3-16-2015

Incentivizing Cost-Effective Reductions in Hospital Readmission Rates

James Cox
Georgia State University

Vjollca Sadiraj
Georgia State University

Kurt Schnier
University of California

John Sweeney
Emory University School of Medicine

Incentivizing Cost-Effective Reductions in Hospital Readmission Rates

James C. Cox[a], Vjollca Sadiraj[a], Kurt E. Schnier[b] and John F. Sweeney[c]

March 16, 2015

Abstract:
The recent regulatory changes enacted by the Centers for Medicare and Medicaid Services (CMS) have identified hospital readmission rates as a critical healthcare quality metric. This research focuses on the utilization of pay-for-performance (P4P) mechanisms to cost effectively reduce hospital readmission rates and meet the regulatory standards set by CMS. Using the experimental economics laboratory we find that both of the P4P mechanisms researched, bonus and bundled payments, cost-effectively meet the performance criteria set forth by CMS. The bundled payment mechanism generates the largest reduction in patient length of stay (LOS) without altering the probability of readmission. Combined these results indicate that utilizing P4P mechanisms incentivizes cost effective reductions in hospital readmission rates.

Keywords: Pay-for-Performance (P4P), Healthcare, Experiment
JEL Codes: C91, D81, I10

Acknowledgement: The research was supported by a grant from the National Institutes of Health, National Institute on Aging (grant number 1RC4AG039071-01). The authors would like to thank two anonymous reviewers and the Editor for useful and insightful comments.

a. Experimental Economics Center (E*x*CEN) and Department of Economics, Andrew Young School of Policy Studies, Georgia State University. Cox: 14 Marietta St. NW, Room 436; Atlanta GA 30303; jccox@gsu.edu; 404-413-0200. Sadiraj: 14 Marietta St., NW, Room 453; Atlanta GA 30303; vsadiraj@gsu.edu; 404-413-0193.
b. Corresponding Author: School of Social Sciences, Humanities and Arts, University of California, Merced; 5200 North Lake Road; Merced CA 95343; kschnier@ucmerced.edu; 209-205-6461
c. Department of Surgery, Emory University School of Medicine; 201 Dowman Drive; Atlanta GA 30322; jfsween@emory.edu; 404-727-5800.

Forthcoming in *Journal of Economic Behavior and Organization*

1. Introduction

Recently the Institute of Medicine estimated the amount of wasted, excess cost of healthcare to be approximately $765 billion in 2009 (Institute of Medicine 2012). The growth in our health expenditures relative to GDP makes the United States a clear global outlier (Chandra and Skinner 2012), but the care being provided merely places us in the middle of the pack (Fuchs and Millstein 2011). The United States is faced with the challenge of not only decreasing the cost of providing care to its population, but also increasing the quality that is provided. The Centers for Medicare and Medicaid Services (CMS) has recently identified hospital readmission rates as a critical healthcare quality metric within the United States and taken regulatory steps to incentivize hospitals to increase their performance. The incentive mechanism utilized by CMS penalizes hospitals that do not meet their performance targets (i.e., readmission rates that exceed expected levels). Recently, the penalties used by CMS amount to a 1% reduction in reimbursement rates for hospitals that have “too many” patients being readmitted within 30 days of hospitalization. The outcome was a total penalty of $280 million in 2013 and the percentage is expected to increase to 3% in 2015 (CMS 2015, Health Affairs 2013).[1]

The penalties enacted by CMS come at a considerable cost to hospitals, as any additional preventive care must be covered by the current prospective payment scheme.[2] This research focuses on the utilization of pay-for-performance (P4P) mechanisms that are intended to lower hospital costs without increasing hospital readmission. The incentive mechanisms are designed to better align the financial interests of the physicians and the hospital. The performance metrics we use are hospital length of stay and readmission rates. Given the current prospective payment scheme used in the United States, a reduction of either one without increasing the other improves the quality of care at lower costs.

We report the results from two experiments. Experiment 1 investigates the efficacy of two alternative P4P mechanisms, bonus and bundled payments, that tie physicians’ payments to performance. We ask whether the P4P mechanisms can be

[1] The current regulations only address hospital readmissions for patients being treated for three medical conditions: heart attack, heart failure and pneumonia. The $280 million in penalties was spread out across over 2,200 hospitals in fiscal year 2013 (Health Affairs 2013). The scope, and therefore the penalties, of the CMS regulations are expected to increase in the future (CMS 2015).

