clinics that either actively or passively declined the GIP invitation were deemed non-participants for the purpose of this study.

Table 1 shows the performance targets for the 2007 GIP. Table 2 shows the incentive and payment scheme for 2007.The financial reward for clinics was based on their overall GDR, GDR for each of six drug classes, and overall compliance with UMP's preferred drug list (PDL). These rates were based on prescriptions written by providers within the participating clinics and subsequently filled at the patient's pharmacy and billed to UMP. The six drug classes evaluated were HMG-CoA reductase inhibitors ("statins"), antidepressants, proton pump inhibitors (PPI), nonsteroidal anti-inflammatory drugs (NSAID), antibiotics, and skeletal muscle relaxants.

Toward the end of the 2007 pilot year, UMP decided to continue the program for 2008. Table 3 shows the targets for the 2008 GIP performance measures. For the 2008 GIP, the class "antihypertensives" replaced the skeletal muscle relaxants, clinics could earn incentives for meeting individual drug class targets, rather than being required to meet targets for multiple classes before

earning any incentive, and individual drug classes had individual targets for each level of incentive. Both the 2007 and 2008 targets were set based on the UMP book of business performance in the previous year. The 2008 dollar amounts quarterly incentive payments were pre-specified by clinic for 2008, based again on their size, to eliminate the need for patient-level evaluation at quarter's end.

Data Sources

We used UMP administrative claims data from January 1, 2006 through December 31, 2009, i.e., including claims for one pre-program year and one post-program year, of continuously enrolled UMP members, aged 18-89. The administrative claims database extract consisted of prescription claims for these continuously enrolled members. Claims from January 1 to December 31, 2006, were used as the pre-program baseline.

A survey of GIP clinic administrators was conducted to elicit factors that may be important predictors of both engagement in and success with this type of program, including clinic size, prescribing information infrastructure, and incentive program characteristics. We sought to survey the GIP participation decision-makers of all GIP invited clinics in order to gain an understanding of motivation for participation as well as drivers of performance of the incentive program. The survey consisted of 11 questions addressing clinic-specific demographics and infrastructure, monetary incentive structure and program-specific questions. See Appendix for the full survey. We developed the survey by first considering clinic-specific demographics and internal support systems that might influence provider prescribing. After drafting the first set of questions, the survey was reviewed by HCA staff and health services researcher colleagues. This feedback was incorporated, and the survey was pilot-tested by one of the participating clinics prior to conducting survey interviews.

Exposure Classification

Clinic UMP GIP participation status was the exposure of interest in this study. Providers were identified by their Drug Enforcement Administration (DEA) number in each prescription claim and

matched to the list of DEA numbers by clinic provided to UMP by participating clinics. For non-participating clinics, providers' DEA numbers and clinic association were taken from the UMP provider database and were matched to prescription claims. For both participating and non-participating clinics, each provider's DEA number within a clinic was assigned a corresponding clinic number to cluster providers within clinics for the purpose of this study.

Outcome

The outcome for this analysis was the clinic's overall GDR. GDR was defined as the number of generic prescriptions dispensed divided by the total prescriptions dispensed. GDR was calculated on the clinic level as a result of prescriptions written by clinic providers and filled by the UMP members, using members' UMP prescription drug benefit. Although UMP evaluated PDL prescribing rates and GDRs for 6 individual drug classes, we focused on the overall GDR. PDLs change frequently and are not consistent between health plans. PDL prescribing adherence, therefore, is a more obscure target for providers. Focusing on the overall GDR, rather than individual drug classes, allowed for a much larger sample size by which to evaluate the impact of the program.

Statistical Analyses

Univariate analysis, using *t*-tests, was used to determine differences between groups at baseline. Bivariate analysis, using *t*-tests and ordinary least squares regression, was used to assess correlation between covariates and outcome to inform the statistical analysis.

Covariates hypothesized to be confounders of the relationship between GIP participation status and annual GDR that were included in the model were: age, gender, provider type (physician or midlevel provider), whether the clinic was of urban or rural geography, and a patient index of comorbid conditions. RxRisk, a risk score attributed to an individual based on weighted values of exposure to prescription drugs among various classes identifying chronic conditions, is a tool used

to predict future health care costs.(13) RxRisk weights were calculated for individual members to create a measure of member comorbidity, which was then converted into a binary high-low indicator using the median risk value (=662) among the entire continuously enrolled population as the threshold for high versus low risk. These covariates were designated at the clinic level for the analysis. Final clinic-level model covariates were: median member age, percent male, percent physician provider type (among physicians and midlevel providers within the clinic), and a time-varying RxRisk indicator to identify whether the average member comorbidity was high or low. Additional model covariates were calendar year, to account for secular changes, and a time-varying participant indicator by calendar year interaction to allow flexibility in annual GDR changes.

We assumed that observations over time for the same clinic were correlated and prescriptions within clinics were clustered by prescriber, so we fit a generalized estimating equation (GEE) model to account for this correlation and clustering in estimating the standard errors of our regression coefficients. GEE estimates the population effect of the intervention, which in this case is the average GDR change across all clinics associated with GIP participant status. The GEE model was fit using robust standard errors and a first-degree autoregressive working correlation matrix. Because robust standard errors might not perform as expected with a small number of clusters, the deletion diagnostic, jackknife, was used to confirm the point estimates.(14) Lastly, we checked for clinic outliers – by number or patients or by number of prescribers – and confirmed the reliability of the GEE estimates by removing the outlier and fitting the GEE model again.

Results

Twenty-five UMP network clinics within Washington State were eligible for this study. Seventeen (70.8%) agreed to participate, but one was an exclusively pediatric practice and thus none of their patients met the eligibility criteria for our study. The final participating clinic population included sixteen clinics (94.1% of all GIP participating groups). For the 2007 calendar year, 15 clinics earned

an award for at least one quarter. Three clinics declined to continue their participation in the second year. For the 2008 calendar year, all participating clinics earned an award for at least one quarter. Of the 87,337 UMP members continuously enrolled from 2006 to 2009 and aged 18-89, 20,996 (24.0%) received prescriptions from a GIP invited clinic. Of the 1,255,741 prescription claims for these members, 3,215 (0.25%) of them did not include a provider DEA, and therefore, could not be attributed to a provider.

Table 4 displays the characteristics of the participating and non-participating groups of clinics. Among the 16 GIP participating clinics that were evaluated were 16,732 unique members who filled a prescription during the study period and 1,977 providers who wrote those prescriptions. Likewise, among the 8 GIP non-participating clinics were 4,458 unique members and 421 unique providers. Members in the comparator groups were similar in gender distribution and in their identified comorbidity risk profile at baseline. Members from participating clinics were significantly older than patients among non-participants. Providers in comparator groups were different in provider type proportion, considering physician versus midlevel designation. And the GDR at study baseline was significantly higher for participating clinics (64.1% vs. 61.7%, $p<0.0001$). Figure 1 displays the raw annual GDR for each clinic, denoting GIP participation status.

Table 5 displays the point estimates and 95% confidence intervals for GDR annual difference between the groups, the interaction between group and year, and the underlying annual changes in GDR. The unadjusted GDR difference between participating and non-participating clinics, accounting for secular trending only, was 3.09 percentage points (95% CI: 0.34, 5.84; $p=0.02$) year over year. Adjusting for covariates, participating clinics' GDR, on average, was 2.61 percentage points (95% CI: -0.20,5.42) higher than that of non-participating clinics in 2006, indicating a systematic difference between the groups. On average, the estimated difference in GDR between participating groups for 2007 (-0.43; 95% CI: -1.46,0.61) and 2008 (-0.66; 95% CI: -2.22,0.90) were

not significant, indicating a non-statistically significant association between program participation and increase in GDR. The underlying (secular) annual changes in generic fill rates were 8.3%, 3.5%, and 2.8%, respectively, for years 2007, 2008 and 2009. Following execution of the GEE model, the results of the subsequent jackknife analysis confirmed the reliability of the point estimates calculated in the respective GEE models. When the largest clinic, considered an outlier, was removed from the analysis, the difference in GDR between groups increased to 2.86 percentage points year over year (95% CI: 0.11,5.60). The year by group interaction in this model was not significant. Secular changes were similar to the full model at 8.3%, 3.6%, and 2.8%, for 2006 to 2007, 2007 to 2008, and 2008 to 2009, respectively and all significant at $p<0001$.