[2] The prospective payment scheme is implemented in the United States using Diagnosis Related Group (DRG) payments. A hospital receives a flat DRG payment for each patient and procedure event with the payment not varying by the patient’s hospital length of stay. An alternative to this is the fee-for-service (FFS) system where a hospital receives payment for each service provided.

used to reduce hospital costs, without increasing readmissions, compared to baseline outcomes with fee-for-service compensation.[3] Experiment 2 investigates the robustness of the P4P incentive effects in an environment with richer information provided to physicians. Our results suggest that either bonus or bundled P4P physician compensation reduces hospital length of stay for patients but the bundled compensation does so without increasing readmission rates. Additional reductions are observed when we combine the bundled payment mechanism with providing physicians information on the likelihood of readmissions.

Design of an efficient healthcare system, including physician compensation and insurance markets, has been extensively studied in the economics literature beginning with the work of Arrow (1963). In light of the asymmetric information and informational uncertainties in the healthcare market, Arrow (1963) highlighted the need for payment of services, either to physicians or incorporated into insurance markets, to be based on the efficacy of a patient's treatment. This form of compensation is rarely if ever used in current practice. P4P mechanisms are an attempt in this direction as many P4P programs are based on the quality of care, which presumably is correlated with patient health outcomes. This said, the most common forms of healthcare payment are fee-for-service, prospective payment (i.e., diagnosis related groups or DRGs), patient-based capitation (i.e., health maintenance organizations or HMOs) and salaries. The existing economic literature, as discussed below, has compared these incentive structures extensively.

The next section focuses on the literature and discusses the contributions of our research. Sections 3 and 4 report on the details and results of Experiment 1 that we conduct to investigate the efficacy of P4P programs to cost-effectively lower hospital readmission rates with patient information from electronic medical records, as currently provided in hospitals. Section 5 reports on the efficacy of P4P mechanisms in a richer information setting. The final section summarizes our research and provides some additional guidance regarding future research needs in this area.

2. Literature Review

In this section we review theoretical, empirical and experimental studies on the effect of payment schedules on physicians' choice of care for their patients. The main finding is

[3] Our baseline treatment is fee-for-service because prospective payments predominately apply to hospital compensation, whereas physicians still receive fees for the services they provide.

that physicians' selections of diagnostic methods, referrals, and care treatments vary greatly across different payment schemes.

2.a Theoretical Studies

Allard, et al. (2011) theoretically investigate the incentives and outcomes of general practitioners (GPs) under three compensation schemes: (1) fee-for-service (FFS), (2) capitation, and (3) fundholding.[4] Fee-for-service pays for all services rendered, capitation pays a flat fee per patient per year with the GPs being responsible for all care costs they provide, whereas fundholding builds on capitation by making GPs financially responsible not only for the care they provide but also for the care provided by specialists. A fee-for-service payment mechanism creates an incentive for physicians to over-treat their patients, which increases treatment costs but not necessarily the quality of care. The capitation payment scheme pays physicians a flat rate for each patient under their care; it was introduced to internalize the incentive problems of over-treatment associated with FFS. A central research question is whether the compensation scheme, combined with GP ability and preferences, alters the treatment and referral rates of "gatekeeper" GPs.

Allard, et al. (2011) show that: (i) under a capitation scheme GPs are better off referring their patients to a specialist to minimize their own treatment costs; (ii) GPs compensated under a fee-for-service system are less likely to refer a patient; and (iii) under some circumstances referral rates are similar for fundholding and fee-for-service. This arises because fundholding compensation induces GPs to reduce the costs that may be incurred if a patient is referred to specialists when the expected costs of the specialists exceed their own. Extrapolating from earlier work on physician partnership revenue sharing, Gaynor and Gertler (1995) predicted that physicians would dramatically reduce their effort levels under a capitation payment model whereas FFS payment encourages excessive effort levels, a moral hazard effect. An extreme form of capitation would be to provide physicians with a flat salary independent of the number of patients they serve. Woodward and Warren-Boulton (1984) offer a theoretical analysis of this

[4] Fundholding was created under the changes to the United Kingdom's health care system in 1991 in an effort to separate the physician and hospital care markets. For a more detailed review of the fundholding program see Croxson, et al. (2001) and Matsaganis and Glennerster (1994).

compensation scheme and show that it induces physicians to provide less than the efficient level of care.[5]

Another payment schedule, known as a prospective payment scheme, was enacted by CMS in the early 1980s in the form of diagnosis related group (DRG) payments. Under this compensation scheme the hospital receives a flat payment that they must use to cover the expenses of treating a patient (i.e., services provided by the physician) but the flat payment differs across DRGs. Ellis and McGuire (1986) theoretically show that the DRG payment scheme leads to an inefficient supply of hospital services unless the physician serves as a perfect agent for the patient and is not influenced by the hospital's profit motive. They argue in favor of a mixed-payment scheme as the most theoretically efficient payment scheme (Ellis and McGuire 1986).[6] The two P4P schemes we experimentally investigate are similar in spirit to the compensation schemes discussed by Ellis and McGuire (1986) as the payments received by hospitals and physicians are more closely linked than under a conventional prospective payment scheme.

2.b Empirical Studies

In an empirical study directly relevant to our research, Moreno-Serra and Wagstaff (2010) analyze the performance of FFS and prospective payment mechanisms, relative to the use of historical budgets, using a panel data set of 28 European and Central Asian countries for 1990-2004. They find higher admission rates to hospitals under the FFS programs but no impact on the average hospital length of stay for patients. On the contrary, the prospective payment schemes had no effect on inpatient admissions but the average length of stay was reduced. In a later study, Clemens and Gotlieb (2014) estimate that a two percent increase in prospective payment compensation rates results in a three percent increase in physician care, defined as the number of relative value units (RVUs) provided to a patient. Their finding provides further evidence that physicians respond positively to increases in compensation rates.

Pay-per-performance (P4P) mechanisms are being increasingly advocated because they are believed to incentivize physicians and hospitals to provide better

[5] The empirical literature in this area reports mixed results. The findings of Barro and Beaulieu (2003) and Hickson et al. (1987) clearly support the hypotheses of Woodward and Warren-Boulton (1984) but the work of Grytten and Sørensen (2001) does not.
[6] In a mixed-payment scheme a portion of hospital expenses is covered via prospective payment whereas the rest is covered from a cost compensation model where payments are received based on the cost of care, a form of FFS.

quality care.[7] However, empirical findings on the quality of care are mixed with some studies reporting support for P4P (Beaulieu and Horrigan 2005, Lindenauer et al. 2007) and others finding no effect (Rosenthal and Frank 2006).[8] Beaulieu and Horrigan (2005) research the effectiveness of quality-based annual bonuses to incentivize higher quality of care for diabetic patients and find that the P4P mechanism increased the patient's health for five of the six reported measures.[9] Lindenauer et al. (2007) study the effectiveness of quality-based bonuses based on a physician's performance relative to their peers. They report improvements in quality for seven of their ten measures of hospital performance. In contrast, Mullen, et al. (2010) report that they failed to find evidence that P4P programs result in any substantial improvements in patient care. They look at whether P4P programs increased the quality of healthcare across a broad suite of P4P incentives in California and the Pacific Northwest. The P4Ps under study provided bonuses to physicians based on whether they met specific clinical quality metrics. An important feature of the Mullen, et al. (2010) study for our research is that one of the metrics they used was hospital readmissions for outpatient care; they find no conclusive evidence that the (bonus) P4P mechanism lowered readmission rates.

Several P4P mechanisms have been implemented in the United Kingdom and Canada. In 2004 the United Kingdom initiated a P4P program that awarded points to practicing physicians based on 146 quality indicators that were used to generate income for the physicians. The government expected physicians to reach 75% of the total achievable points, but the median physician hit 96.7% of the total (Doran, et al. 2006). This resulted in a substantially higher cost for the government than anticipated but the incentives did generate a change in physician practice consistent with the objectives of the program. In Ontario, Canada eleven different P4P mechanisms were introduced between 1998 and 2008 (Li, et al. 2014). Six of these P4P mechanisms were special payments made for taking specific actions whereas five were bonus mechanisms based on performance targets. The results were mixed as the bonuses were effective for some procedures (i.e., pap smears, mammograms, flu shots) but not for others (i.e., palliative care, prenatal care, home visits) (Li, et al. 2014).

---

[7] Performance incentives are not limited to the health care literature. For example, Bandiera, et al. (2005, 2009) investigate the impact that alternative payment mechanisms have on workforce productivity using field experiments.

[8] Despite these mixed results, Hemenway et al. (1990) find that P4P mechanisms were effective in incentivizing physicians to increase billings at for-profit emergency rooms. Physicians in their study strongly responded to financial incentives that affected their earnings.

[9] A caution is warranted as enrollment in the P4P was not random.