Survey

Of the twenty-five clinics originally invited to participate in the UMP GIP and solicited for survey, 17 responded, all of which were located in urban areas of Washington.(15) Of the respondents, 2 did not participate in the program either year, 3 participated in the GIP in 2007 only, and the remaining 12 participated in both 2007 and 2008. Answers to survey questions reflect the clinic's information at the time of the UMP GIP program. The representatives responding on behalf of the two non-participating clinics/systems were not aware of the reason the clinics chose not to participate in the program and their respondent information only included clinic demographics. Therefore, program-specific information was obtained from 15 clinics only. Two of the three interviewees for clinics that discontinued their participation for 2008 were not aware of the reason they discontinued participation, while the remaining 2008 non-participant stated the clinic did not have enough staff support to meet program requirements.

Clinic size or book of business ranged from approximately 15,000 annual patient visits to hundreds of thousands of active charts. Of a clinic's book of business, UMP's members were a small proportion of that population, representing from <1% to 2% with a couple of clinics identifying up

to 10% of their population as UMP membership. One clinic indicated up to 60% of its Medicare patients were covered by UMP.

Table 6 displays the results of the clinic demographic and infrastructure questions of the 17 respondents. Eleven clinics (65%) had incentive plans or payment models other than UMP's program at the time of the UMP GIP that could affect prescribing practices. Sixteen clinics (94%) had technology to improve prescribing, such as electronic prescribing via electronic medical record or some other type of electronic drug resource. Fifteen clinics (88%) had some system-level policy that restricted distribution of samples and/or visits by pharmaceutical sales representatives. Clinic providers were educated on new drugs in a variety of ways across respondent clinics/systems. Methods included in-house newsletters, embedded or circulating pharmacists or associated hospital staff giving talks to provider groups, published in-house Pharmacy and Therapeutics Committee recommendations, specified pharmaceutical company educational programs, as well as those clinics with purely self-motivated provider education via peer-reviewed journals, electronic drug resources, continuing education materials and programs, and other literature available on clinic premises.

Table 7 displays the results of the UMP GIP program-specific question among the 15 respondents who participated in the GIP. Only three clinics/systems (20%) made their providers aware that their prescribing performance was being evaluated as part of UMP's program, all of which were made aware sometime prior to the onset of the program. Four clinics (27%) shared earned incentives from the UMP program directly with individual providers in their salaries, and five clinics (33%) rewarded providers non-financially, such as with new office equipment or education or verbal or written recognition among peers. When respondents were solicited for the dollar threshold (per patient or flat dollar) that would make prescribing habits a priority at their clinic/system, 8 respondents (53%) stated that either no specific dollar amount was motivating or that generic or low-cost prescribing was already a priority at the clinic/system, sometimes related specifically to the

patient's financial resources and/or formulary. For respondents who did think financial incentive would motivate clinic/system providers, suggested amounts ranged from 2% of reimbursement to $1,000-5,000 per provider per year to up to $100,000 annually to the clinic.

Three clinics (20%) discussed with individual providers the clinic-level and provider-level quarterly report cards sent by UMP. Eleven clinics (73%) believed the UMP program did not affect providers' prescribing behavior and only one clinic (7%) believed the UMP program did affect providers' prescribing behavior.

Discussion

In evaluating prescribers' response to a clinic-level P4P program, we sought to determine if program participation was associated with an annual increase in GDR that exceeded that of non-participants'. Further, we sought to describe clinic characteristics and program elements that may influence prescribing patterns and predict program engagement and success by clinics and their providers. Clinic-level P4P program participation and increased GDR were not related. Removing the outlier clinic did not meaningfully change the results. This non-significant finding can be a result of the incentive size or direction, participation cost, or lack of coordination, all of which are known considerations for structuring P4P programs.(16-18)

Against a backdrop of large secular increases in GDR for both participants and nonparticipants, the incentive may have been more motivating to behavior change had it been larger or directed to individual physicians rather than the group. A participation cost—financial or non-financial—may have been too large for participation: one clinic dropped out for the 2008 year for this very reason. Lastly, a lack of coordination among local health plans with competing incentive programs, causing clinic or provider multitasking, as well as a lack of coordination within individual clinics could have prevented attention toward performance in UMP's GIP. Additionally, it has been acknowledged that

reporting providers' performance to them in light of the available prescribing choices and compared to their peers, with no incentive, changes behavior.(19)

A strength of the study design is that new generics to market, changes in UMP benefit design and other tier shifts of drugs would non-differentially impact the difference in the change in generic prescribing between participants and non-participants. Furthermore, groups' baseline GDRs and all secular trends and are taken into account when estimating the difference in annual effect between groups with this GEE model. The published data on U.S. physician incentives for generic prescribing found a non-significant negative effect on generic prescribing. They closely matched a control group to their intervention group, which had similar demographic characteristics to our group of invited clinics.(10) Because UMP members were already incentivized via their out-of-pocket cost differential between brands and generics to purchase generics and because the program did not directly communicate with members, member behavior is not hypothesized to be an unmeasurable confounder of the result. Even where $4 generic prescriptions were offered, UMP members were still incentivized to use their UMP benefit since they paid a 10% coinsurance for generics, leading to a 40-cent out-of-pocket cost. However, we don't know if there were pharmacy-level incentives, financial or in conserving time or resources, that may have led to fewer or greater generic interchanges. These incentives may have been spillover effects related to pharmacy profit margins for generics versus brands or in dispensing simultaneous prescriptions for Medicare Part D patients, as Part D was still considered a new program at that time.

Participating clinics were more likely to respond to the survey regarding the program, with 16 of the 17 respondents being GIP participants for at least one of the two years the program ran. This differential response by participants may be a result of the same unobservable confounder that led the participants to self-select into the UMP GIP, leading to their borderline significant 2.61% higher annual GDRs year over year compared to non-participants..

Because the survey responses indicated that eleven participating clinics had concurrent incentive programs and only one clinic thought the program affected prescribing behavior, it is possible that the increase in generic prescribing among participating clinics during the time of the GIP was a combined result with of other local market forces. However, depending on the requirements of contemporaneous incentive programs, the UMP GIP non-participating clinics may have been participants in other programs, which would allow for a greater association of increased generic prescribing among UMP GIP participants to be a result of the UMP program. Finally, contemporaneous incentive programs may have contributed to the multitasking problem within clinics, preventing clinics from directly full effort to their performance in the UMP GIP. We are aware of one local P4P program that was in place near the same time as the UMP GIP. The concurrent P4P program included a measure of generic prescribing as well as plan preferred drug list prescribing compliance.(20)

Few clinic administrators communicated their participation in the UMP GIP to their individual providers or shared reward—monetarily or non-monetarily—with their providers, few discussed the prescribing report cards provided quarterly by UMP, and most respondents had technology to improve prescribing as well as education on new drugs that was not necessarily directly from the respective manufacturer. It is possible that the individual providers had some intrinsic motivation to increase their generic prescribing. This argument is supported by the fact that half of survey respondents indicated no dollar amount would motivate an increase in generic prescribing or that generic prescribing was already a priority. Communication and coordination among providers within a given clinic regarding program participation is essential to move beyond intrinsic motivation to achieve higher performance. A 2006 review of P4P program evaluations mentioned the lack of providers' awareness of their participation in these programs and discussed the importance of

communication with and among providers participating in these programs to increase their chance of success.(21)

The UMP GIP program framework can be described in light of the applications of the economic incentives described in Chapter 1. The strengths of the UMP GIP were: its few, specific and easily measured targets; its regular reporting, reward, and measure review for appropriateness; its inclusion of improvement targets rather than just threshold targets: its root in clinical evidence; the reward-only (no penalty) payment scheme, a low cost for participation; and a type of "risk adjustment" by not setting 100% GDR targets for drug classes without 100% generic availability. The potential limitations of the UMP GIP were: the lack of participation awareness among clinic providers, group versus individual rewards, and no public reporting of results. These possible limitations all reflect the level of intrinsic motivation of individual provider performance.

Limitations

Just as there were systematic differences between UMP GIP invited and uninvited clinics, there were possibly systematic differences between UMP GIP participating and non-participating clinics, which may be related to non-participants' active or passive decision to decline the GIP. The ideal experiment to evaluate this type of program would be to randomize clinics as participants or non-participants with perfectly measured clinic demographics and infrastructure components not otherwise captured in claims data. Because the clinics were not required to participate, those who accepted the invitation may have unobservable characteristics that would cause them to participate, some of which may be related to their proclivity to higher performance. Self-selection differentially affects the result and cannot be accounted for by covariates, because of these unobservable factors. Generalizability is limited by the self-selection of the clinics to the program.

Although administrative claims data are of high quality for classifying drug exposure and prescription drug claims are considered the gold standard,(22, 23) the limitations associated with the

use of claims data must be considered in light of these research findings. Prescription drug samples given to patients by prescribers, drugs given incident to a hospitalization, and prescriptions written but not filled are not recorded in prescription claims. Additionally, GDR is a proxy of written prescriptions, as some prescriptions written may not have been subsequently filled or billed to UMP, although the latter is less likely than the former. Although supply days is the most accurate representation of drug quantity exposure, compared to prescription count, supply days, in rare instances, may be incorrectly hand-calculated at the pharmacy.(23) Not all prescriptions were able to be attributed to providers when provider DEA was missing from the claims data, and therefore those 0.25% of prescription claims were not included in the generic prescribing rate. Although small, we cannot assume this missingness is non-differential.

Conclusion and Implications

Against a backdrop of large secular increases in GDR for both participants and nonparticipants, we did not find a statistically significant association between P4P program participation and increased GDR, it is possible that with a larger or more diverse incentives and with greater collaboration and coordination among providers and local health plans, this type of program would be a successful health plan cost containment strategy. Prescription P4P programs may influence prescribing, particularly among providers who are already intrinsically motivated to prescribe according to the goal(s) of the incentive plan and/or have strong clinic infrastructure and communication to support participation and success in these types of programs. Program developers should consider the application of P4P's incentive contracting history and economic principle application to optimize program features and maximize success. More research to understand and characterize clinic administrator and individual provider response to prescription P4P programs may better inform payers to create future successful programs in a payer-provider partnership to optimize prescription drug expenditures for health plans.

Table 1: 2007 UMP GIP Performance Targets

Performance Measure	Target
Overall GDR	65%
Overall PDL Compliance	90%
Individual Drug Class Performance Targets	
Statin	60%
Antidepressant	87%
PPI	50%
NSAID	97%
Antibiotics	95%
Skeletal Muscle Relaxants	100%

Table 2: 2007 UMP GIP Reward Scheme

Award Level	Gold	Silver	Bronze	Needs Improvement
Quarterly Incentive Payment	$1,250 + $1.25/ UMP Primary Care Patient	$875 + $1/ UMP Primary Care Patient	$500 + $0.75/ UMP Primary Care Patient	No Incentive Payment
Performance Targets				
1. GDR	≥ 65%	≥ 65%	≥ 65% or 3% improvement	Clinic performance is below minimum standards for incentive payment.
2. PDL Compliance	≥ 90%	≥ 90% or 1% improvement	≥ 90% or 1% improvement	
3. Target Drug Classes GDR	Meet or exceed targets for all 6 classes	Meet or exceed targets for 4 of 6 classes	Meet or exceed targets for 3 of 6 classes	

Table 3: 2008 UMP GIP Performance Targets

Performance Measure	Gold	Silver	Bronze	2008 vs. 2007 Performance Target
Overall Generic Dispensing Rate	77%	75%	73%	↑
Overall Preferred Drug List Compliance	95%	93%	91%	↑
Individual Drug Class Performance Targets				
Statin	85%	80%	75%	↑
Antidepressant	80%	77%	74%	↓
PPI	68%	65%	62%	↑
NSAID	97%	94%	91%	same
Antibiotics	97%	95%	93%	↑
Antihypertensives	80%	75%	70%	new class

Table 4: Group Characteristics

Variable	2007 Participants (16 clinics)	2007 Non-Participants (8 clinics)	*p*-value
No. Unique Patients	16,732	4,458	-
Female Gender	8,510 (60.0%)	2,570 (60.7%)	0.4018
Mean Age	56.2	55.6	0.0164
% High Risk	70.4%	71.3%	0.2713
No. Unique Providers	1,977	421	-
% Physician Provider Type	85.3%	93.3%	<0.0001
Generic Dispensing Rate	64.1%	61.7%	<0.0001

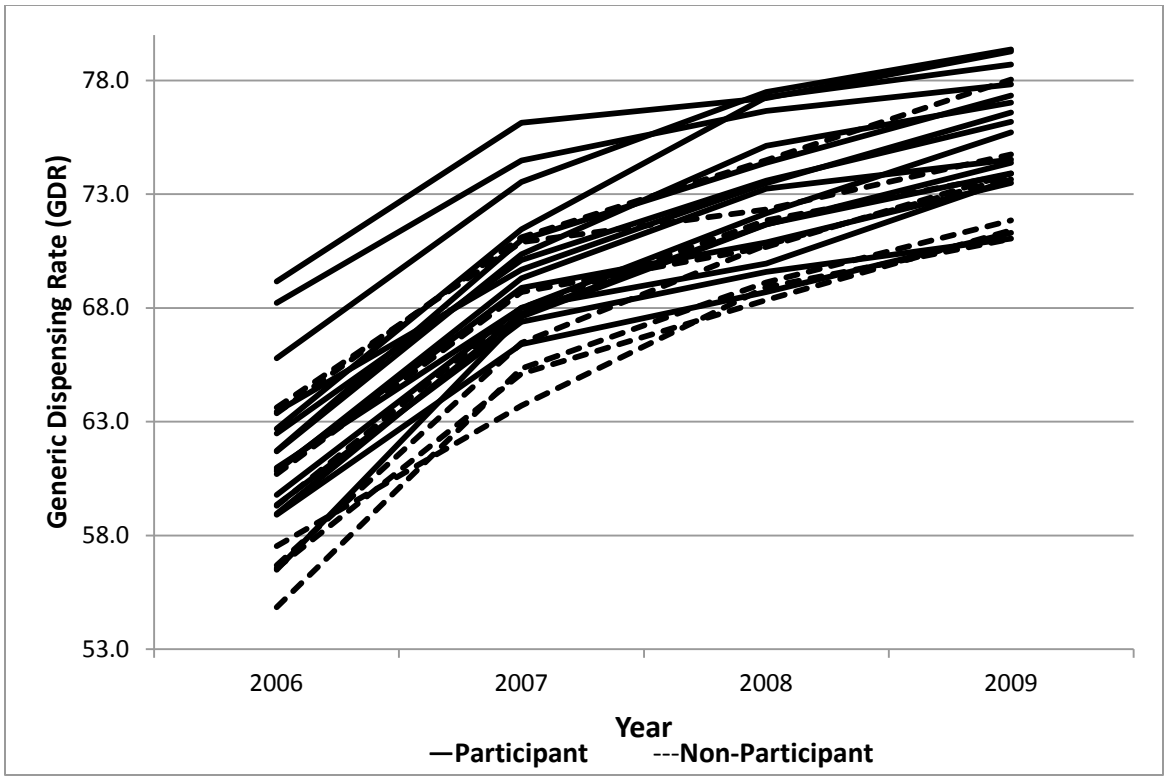

Figure 1: Raw Annual GDR of each Clinic

Table 5: GEE Model Point Estimates for GDR Annual Changes

Variable	Point Estimate (95% Confidence Interval)
Participant Indicator[1]	2.61 (-0.20,5.42)
Participant*Year 2007 Interaction	-0.43 (-1.46,0.61)
Participant*Year 2008 Interaction	-0.66 (-2.22,0.90)
Participant*Year 2009 Interaction	-0.96 (-2.59,0.67)
Year 2007 Indicator[2]	8.3 (7.51,9.16)
Year 2008 Indicator[2]	11.8 (10.61,13.07)
Year 2009 Indicator[2]	14.59 (13.19,15.98)

[1]p=0.07; [2]p<0.0001

Table 6: Clinic Demographics from Survey Data (N=17)

Survey Item	Yes	No
Clinic had other incentive plans or payment models that impacted prescribing practices.[1]	11	6
Clinic had technology to improve prescribing, such as electronic prescribing or other electronic drug resource.[2]	16	1
Clinic had a policy that restricted distribution of samples and/or visits by pharmaceutical representatives.[1]	15	2

[1]Of the 2 respondent non-participants, 1 responded "yes," and 1 responded "no."
[2]Both respondent non-participants responded "yes."