There are a number of limitations of P4P programs that may explain why the results have been mixed. Many of these studies are small in scale, based on Medicare patients, and have short intervention time periods (Rosenthal and Frank 2006). In addition, in the U.S. health care system physicians receive compensation for their services from a large number of providers, and it is possible that the failure to incentivize the physicians can be attributed to interventions being too small relative to other payments received (Rosenthal and Frank 2006). The experiments we conduct will allow us to control for some of these factors (i.e., tangible incentives, representative patient sample, well constructed treatment and control groups) and to better isolate the efficacy of P4P programs. Our experiments use a sample of experimental healthcare providers randomized over three different compensation environments tied to effective care of the patients and relative performance.

There are two additional limitations of P4P that we will not be able to address with our experiments but that we can control for in our design: the multitasking nature of a physician's decision space and the imperfect matching of patients and physicians. Eggleston (2005) illustrates that a physician has an incentive to increase the quality of care in the rewarded dimensions of a P4P program at a cost of reducing the quality of care in other dimensions of the service they provide because they are multitasking. This has the potential to have a negative welfare effect. Imperfect matching between patients and physicians reduces the ability of the P4P mechanism to incentivize the appropriate physicians that the program targets. Using Medicare claims data, Pham et al. (2007) show that no more than half of patients' visits are with physicians who would be responsible for them under a P4P program; this is attributed to instability in the patient-physician matches. Therefore, it may be difficult under current P4P programs to account for dispersion of care. The experiment we conduct controls for these factors. In the experiment our physician subjects are responsible for making only one decision, to discharge or not to discharge a patient, thus eliminating the effects of physician multitasking on outcomes. Furthermore, each physician subject in the experiment is responsible for all the patients under her care thus ensuring stability in the patient-physician pairing.

2.c Experimental Studies

The experiments we conduct are not the first to investigate the effect of alternative payment mechanisms on physician behavior (see, for example, Hennig-Schmidt, et al.

2011; Fan, et al. 1998). However, to the best of our knowledge, we are the first to investigate the efficacy of P4P mechanisms that attempt to more closely align the physician and hospital incentives that are being increasingly proposed as a solution to the rift between the cost of patient care and the quality delivered.

Fan, et al. (1998) conduct an experimental analysis of physician decisions under two alternative payment mechanisms: (1) expenditure targets and (2) expenditure caps. Under an expenditure target payment mechanism physicians individually profit maximize whereas under an expenditure cap model physician payment is determined also by the actions of other physicians. Their results indicate that the level of health services is greatest under the cap system given the same level of budget. Hennig-Schmidt, et al. (2011) conduct an experiment using 42 medical students to investigate the impact that a FFS and capitation (CAP) payment mechanism have on: (1) the supply of medical services; (2) the relationship between patient health status and selected treatments; and (3) the resulting health status of patients. There were 15 hypothetical patients in their experiments divided into 3 types of healthcare needs: low, moderate, and high need. They find that more care is provided under a FFS than CAP system, implying that patients are under-served under the CAP and over-served under FFS, relative to the induced optimal treatment benefits for the patient.[10] The finding is consistent with the predictions of Gaynor and Gertler (1995). They also find that a patient's health status influences the provision of care: patients with low to moderate health needs are over-served under a FFS and those with high needs are underserved under the CAP system.

3. Experiment 1: P4P Compensation and EMR Information

We experiment with hospital discharge decision making using medical students enrolled in a large medical school within the southeastern United States. At the beginning of each experimental session, the subjects were welcomed to the decision laboratory by one of the researchers who explained that the research was supported by an NIH grant with the purpose of investigating discharge decision making.[11] Subjects read and signed the IRB-approved consent form and subsequently began reading the subject instructions on their

[10] In the Hennig-Schmidt, et al. (2011) experiments a patient benefit function has diminishing marginal benefits. All subjects are aware of the benefits each treatment will provide to their hypothetical patient as well as their internal costs and benefits.

[11] The lead point of contact for the experiments was the PI on the NIH grant funding our research at the southeastern United States medical school where the experiments were conducted.

computer monitors.[12] For each "experimental day", subjects reviewed individual patient data from a list on the screen that displayed fictitious patient IDs and names along with real summary information from *three* distinct patient charts. The real summary information included the patient's age, sex, and current length of stay in the hospital (up to the current experimental day) taken from de-identified electronic medical records (EMRs). For any patient on the list, the subject was presented with a series of charts that were facsimiles of the patient charts in the EMR, retaining the de-identified nature of the data. Hospital records for patients used in the experiment were obtained from the "data warehouse" of a large southeastern medical school where the experiments were conducted. The patient charts used in the experiment are de-identified electronic medical records for patients served at this hospital. EMR facsimile screens with data for a sample de-identified patient used in the experiment are shown in appendix 1. The appendix 1 screen shots show a single static view of a patient's EMR. Software used in the experiment dynamically updated the chart information reported in the EMR screens with each additional "experimental day" adding the new EMR information from another 24 hour period included in the hospital's EMR for a patient.