Table 7: UMP GIP Specific Responses from Survey Data (N=15)

Survey Item	Yes	No	Unknown
Clinic management notified providers that their prescribing performance was being evaluated as part of the program.	3	10	2
Clinic shared earned incentives with individual providers directly in their salaries.	4	8	3
Clinic provided nonfinancial incentives to providers based on their performance. (Examples: new office equipment, a plaque or other visible display of achievement, etc.)	6	5	4
Clinic discussed report cards sent by UMP at the end of each quarter with individual providers.	3	7	5
Do you think the UMP program affected providers' prescribing behavior?	1	11	3

Chapter 2 References

1. Hartman M, Martin AB, Lassman D, Catlin A, National Health Expenditure Accounts T. National health spending in 2013: growth slows, remains in step with the overall economy. Health Aff (Millwood). 2015;34(1):150-60.
2. 2014;Pages. Accessed at US Food and Drug Administration at http://www.fda.gov/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/ApprovalApplications/AbbreviatedNewDrugApplicationANDAGenerics/ on May 27 2015.
3. Drug Price Competition and Patent Term Restoration Act of 1984. 98th Congress ed; 1984.
4. Duckett SJ. Design of price incentives for adjunct policy goals in formula funding for hospitals and health services. BMC Health Serv Res. 2008;8:72.
5. Dunn JD. Pharmacy management approach: how do we align all the incentives? J Manag Care Pharm. 2007;13(2 Suppl B):S16-9.
6. Shrank WH, Choudhry NK, Liberman JN, Brennan TA. The use of generic drugs in prevention of chronic disease is far more cost-effective than thought, and may save money. Health Aff (Millwood). 2011;30(7):1351-7.
7. Gosden T, Torgerson DJ. The effect of fundholding on prescribing and referral costs: a review of the evidence. Health Policy. 1997;40(2):103-14.
8. Sturm H, Austvoll-Dahlgren A, Aaserud M, Oxman AD, Ramsay C, Vernby A, et al. Pharmaceutical policies: effects of financial incentives for prescribers. Cochrane Database Syst Rev. 2007(3):CD006731.
9. Dylst P, Vulto A, Simoens S. Demand-side policies to encourage the use of generic medicines: an overview. Expert Rev Pharmacoecon Outcomes Res. 2013;13(1):59-72.
10. O'Malley AJ, Frank RG, Kaddis A, Rothenberg BM, McNeil BJ. Impact of alternative interventions on changes in generic dispensing rates. Health Serv Res. 2006;41(5):1876-94.
11. Health Care Costs: A Primer. Key Information on Health Care Costs and Their Impact: The Henry J. Kaiser Family Foundation; 2012:34.
12. Engrossed Substitute Senate Bill 6090: An Act Relating to Fiscal Matters. 2005.
13. Fishman PA, Goodman MJ, Hornbrook MC, Meenan RT, Bachman DJ, O'Keeffe Rosetti MC. Risk adjustment using automated ambulatory pharmacy data: the RxRisk model. Med Care. 2003;41(1):84-99.
14. Dodge Y. The concise encyclopedia of statistics. 1st. ed. New York: Springer; 2008.
15. Rural Health Clinics and Urbanized Areas. Washington Department of Health; 2012.
16. Halladay JR, Stearns SC, Wroth T, Spragens L, Hofstetter S, Zimmerman S, et al. Cost to primary care practices of responding to payer requests for quality and performance data. Ann Fam Med. 2009;7(6):495-503.
17. Blecker E. The challenges in achieving successful P4P programs. Find Brief. 2014;42(2):1-3.
18. Chung S, Palaniappan LP, Trujillo LM, Rubin HR, Luft HS. Effect of physician-specific pay-for-performance incentives in a large group practice. Am J Manag Care. 2010;16(2):e35-42.
19. O'Neal BC, Couldry RJ, Wilkinson ST, Cannella CA, Williams CB, Scott LA, et al. Leveraging drug-utilization and external benchmarking data to drive change in prescribing behaviors. Am J Health Syst Pharm. 2012;69(21):1916-22.

20. Loffgren D. Overview of Selected Pay for Performance Programs. Inland Northwest Business Coalition on Health; 2005.
21. Christianson JB, Leatherman S, Sutherland K. Lessons from evaluations of purchaser pay-for-performance programs: a review of the evidence. Med Care Res Rev. 2008;65(6 Suppl):5S-35S.
22. Strom BL. Data validity issues in using claims data. Pharmacoepidemiol Drug Saf. 2001;10(5):389-92.
23. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005;58(4):323-37.

Chapter 3:

Program Evaluation of a Clinic-Level Pay-for-Performance Program

Abstract

Background

Provider incentive programs focused on generic prescribing encourage the use of higher value medications, aiming to realize the potential cost savings between brand-name and therapeutically equivalent generic drugs. Research has suggested that increased cost-sharing for prescription drugs is associated with lower adherence and higher utilization of non-drug health care services. One potential way to lower overall health care costs would be to optimize prescription drug expenditures by promoting an increased proportion of generic drug use. This study evaluated a two-year, clinic-level prescription drug pay-for-performance (P4P) program's impact on the prescription drug budget and patient outcomes.

Methods

We performed a prescription drug budget analysis and a retrospective cohort study evaluating Uniform Medical Plan's Generic Incentive Program (GIP). We calculated budget savings as a result of increased generic dispensing rates (GDRs) by GIP clinics compared to the costs of administering the program. Using administrative claims data in a multivariate analysis, we compared emergency department visits and hospitalizations for patients who were continuously enrolled members aged 18-89 years and sought care from participating and non-participating GIP clinics. For the multivariate analysis, we used generalized estimating equations, adjusted for baseline demographics and secular trends

Results

Two-year program budget impact constituted $250,795 in administrative costs amounted. Two-year program savings resulting from systematically increased GDRs among participating clinics during the two years the GIP was in place amounted to $1,998,633. After adjusting for available

confounders, we found no association between program participation and increased odds of emergency department visits or hospitalizations.

Discussion

Overall, the results of this study may indicate that methods to incentivize increased generic prescribing can lead to substantial savings without compromising patient health-related outcomes. A program that is cost-saving while maintaining or improving outcomes could be considered a cost-effective program and a productive investment.

Background

As U.S. health care expenditures continue to increase, payers respond by developing and implementing various programs and cost-containment strategies to mitigate their burden of the increase. As discussed in Chapter 2, provider incentive programs focused on generic prescribing do not create a barrier between a patient and medication as can other payer prescription cost-containment strategies. These programs encourage the use of higher value medications, realizing the difference in value between brand-name and generic drugs.(1) Compared to medical benefit incentive programs, which require not only provider involvement but patient follow-through with laboratory evaluations or diagnostic procedures, prescription benefit incentive programs only require patient follow-through by filling the prescriptions written for them. However, as with the introduction of any program that intends to impact prescriber behavior, there is always the concern that patient outcomes may be adversely impacted.

Several studies have suggested an association between increased prescription drug cost-sharing by patients, which leads to less prescription drug use, and an increase in non-prescription drug health care services among low-income or vulnerable populations.(2-5) A review of 136 articles by Goldman and colleagues concluded this inverse relationship exists for chronic diseases in particular when looking at cost-sharing changes for and adherence to medications.(6) More recent research suggests an association between increased prescription drug expenditures and decreased non-prescription drug health care utilization among privately insured members.(7) Additionally, other recent research demonstrated an association between less access to needed prescription drugs and increased emergency department costs among both publicly and privately insured individuals.(8) Conversely, recent research demonstrated increased adherence to generic statins compared to their brand-name counterparts, with an associated reduction in cardiovascular morbidity and mortality.(9) Furthermore, greater adherence has been found to be associated with generic prescribing overall for

chronic conditions compared to brand-name prescribing.(10) In light of this information, one potential way to lower overall health care costs would be to optimize prescription drug expenditures. Given the research demonstrating an association between greater generic drug use and higher adherence and the known positive effect of adherence, one important cost control tool is to promote a greater proportion of generic drug use.

We sought to determine whether Uniform Medical Plan's (UMP) Generic Incentive Program (GIP) was a cost-saving program, given the known difference in value between brand-name and generic drugs, and whether it at least maintained patient outcomes. To our knowledge, there is no published literature evaluating a prescription P4P program for its budget impact and impact on patient outcomes. Similarly to that of Chapter 2, the value in answers to these questions is for development of future prescription P4P models, including newer models, such as Accountable Care Organizations, Patient-Centered Medical Homes, and Coordinated Care Organizations.

Methods

We used a mixed methods approach to evaluate UMP's Generic Incentive Program (GIP) that included a budget impact analysis and statistical analyses of administrative claims data. This study was reviewed and approved by the University of Washington Human Subjects Division (#35845). All statistical analyses were completed in Stata/IC 13.

Setting

The UMP GIP, described in Chapter 2, is the setting for this program evaluation.

Data Source

We used UMP administrative claims data from January 1, 2006 through December 31, 2009, i.e., including claims for one pre-program year and one post-program year, of continuously enrolled UMP members, aged 18-89 years. The administrative claims database extract consisted of

prescription, emergency department visit, and hospitalization claims for these continuously enrolled members. Claims from January 1 to December 31, 2006, were used as the pre-program baseline.

Prescription Drug Budget Impact Analysis

We calculated the UMP prescription drug budget impact of the GIP for calendar years 2007 and 2008. The cost or savings of this program was calculated by comparing the savings from the GIP, resulting from increased generic prescribing, to the costs incurred by administrative work, clinical expertise time, reporting to clinics, and award payout. Because the perspective of analysis is the prescription drug budget, net financial impact did not include non-drug medical costs.

To control for changes in patient mix, drug class utilization, annual benefit changes, and underlying trends, the GDR difference (found in Chapter 2) between GIP participating and non-participating clinics was used to calculate the savings attributed to the program. That is, the program savings was calculated as the expenditure savings resulting from the increase in GDR among participating clinics minus the expenditure savings resulting from the increase in GDR among non-participating clinics. Drug expenditure plan savings was based on the difference between the average amount the plan paid for a 30-day supply of a brand-name drug and the average amount the plan paid for a 30-day supply of a generic drug. These calculations are displayed in Figure 1. Per member per month (PMPM) savings calculations were based on the average UMP enrollment during the UMP GIP. We used 2006 prescription claims from the 2006-2009 continuously enrolled population to calculate average brand-name and generic drug plan expenditures.

Health-Related Patient Outcomes Analyses

We tested the hypotheses that there was no association between GIP clinic participation status and increased patient odds of annual hospitalizations, total annual days in the hospital, and annual emergency department visits. These hypotheses were based upon the assumption of therapeutic interchangeability of generic versus brand-name drugs, indicating that a higher increase in annual

GDR by GIP participating clinics compared to non-participating clinics would have no adverse impact on health outcomes. This constituted a longitudinal study of a natural experiment.

Exposure Classification

Clinic UMP GIP participation status was the exposure of interest. Classification into participation groups was outlined in Chapter 2.

Outcomes

Annual patient odds of being admitted to the hospital, spending days in the hospital, and visiting the emergency department were the primary outcomes.

Statistical Analyses

Bivariate analysis, using *t*-tests and ordinary least squares regression, was used to assess the correlation between covariates and outcomes.

Covariates hypothesized to be confounders of the relationship between clinic participant status and primary outcomes were: age, gender, provider type (physician or midlevel provider), patient income quartile, patient's associated clinic, and an index of comorbid conditions. The same binary high-low, time-varying RxRisk indicator used in the Chapter 2 GDR GEE model was used in these models as the index of patient comorbidity. Patient median income was approximated using the crosswalk of 2009 US Census data(11) to zip code tabulation area to patient zip code. Each member was designated to an income quartile based on the range of the 2006-2009 continuously enrolled members in the claims data extract. Calendar year was included in the model to allow flexibility in annual (secular) changes in hospitalizations, hospital days, and emergency department visits.

We assumed that observations over time for the same patient were correlated, and so we fit generalized estimating equation (GEE) models to account for this correlation in estimating the standard errors of our regression coefficients. Because the outcomes in these analyses were binary,

we fit a binomial family GEE model with a logit link for each outcome. The models were fit using robust standard errors and an unstructured working correlation matrix.

Results

Demographic and baseline characteristics of the overall study population are described in Chapter 2. Additional study population data used in this study are shown in Table 1.

Comparator groups were similar in mean numbers of hospitalizations, days in the hospital and emergency department visits. Participating clinics' patients had significantly higher median income.

Prescription Drug Budget Impact Analysis

Program Costs

The two-year program costs included staff costs for program development, contracting, implementation, maintenance, and closure. Staff costs were valued at $50 per hour, according to UMP.(12) Vendor costs borne by UMP included quarterly report card generation and intermittent in-person meetings with clinics. Most of the costs to implement and maintain the program were considered "overhead costs" and were fixed and/or sunk in the first year.(13, 14) Staff costs totaled $20,050. Report card generation and meeting costs totaled $63,690. Reward payout totaled $166,785. Table 2 displays the range of quarterly reward payments made to clinics, as well as the total quarterly payout by UMP. The total program costs were thus $250,795. Table 3 displays cost categories and totals.

Program Savings

The 2006 average plan expenditure for a 30-day supply of a brand-name drug was $73.63. The 2006 average plan expenditure for a 30-day supply of a generic drug was $16.63. The average plan savings, therefore, for purchase of a generic versus a brand-name drug in 2006 was $57.00. The plan savings per 30-day supply increased year by year. Because there was no significant increase in GDR by

participants beyond their systematically greater GDRs at baseline during the years of the GIP, there was no program savings to calculate.

Budget Impact

Total budget impact to UMP prescription drug expenditures comparing GIP program costs to GDR savings over 2007-8 was a net cost of $250,795.

Health-Related Patient Outcomes

All 20,996 patients who received care from a GIP-invited clinic were included in the analyses. Upon evaluation in bivariate analysis, patient income quartile was found to be correlated with the primary outcomes and was, therefore, included in the model. Additionally, upon bivariate analysis, supply days quartile (categorical designation of total annual supply days of all patient prescriptions) was found to be correlated with each outcome of interest. Final model covariates were: age, gender, a binary indicator of the patient's provider type (physician or midlevel provider), patient's associated clinic, calendar year, supply days quartile, income quartile, the patient's time-varying RxRisk indicator, and a participant by year interaction term.