To conduct our experiment we selected 30 representative patient charts with varying degrees of patient risk characteristics. We selected an equal balance of 10 low risk, 10 medium risk and 10 high risk patients based on the historical readmission rates for the procedure the patient was submitted to the hospital. In all treatments, the subjects were informed that they should assume that a patient was being managed at the appropriate standard of care while in the hospital and that the subjects were *not* being asked to speculate about additional tests or procedures that they might want to order. Instead, they were asked only to make a decision on whether to discharge each patient served on an experimental day.

On each experimental day, representing EMR patient data for one calendar day in the hospital, subjects see three patients and when one patient is successfully discharged (i.e., does not get readmitted) another patient is placed under the control of the subject. Each subject makes a total of 30 discharge decisions (including unsuccessful ones). However, the number of patients being seen during the experiment, the duration of the experiment and the number of experimental days are all endogenously determined by the subject's decisions. Fictitious names were used for the patients to ensure anonymity.

[12] Subject instructions for the experiment can be found at http://excen.gsu.edu/jccox/instructions.

The three treatments, representing alternative physician compensation schemes, in Experiment 1 are listed in the top row of Table 1: flat fee for service (FFS), instantaneous profit-sharing bundled (BU) payments, and deferred profit-sharing bonus (BO) payments. All three schedules offer payment only for successful discharges. A "successful" discharge decision is one that is not followed by a readmission within 30 days.

The first payment plan, flat fee-for-service (FFS) pays each subject $5 for each successful discharge decision.[13] With this payment mechanism there are no rewards or penalties for keeping the patient fewer or more days. This payment schedule offers incentives for over treatment because the payment is received only for successful discharges while readmission probabilities may decrease with longer hospital stays. To calibrate the other two payment mechanisms, we used the subject choices under the FFS treatment to generate payments under the P4P mechanisms that would potentially incentivize subjects to cost-effectively reduce the patient's length of stay while meeting the CMS targeted reductions in readmission rates. We calculated the average length of stay (LOS) within the FFS treatment for each patient. We then used this average LOS to parameterize our other two payment mechanisms in a way that the payment received would be identical across the three treatments if the patient were to be discharged on the day that corresponds to its average LOS in FFS.

Under the bonus (BO) payment treatment subjects received a flat fee of $3 for each successful discharge and then a bonus payment that was awarded to them at the end of a period of 10 discharges as the lump-sum of all the bonuses derived for all of the successfully discharged patients during that period.[14] The bonus was calculated such that each day the physician subject shortened a patient's length of stay below the average LOS (across subjects) for that patient under the FFS treatment pays an extra $0.50 and every day beyond the average LOS under the FFS treatment comes with a loss of $0.50. This information was provided to subjects on their decision screens so they could see the schedule of bonuses. This treatment represents a deferred payment P4P mechanism as the performance-adjusted earnings are obtained after every 10 discharges.

[13] Accumulated earnings were updated at the end of each successful discharge. The possible wealth effects of this payoff protocol were minimized by the independent random order of patients for each subject.
[14] There is a portfolio or hedging incentive within each ten periods of successful discharges. This incentive in the experiments reflects the incentives created by the P4P mechanism itself.

The bundled (BU) payment mechanism is similar in the final payment to the one offered under the BO treatment but the fixed payment and bonus payments are not separated; the payments are combined into one payment. In the BU treatment each subject receives an instantaneous reward or penalty for decisions within the experiment. As was the case under the FFS incentive structure subjects were only paid for successful discharges under the BO and BU P4P mechanisms. Both mechanisms, BO and BU provide incentives not only for good patient care, as payment is received only for successfully discharged patients, but also for avoiding over treatment as the payment decreases for every day the patient is kept in the hospital. In addition each readmission costs at least $1 because a readmitted patient cannot be discharged earlier than the third day after readmission, providing stronger incentives for higher quality of care while at the same time making the subject accountable for an unnecessary patient stay in the hospital. The average individual subject payoff in Experiment 1 was $147 and the experiment lasted, on average, 90 minutes.