Table 4 displays the odds ratios for the underlying, annual trends. Adjusting for covariates, on average across all patients, the odds ratio of incurring an annual hospitalization or days in the hospital in 2007 compared to 2006 was 1.25 (95% CI: 1.08,1.45; p=0.004). Adjusting for covariates, on average across all patients, the odds of incurring an annual ER visit was 1.13 (95% CI: 0.98,1.31; p=0.09) in 2007 compared to 2006. Adjusting for covariates, across all patients, there was no increased odds of incurring a hospitalization, days in the hospital or an ER visit in 2008 or 2009 compared to 2006. Table 5 displays the coefficients for patient outcomes and participant by year interaction terms. Table 6 displays the odds ratios (ORs) of each outcome comparing participants to non-participants. We tested linear combinations of the systematic difference between groups and each participant by year interaction to calculate odds ratios for patient outcomes by group for each

year. Those values are also included in Table 6. Adjusting for covariates, on average, patients of GIP participants were no more likely than those of non-participants to incur annual hospitalizations or days in the hospital year over year. Adjusting for covariates, on average, patients of GIP participants trended toward lower odds of emergency department visits compared to patients of non-participants in 2006 with an OR of 0.77 (95% CI 0.59,1.01; p=0.06) . The odds of incurring a hospitalization or days in the hospital in 2007 was 31% lower for patients of participating clinics compared to those of non-participating clinics (95% CI: 0.53,0.89; p=0.004) and 30% lower for incurring an ER visit (95% CI: 0.53,0.91; p=0.008). The odds of incurring an ER visit in 2008 was 32% lower for patients of participating clinics compared to those of non-participating clinics (95% CI: 0.52,0.89; p=0.05). The odds of incurring an ER visit in 2009 were 30% lower for patients of participating clinics compared to those of non-participating clinics (95% CI: 0.53,0.92; p=0.01). There was no difference in odds for patients in comparator groups of incurring hospitalizations or days in the hospital in 2008 or 2009 compared to 2006.

Discussion

We evaluated the UMP GIP for its impact on the prescription drug budget and health-related patient outcomes. As discussed in Chapter 2, UMP GIP clinic-level P4P program participation and increased GDR were not related. Therefore, the budget impact to the State of Washington was a cost of over $250,000, which is an average PMPM cost to the State of $0.16. Finally, participation in UMP's GIP and the associated increase in GDR was not associated with an increase in annual hospitalizations, hospital days, or emergency department visits. A secular trend indicated increased odds of hospitalizations, days in the hospital and emergency department visits across all patients, and in a few cases, trends of these outcomes were lower among patients of participants than those of non-participants. A strength of the design of the outcomes portion of this study is that groups'

baseline and all secular trends are taken into account when estimating the difference in annual effect between groups with the GEE models.

The 2.61% systematic difference in GDR (discussed in Chapter 2) by participants compared to non-participants year over year amounted to a two-year savings of $1,998,633, or an average two-year PMPM savings of $0.46, indicating the budget impact increased generic prescribing has when characteristics unobserved in our study impact clinic-level generic prescribing. Using the same savings calculations as for plan savings, replacing plan expenditures with member expenditures, members who were patients of participating clinics overall saved $361,554 in 2007 and $399,692 in 2008. Members with prescription drug benefit coverage save with generic prescriptions in two ways: members who purchase generic prescriptions save at the point of service with lower out-of-pocket costs, and all members save money by way of premiums when plan savings is taken into account actuarially.

We did not include use of outpatient or office visits in our study of outcomes. These are considered essential and appropriate resource utilization in health care, rather than the non-essential use of the hospital and emergency room. In their Data Brief "Recommended Core Measures for Evaluating the Patient-Centered Medical Home," Rosenthal and colleagues do not include outpatient visits as one of the core measures of cost and utilization.(15) Outpatient visit utilization may be confounded by unobservable factors that are not an indication of poor outcomes, such as physician preference for visit frequency, visits to change, withdraw or refill prescription drugs, and visits requested or required by providers for laboratory orders.

Published studies of incentivized prescribing—of any kind—within the US are few, focusing on incentives paid directly to the patient in the form of vouchers (16), clinic-based pharmacist impact on generic prescribing, both educational and prescriptive (17, 18), and investigation of program impact on poor performers.(19) As noted in Chapter 2, the one research publication focused on the

evaluation of a US health plan's P4P program incented providers with shared savings, rather than payment for reaching preset targets.(20) Our study adds to the existing research on prescribing pay-for-performance evaluation by investigating response to set targets, a calculation of explicit budget impact, and a review of associated health-related patient outcomes.

Limitations

Limitations associated with the self-selection of clinics into participation groups in this natural experiment evaluation and uses of prescription claims data for research are described in Chapter 2. Additionally, while administrative claims data may be of high quality for evaluating policy changes, given the sample size of providers and patients, there are patient-level confounders and indicators of and impacts on health outcomes, such as smoking status and other indicators of health, such as laboratory values and provider progress notes, that are not included.(21)

In the budget impact analysis, non-drug health care utilization was not taken into account, based on the perspective of the analysis and the assumption that a program focused on generic prescribing would not change patient outcomes. However, non-drug health care utilization may be impacted as a result of patient drug regimen changes. We also did not take drug manufacturer rebates received by the plan into account which could decrease the plan paid difference between brand-name and generic drugs. Health-related patient outcomes are affected by many factors that may confound trending, not all of which will be measurable, such as patient hygiene, family history, diet, and exercise. Additionally, some outcomes of chronic diseases have long induction times that will not be captured in the evaluated time frame.

Conclusion

The non-significant increase in annual GDRs by UMP GIP participants resulted in a cost to the State of Washington of $250,795 for the GIP. However, the systematically increased GDRs by

participating clinics amounted to a $2 million savings during the two years of the GIP. Increased generic dispensing also resulted in savings for members at the point of service.

Overall, the results of this study may indicate that methods to incentivize increased generic prescribing can lead to substantial savings without compromising patient health-related outcomes. A program that is cost-saving while maintaining or improving outcomes could be considered a cost-effective program and a productive investment.

GIP Program Savings =	Program Savings–Program Costs
Program Savings =	Average Savings per 30-Day Supply of Generic Drug x Total Supply Days of Drug in Year x %GDR Excess by GIP Participants
Average Savings per 30-Day Supply of Generic Drug =	Average UMP Paid per Brand per 30-Day Supply– Average UMP Paid per Generic per 30-Day Supply
Average UMP Paid per 30-Day Supply of Brand =	Total Paid for Brands/Total Brand Days x 30
Average UMP Paid per 30-Day Supply of Generic =	Total Paid for Generics/Total Generic Days x 30

Figure 1: Program Savings Calculation

Table 1: Group Baseline Characteristics

Variable	2007 Participants (16 clinics)	2007 Non-Participants (8 clinics)	*p-value*
Patient Income Quartile (Median Income)	1.75 ($65,412.42)	1.63 ($64,770.87)	<0.0001 (0.03)
Mean No. Hospitalizations	0.085	0.085	0.9761
Mean No. Hospital Days	0.355	0.324	0.5390
Mean No. Emergency Department Visits	0.124	0.122	0.7808

Table 2: GIP Reward Payout

Year	Quarter	Range Among Clinics	Total Payout
2007	1	$0-1,854	$2,957
	2	$0-2,129	$13,989
	3	$0-3,138	$15,048
	4	$0-3,715	$26,201
2008	1	$400-4,400	$24,395
	2	$560-4,680	$26,585
	3	$600-4,840	$27,615
	4	$740-4,590	$29,915

Table 3: GIP Costs

Cost Category	Cost
UMP Staff Costs	$20,050
Reporting and Meetings	$63,960
2007 Reward Payout	$58,195
2008 Reward Payout	$108,590
Total Cost	**$250,795**

Table 4: Secular Trends for Annual Health-Related Patient Outcomes

Primary Outcome	2007 vs 2006 Secular Trend Odds Ratio (95% Confidence Interval)	2008 vs 2006 Secular Trend Odds Ratio (95% Confidence Interval)	2009 vs 2006 Secular Trend Odds Ratio (95% Confidence Interval)
Hospitalizations	1.25[1] (1.08,1.45)	1.10 (0.94,1.28)	1.05 (0.90,1.23)
Hospital Days	1.25[1] (1.08,1.45)	1.10 (0.94,1.28)	1.05 (0.90,1.23)
Emergency Department Visits	1.13[2] (0.98,1.31)	1.12 (0.97, 1.29)	1.05 (0.90,1.22)

[1]p=0.004; [2]p=0.09

Table 5: Coefficients for Annual Health-Related Patient Outcomes: GIP Participants vs Non-Participants

Primary Outcome	Coefficient (95% Confidence Interval)	Participant*2007 Interaction (95% Confidence Interval)	Participant*2008 Interaction (95% Confidence Interval)	Participant*2009 Interaction (95% Confidence Interval)
Hospitalizations	-0.11 (-0.37,0.15)	-0.26[1] (-0.44,-0.09)	-0.10 (-0.28,0.07)	-0.05 (-0.23,0.12)
Hospital Days	-0.11 (-0.37,0.15)	-0.26[1] (-0.44,-0.09)	-0.10 (-0.28,0.07)	-0.05 (-0.23,0.12)
Emergency Department Visits	-0.26[2] (-0.54,0.01)	-0.10 (-0.26,0.06)	-0.12 (-0.29,0.05)	-0.09 (-0.26,0.08)

[1]p=0.03; [2]p=0.06

Table 6: Odds Ratios for Annual Health-Related Patient Outcomes: GIP Participants vs Non-Participants

Primary Outcome	Odds Ratio (95% Confidence Interval)	2007 Odds Ratio (95% Confidence Interval)	2008 Odds Ratio (95% Confidence Interval)	2009 Odds Ratio (95% Confidence Interval)
Hospitalizations	0.90 (0.69,1.16)	0.69[1] (0.53,0.89)	0.81 (0.62,1.05)	0.85 (0.66,1.10)
Hospital Days	0.90 (0.69,1.16)	0.69[1] (0.53,0.89)	0.81 (0.62,1.05)	0.85 (0.66,1.10)
Emergency Department Visits	0.77[2] (0.59,1.01)	0.70[3] (0.53,0.91)	0.68[4] (0.52,0.89)	0.70[5] (0.53,0.92)

[1]p=0.004; [2]p=0.06; [3]p=0.008; [4]p=0.005; [5]p=0.01

Chapter 3 References

1. Shrank WH, Choudhry NK, Liberman JN, Brennan TA. The use of generic drugs in prevention of chronic disease is far more cost-effective than thought, and may save money. Health Aff (Millwood). 2011;30(7):1351-7.
2. Soumerai SB. Benefits and risks of increasing restrictions on access to costly drugs in Medicaid. Health Aff (Millwood). 2004;23(1):135-46.
3. Horn SD. Unintended consequences of drug formularies. Am J Health Syst Pharm. 1996;53(18):2204-6.
4. Soumerai SB, Ross-Degnan D, Avorn J, McLaughlin T, Choodnovskiy I. Effects of Medicaid drug-payment limits on admission to hospitals and nursing homes. N Engl J Med. 1991;325(15):1072-7.
5. Heisler M, Langa KM, Eby EL, Fendrick AM, Kabeto MU, Piette JD. The health effects of restricting prescription medication use because of cost. Med Care. 2004;42(7):626-34.
6. Goldman DP, Joyce GF, Zheng Y. Prescription drug cost sharing: associations with medication and medical utilization and spending and health. JAMA. 2007;298(1):61-9.
7. Hazlet T, James E, Blough D, Sullivan D. Longitudinal Health Outcomes in a Preferred Provider Organization. 22nd International Conference on Pharmacoepidemiology & Therapeutic Risk Management. Lisbon, Portugal.
8. Watanabe JH, Ney JP. Association of increased emergency rooms costs for patients without access to necessary medications. Res Social Adm Pharm. 2014.
9. Gagne JJ, Choudhry NK, Kesselheim AS, Polinski JM, Hutchins D, Matlin OS, et al. Comparative effectiveness of generic and brand-name statins on patient outcomes: a cohort study. Ann Intern Med. 2014;161(6):400-7.
10. Shrank WH, Hoang T, Ettner SL, Glassman PA, Nair K, DeLapp D, et al. The implications of choice: prescribing generic or preferred pharmaceuticals improves medication adherence for chronic conditions. Arch Intern Med. 2006;166(3):332-7.
11. 2005-2009 American Community Survey Data.
12. Sullivan D. Personal Communication. Seattle; 2008.
13. Gold MR. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996.
14. Drummond M. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford ; New York: Oxford University Press; 2005.
15. Rosenthal MB, Abrams MK, Bitton A, Collaborative P-CMHE. Recommended Core Measures for Evaluating the Patient-Centered Medical Home: Cost, Utilization, and Clinical Quality. In: Fund TC, ed. Data Brief ed; 2012.
16. Bhargava V, Greg ME, Shields MC. Addition of generic medication vouchers to a pharmacist academic detailing program: effects on the generic dispensing ratio in a physician-hospital organization. J Manag Care Pharm. 2010;16(6):384-92.
17. Devine EB, Hoang S, Fisk AW, Wilson-Norton JL, Lawless NM, Louie C. Strategies to optimize medication use in the physician group practice: the role of the clinical pharmacist. J Am Pharm Assoc (2003). 2009;49(2):181-91.
18. Koenigsfeld CF, Horning KK, Logemann CD, Schmidt GA. Medication therapy management in the primary care setting: a pharmacist-based pay-for-performance project. J Pharm Pract. 2012;25(1):89-95.

19. Dowd B, Feldman R, Nersesian W. Setting pay for performance targets: do poor performers give up? Health Econ. 2013;22(2):168-79.
20. O'Malley AJ, Frank RG, Kaddis A, Rothenberg BM, McNeil BJ. Impact of alternative interventions on changes in generic dispensing rates. Health Serv Res. 2006;41(5):1876-94.
21. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005;58(4):323-37.

Summary of Findings and Implications

Summary of Findings

The fundamental economic paradigm behind incentive contracting is the principal-agent relationship and its inherent information asymmetry. The goals of a pay-for-performance (P4P) program are to improve and to reduce variation in health care and outcomes. The Institute of Medicine (IOM) has been a catalyst for movement in the quality of health care, with its work leading to payers focusing on quality of their contracted providers through P4P programs.(1) Number and type of measures, incentive size and distribution, risk adjustment, and reporting are all aspects of program framework to be considered during development. Ultimately, the cost of the performance measure(s) should be less than the benefit both from the payer and provider's/clinic's perspective.

Providers are "double agents" in P4P programs, as they are expected to act in the best interest of both the payer and their patients. Within this double agency exists the persistent tension between population-level health care and patient-centered care, of which providers are expected to deliver both efficiently. Eijkenaar and colleagues' international systematic review of P4P systematic reviews found that the more effective P4P programs were developed collaboratively between payers and providers.(2) This tactic could serve to reduce the information asymmetry between payers and providers and to encourage better alignment between the payer and provider(s) performance functions.

In assessing P4P program effectiveness, the most recent systematic review concluded that we are still awaiting conclusive findings on the long-term impact to payer budgets and health-related patient outcomes.(2)

In our study of the Uniform Medical Plan (UMP) 2007-2008 Generic Incentive Program (GIP), we found that clinic-level P4P program participation and increased GDR were not related, though there was a 2.61% systematic difference in GDRs between participants and non-participants. The

total budget impact of the GIP was the total cost of the program: just over $250,000. However, the systematically increased GDRs by participating clinics amounted to a $2 million savings during the two years of the GIP. This savings was accrued without compromising patient health-related outcomes. Increased generic dispensing also resulted in savings for members at the point of service.

Implications

Multiple players, multiple stakeholders, and a multi-faceted, fragmented health care system make the U.S. health care a complex environment for P4P programs. Today, payer to health system payment for quality is seen in platforms such as patient-centered medical homes, accountable care organizations, and coordinated care organizations, as well as in hospital contracting that audits for the presence of never-events and 30-day readmissions. For the best chance of successful program development and effectiveness, program developers should refer to the IOM's dimensions influencing health care quality, Town and colleagues' principal-agent model of health care applications, and Forrest and colleagues' provider preference lists during conceptual program development.(1, 3, 4) These concepts reflect the basic principles behind incentive contracts, provide a useful framework for P4P program, and may lead to more successful efforts to implement P4P.