The probit econometric model used to construct the clinical decision support system developed in Cox, et al. (2014) provided an anchoring point for the experimental data by providing an objective measure of when a patient should be discharged regardless of the P4P mechanism utilized. Given the heterogeneity in the risk types among the 30 patient records used in the experiment we randomly selected the first *experimental* day on which the subject was asked to make a discharge decision to be between one and four days before the first day that the patient should have been discharged according to the decision support system.[15] The number of preceding days was independently randomly selected for each of the 30 patients. Furthermore, the ordering of the 30 patient charts was independently randomly selected for each of the subjects within the experiment.[16] Any patient who was discharged could either be

[15] A patient's actual hospital length of stay is correlated with their readmission risk. Our randomization process ensures that all patient types are treated equally with regard to when a subject evaluates the patient chart information within the experiment. The correlation between readmission risk and the hospital length of stay is controlled for in our regression analyzes with the inclusion of our target probability variable.

[16] To ensure that the experimental subjects' discharge decisions were not influenced by the decisions of the practicing physician who actually took care of the patient, we removed the dates of actual discharge from the patient charts. This created the possibility for the *experimental* LOS to exceed the actual number of days observed in the EMR for a patient. To address this we created continuation charts for all 30 patients that imputed an extra five days of data assuming the data recorded in the EMR for the last day remained stable over this time period.

successfully discharged or readmitted.[17] In the case a patient was readmitted the patient was returned to the subject's patient queue with a new set of patient data generated to reflect a particular complication that would generate a readmission. Readmitted patients were required to remain in the queue for at least two days before the subject could choose to discharge them again and the subject was required to review the patient's chart for each of those days. Each subject was only allowed to make a total of 30 discharge decisions during the experiment.

After making their discharge decisions subjects completed an online questionnaire that was embedded in the experiment software. The questionnaire elicited demographic information such as subject age, gender, academic performance and non-academic interests (i.e., athletics, musical instruments). The questionnaire also included hypothetical response questions about risk attitudes to get some information about subjects attitudes toward risk that could affect their decision making in the presence of health related risks.[18] After completing the questionnaire, subjects exited the lab one at a time to be paid in private.

4. Data from Experiment 1

A total of 100 subjects participated in Experiment 1. The distribution of subjects over the three treatment cells is reported in the first row of Table 1: 24 subjects participated in FFS payment treatment and 36 and 40 subjects participated in BO and BU payment treatments respectively. All of the subjects participating in the experiment were third or fourth year medical students at a large southeastern medical school who were in the clinical training phase of their medical education.[19] Each subject participated in only one of the three treatment cells; we utilized a between-subjects design to investigate the treatment effects of our P4P payments, BO and BU.

The characteristics of the subjects participating in each of the three treatments were similar. The overall number (48) of female participants was similar to the number (52) of male subjects within the experiment. The composition of males and females across the three P4P mechanisms were not significantly different from one another.

[17] The daily likelihood of readmission was calculated using the probit model developed by Cox et al. (2014).
[18] The questionnaire can be found at http://excen.gsu.edu/restricted/subjectInstructions/cer/PostExperimentQuestionnaire.pdf
[19] We restricted our sample of medical students to only those in the clinical phase of their training to ensure they all possessed the necessary baseline information to make an informed discharge decision.

There were 13 males and 11 female subjects in the FFS treatment, 22 male and 14 female subjects in the BO treatment and 17 male and 23 female subjects in the BU treatment (Pearson chi2(2)=2.69, p-value=0.261). Academic performance of subjects who participated in different treatments was at comparable levels. The reported average grade point in medical school of subjects in the FFS, BO and BU payment treatments were 3.58 (std=0.219), 3.57 (std=0.260) and 3.55 (std=0.284) respectively. A statistical test of differences across the three treatment cells failed to reject the null hypothesis that academic performance grades across the treatments were drawn from the same distribution (Kruskal-Wallis test: chi2(2) = 0.250, p-value=0.883).

There are three measures of performance across the three treatment cells that we look at in our analysis: subject earnings, quality of service (readmission rate), and hospital length of stay. Table 2 provides descriptive statistics for each of the three groups of subjects in the experiment for the key variables of interest as well as a few demographic characteristics (collected in the post-experiment survey and referenced above). We define LOS as the number of days that the physician subject retained the patient in their queue before discharging. Furthermore, we do not include readmitted patients in the LOS calculations for two reasons: (1) these data does not precisely match up with the data obtained from the data warehouse due to our imputed medical complications, and (2) readmitted patients must be retained at least two days before a physician subject can consider discharging the patient. Both of these factors may introduce bias in the LOS for readmitted patients.