In addition to targets being designed as accurate representations of true morbidity, mortality and/or disability, targets should be set so that poor performers are engaged to participate and so that higher performers do not quickly reach a plateau.(5, 6) Measures should be continually evaluated for their appropriateness in invoking performance that improves overall care. Program advisors to the Quality and Outcomes Framework P4P program in the United Kingdom are aware of the need to continually evaluate metrics and their targets to ensure their appropriateness and meaningfulness to practice, including whether or not metrics should be removed altogether.(6-9) This awareness led to the 2012 removal of all organizational indicators that were not able to invoke continued improvement in performance.(6)

More effective programs have been found to be directed to smaller groups and even individual physicians.(2) For larger groups, perhaps a hybrid of incentive pay given to the group as well to as individual contributors could be considered. Individual incentive pay could be distributed by the health plan or by the practice as a contractual requirement of participation in the P4P program. An individual contributor aspect to incentive pay would also help to reduce the free rider problem in group practices. While significant financial incentives have been found to improve performance in P4P programs,(2) payers must also consider that increased payment potential may result in additional unintended or negative consequences, as providers "teach to the test" or focus on the high(er) pay associated with given performance targets.(6) Regular communication between payers and providers, including payer distribution of high quality feedback reports, could promote greater success. Because of the importance of provider reputation, feedback to providers elucidating their performance compared to peers, best practices, or high-performing institutions may encourage continued provider engagement toward meeting performance targets.(10)

Although we did not find a large or statistically significant association between P4P program participation and increased GDR, it is possible that with larger and more diverse or different incentives and greater collaboration, communication and coordination among providers and local health plans, this type of program would be a successful health plan cost containment strategy. Prescription P4P programs may influence prescribing, particularly among providers who are already intrinsically motivated to prescribe according to the goal(s) of the incentive plan and/or have strong clinic infrastructure and communication to support participation and success in these types of programs. More research to understand and characterize clinic administrator and individual provider response to prescription P4P programs may better inform payers to create future successful programs in a payer-provider partnership to optimize prescription drug expenditures for health plans. Overall, the results of this study suggest that stronger incentives to increase generic

prescribing could produce substantial cost-savings without compromising patient health-related outcomes. A program that is cost-saving while maintaining or improving outcomes could be considered a cost-effective program and a productive investment.

Summary of Findings and Implications References

1. Berwick DM. A user's manual for the IOM's 'Quality Chasm' report. Health Aff (Millwood). 2002;21(3):80-90.
2. Eijkenaar F, Emmert M, Scheppach M, Schoffski O. Effects of pay for performance in health care: a systematic review of systematic reviews. Health Policy. 2013;110(2-3):115-30.
3. Town R, Wholey DR, Kralewski J, Dowd B. Assessing the influence of incentives on physicians and medical groups. Med Care Res Rev. 2004;61(3 Suppl):80S-118S.
4. Forrest CB, Villagra VV, Pope JE. Managing the metric vs managing the patient: the physician's view of pay for performance. Am J Manag Care. 2006;12(2):83-5.
5. Dowd B, Feldman R, Nersesian W. Setting pay for performance targets: do poor performers give up? Health Econ. 2013;22(2):168-79.
6. Harrison MJ, Dusheiko M, Sutton M, Gravelle H, Doran T, Roland M. Effect of a national primary care pay for performance scheme on emergency hospital admissions for ambulatory care sensitive conditions: controlled longitudinal study. BMJ. 2014;349:g6423.
7. Lester H, Schmittdiel J, Selby J, Fireman B, Campbell S, Lee J, et al. The impact of removing financial incentives from clinical quality indicators: longitudinal analysis of four Kaiser Permanente indicators. BMJ. 2010;340:c1898.
8. Doran T, Kontopantelis E, Reeves D, Sutton M, Ryan AM. Setting performance targets in pay for performance programmes: what can we learn from QOF? BMJ. 2014;348:g1595.
9. Caley M, Burn S, Marshall T, Rouse A. Increasing the QOF upper payment threshold in general practices in England: impact of implementing government proposals. Br J Gen Pract. 2014;64(618):e54-9.
10. O'Neal BC, Couldry RJ, Wilkinson ST, Cannella CA, Williams CB, Scott LA, et al. Leveraging drug-utilization and external benchmarking data to drive change in prescribing behaviors. Am J Health Syst Pharm. 2012;69(21):1916-22.

Uniform Medical Plan Generic Incentive Program
Clinic Interview Questions Script

1. (*Only for clinics that declined participation for the 2007 calendar year*) Why did [clinic or group name] decline participation in the Uniform Medical Plan Generic Incentive Program?
    a. Benefit did not outweigh effort involved to participate
    b. Not enough staff support for clinic requirements such as DEA lists
    c. Incentive not large enough
    d. Lack of interest
    e. Didn't like health plan telling us how to prescribe drugs
    f. Other:

2. (*Only for clinics that participated in the 2007 calendar year but declined to continue participation in the 2008 calendar year*) Why did [clinic or group name] discontinue participation in the Uniform Medical Plan Generic Incentive Program for the 2008 calendar year?
    a. Benefit did not outweigh effort involved to continue
    b. Not enough staff support for clinic requirements such as DEA lists
    c. Incentive not large enough
    d. Lack of interest
    e. Didn't like health plan telling us how to prescribe drugs
    f. Other:

"Would you mind if I ask you a few more questions?" (Purpose: those clinics that did not participate or dropped out may not be willing to spend more time answering questions.)

3. Within [clinic or group name], do you have other incentive plans or payment models that impact prescribing practices?
    a. Yes
    b. No

4. Does [clinic or group name] have technology to help improve prescribing such as electronic prescribing, perhaps including prompts with medication alternatives, Epocrates or something like it, RxHub/SureScripts or something else with a patient-level formulary, etc.?
    a. Yes
    b. No

5. Does [clinic or group name] have a policy that restricts distribution of samples and/or visits by pharmaceutical representatives?
    a. Yes
    b. No

6. How are your prescribers educated on new drugs, including comparison to existing drugs? Another way to ask this: do you have a "counter-detailing" type of education in the clinic?

7. What dollar threshold (per patient or flat dollar) would make prescribing habits a priority in [clinic or group name]?

8. Demographics questions
    a. How big is your patient family/book of business (number of patients)?
    b. What percent of your patient family/book of business is UMP (number of patients)?

STOP HERE for clinics that never participated in the UMP Generic Incentive Program.

9. The Uniform Medical Plan Generic Incentive Program ran from January 1, 2007, to December 31, 2008.
    a. Did you notify your providers that their prescribing performance was being evaluated as part of this program?
        i. Yes (if yes, go to 1b)
        ii. No
    b. If yes, in what month and year were the providers notified they were participating in the Uniform Medical Plan Generic Incentive Program? (*Please note that rewards were calculated at the end of each calendar quarter, i.e., end of March, June, September, and December each year.*)
        i. [Month] of 200___
        ii. Sometime prior to January 1, 2007

10. Questions regarding incentives provided to individual providers:
    a. Did your clinic share any earned incentives with individual providers directly in their salaries?
        i. Yes
        ii. No; go to (d)
    b. Did you reward providers based on their individual performance or overall clinic performance?
        i. Individual performance. Formula:
        ii. Overall clinic performance. Formula:
    c. Did you make providers aware of the financial incentive ahead of time?
        i. Yes
        ii. No
    d. Did you provide nonfinancial incentives to providers based on their performance? For example, did providers receive gift cards, new office equipment, a plaque or other visible display of achievement, etc.?
        i. No
        ii. Yes. Incentive:

e. Did you discuss the report cards UMP sent to your office at the end of each quarter with the individual providers?
    i. Yes
    ii. No

        If no,

        1. Do you do any other type of peer review among prescribers?
        2. How often?
        3. Who is in a peer group? (ex. Comparison of surgeons to family practice docs or surgeons to surgeons?)

11. Do you think the UMP program affected providers' prescribing behavior?
    a. Yes
    b. No