Comparison across payment schemes shows that both P4P compensation schemes reduce LOS per patient by about 1 day, increase earnings for subjects by about $12 while increasing readmission rates by about two to three percent.[20] We next report several ways of describing the data and statistical analysis for significance of treatment effects at the individual level.

4.a Readmissions as an Indicator of the Quality of Care

An earlier discharge is not an indicator of better discharge decision making if it decreases the quality of care. An indicator of the quality of care is the readmission rate since a premature discharge increases the likelihood of an unplanned but necessary

[20] With respect to readmission rates, however, the null hypothesis of no effect of payment mechanisms on the readmission rates is not rejected by our data (Kruskal-Wallis test, chi2(2)=3.81, p=0.149).

readmission. The average readmission rates were 9% (FFS), 11% (BU) and 12% (BO). To investigate the treatment effects on the readmission rate we ran probit regressions with a binary dependent variable that takes a value of 1 if a patient is readmitted. The results from the regression are contained in the Experiment 1 columns of Table 3 (first three columns). We report the estimated marginal coefficients (and p-values in parentheses) with clusters at the subject level; the omitted reference group is the FFS treatment. The first column contains the regression results using just the treatment dummies within the experiment. The second column adds demographic information, which includes gender, whether or not the subject was a student-athlete in college or plays a musical instrument, the current grade point average (GPA) in medical school, the undergraduate GPA and the risk attitude index. The third column includes an additional variable, the target probability. This variable represents a 10% reduction in the readmission rate relative to the historically observed procedure-specific readmission rate and is used to define our high, medium and low risk patient types.[21] Our prior is that patients with a higher target readmission rate may require longer LOS as their surgical procedures are more complex. Our main finding is that BU compensation does not induce lower quality of care as the probability of readmissions is (statistically) similar to the one observed under FFS but the BO compensation comes with slightly higher readmission rates. This gives us our first primary result:

Result 1. *Use of the BU payment mechanism does not significantly increase readmission rates but the BO induces higher readmissions.*

The demographic controls used in the probit regression reveal that subjects with an athletic or musical background had lower rates of patient readmission whereas a higher medical GPA is correlated with higher readmission rates.[22] We next turn our attention to the LOS across treatments.

4.b Hospital Length of Stay

Figure 1 shows (Gaussian kernel) densities of the distributions of LOS across the three payment mechanisms. It suggests that the distribution of FFS has a fatter right tail than

[21] The 10% reduction is based on the targets set by CMS discussed earlier. The targeted readmission rate is less than 10% for the low risk patients, between 10% and 17% for the medium risk patients and greater than 17% for the high risk patients.
[22] A record as a competitive athlete or musician is selected for in medical school admissions.

the other two, implying a longer LOS in the FFS treatment than the BU and BO treatments. The 95% C.I. of the mean of LOS for FFS ([3.85, 4.37]) does not overlap with the 95% C.I. of the means of LOS for the two P4P mechanisms, BU ([2.94, 3.26]) and BO ([2.81, 3.17]).[23]

To further investigate the treatment effects on LOS we ran Censored-Normal regressions on the observed patient LOS using the same covariates in the probit regression as well as the "Start Date" for a patient as this information is observed by subjects within the experiment. The Start Date is the patient's current number of days within the hospital before the physician subject observed them within the experiment. The results from this regression with robust standard errors are reported in the left three columns of Table 4. The regression results indicate that both P4P compensations decrease LOS by slightly over one day; the estimates are robust across the three specifications. This generates our second primary result:

Result 2. *Use of P4P compensation reduces hospital length of stay.*

Our data provide evidence that altering the compensation mechanism to better align incentives of physicians, hospitals and patients, either BU or BO, reduces the hospital length of stay. However, the BU payment mechanism provides a more cost-effective method to meet the quality standards set forth by CMS as it reduces LOS without increasing readmission rates. The BO compensation scheme reduces LOS but increases the rate of hospital readmission by 3% (statistically significant at the 10% level after controlling for subject's idiosyncratic characteristics).

5. Experiment 2: P4P Compensation and EMR+CDSS Information

Rapid technological progress facilitated by the use of automated processes and computers enhances the opportunities for providing physicians with richer information on the health trajectories of their patients on a daily basis. A remaining question is whether in the presence of a richer information structure the effects of the compensation mechanisms on the quality of physicians' decisions remains. To get some insights on the robustness of the results reported earlier, we conducted a second experiment that is

[23]The 95% C.I. of the mean of total LOS (sum of the time before a patient was observed by a physician subject and the number of days that the patient was retained in the queue) for FFS ([7.73, 8.30]) does not overlap with the 95% C.I. of the means of LOS for the two P4P mechanisms, BU ([6.77, 7.14]) and BO ([6.63, 7.07]).

identical to Experiment 1 except that the subjects were provided discharge recommendations and information on the likelihood of readmissions for each patient on each day in the hospital.

All features of Experiment 2 are the same as Experiment 1 except the way in which information was presented to the subjects. Experiment 2 was conducted in an environment that included the clinical decision support system (CDSS) described in Cox, et al. (2014). In addition to the EMR information presented to subjects in Experiment 1, the information presented to subjects in Experiment 2 included the CDSS information screens and dynamically updated recommendations on whether to discharge a patient on each "experiment day" which corresponded to a 24 hour day in the electronic medical record for a patient. The CDSS makes one of three recommendations for each patient on each experiment day: (1) do not discharge the patient, (2) physician judgment, or (3) discharge the patient. Information and discharge recommendation screens for the CDSS are shown in Cox, et al. (2014).[24]

A total of 109 subjects participated in Experiment 2. They were distinct from the subjects in Experiment 1. The distribution of subjects over the three treatment cells is reported in the second row of Table 1. All of the subjects participating in the experiment were third or fourth year medical students at a large southeastern medical school who were in the clinical training phase of their medical education. Each subject participated in only one of the treatment cells and we utilize a between-subjects design to investigate the treatment effects of the P4P mechanisms and their interaction with the two (EMR or EMR+CDSS) information treatments. The average earning for physician subjects in Experiment 2 was $159.

5.a Readmissions *as an indicator of the quality of care*

The average readmission rate for the FFS treatment was 8.74% and 7.69% for the two respective information treatments, EMR and EMR+CDSS respectively. The average readmission rate increased (by about 2%) for both the BU and BO payment treatments, but by less than that observed in the absence of information (Experiment 1). The null hypothesis of the readmission rates coming from the same distribution is not rejected by data from EMR+CDSS cells (chi2(2)=2.751, p=0.253). In the BU treatment the average readmission rate was 10.96% in the EMR information treatment and 9.91% under

[24] These screens are available at: http://excen.gsu.edu/jccox/docs/CDSS-Information-and-Decision-Screens.pdf

EMR+CDSS, and in the BO treatment the average readmission rate was 12.04% under the EMR and 9.27% under EMR+CDSS. The data reject the null hypothesis that the readmission rates observed in the EMR and EMR+CDSS cells come from the same distribution in favor of the alternative hypothesis that the CDSS decreases readmissions (Mann-Whitney, p=0.067; t-test, p=0.041; means are 10.80% (EMR) and 9.18% (EMR+CDSS)). The EMR+CDSS treatment significantly decreases readmissions compared to the EMR treatment. This finding is consistent with the research findings of Cox, et al. (2014).

To further investigate the treatment effects on the readmission rate we ran probit regressions with a binary dependent variable that takes a value of 1 if a patient is readmitted. The results from the regression are reported in the last four columns of Table 3. We report the estimated marginal coefficients (and p-values in parentheses) with clusters at the subject level. The omitted reference group is the FFS enhanced with information on the daily likelihood of readmissions and CDSS recommendations (FFS-EMR+CDSS treatment). The covariates utilized are the same covariates used in our analysis of Experiment 1 except for the model reported in the fourth column. In the fourth column we add two additional covariates, "Understay" and "Overstay." Understay is the number of days a patient is discharged before the software recommended LOS, whereas Overstay is the number of days a patient was retained by the physician subject after the recommended LOS.[25] Focusing on the specification that utilizes all of the covariates in the model, our preferred specification, we find that relative to the readmissions observed under FFS-EMR+CDSS the probability of readmission did not increase when we utilized either the BU or BO P4P mechanisms.

Referring to parameter estimates for the Understay and Overstay variables, we find that Understay and Overstay, respectively, increase and decrease readmissions; keeping patients one day less than recommended by the discharge recommendation software increases the likelihood of readmissions by 1.9% whereas keeping the patient one more day than recommended decreases the likelihood of readmissions by 1.1%. The recommended hospital length of stay (Recommended LOS) also has a significantly negative effect on readmissions. This gives us our third result:

[25] We do not include this information in the probit regression analysis for experiment 1 because the recommended LOS (i.e., the CDSS) was not provided to our subject physicians in that experiment.

Result 3. *Discharging a patient earlier than recommended significantly increases the likelihood of overall unplanned readmission; the magnitude of the effect is stronger than the effect of a later-than-recommended discharge.*

5.b Hospital Length of Stay

We find that with the EMR+CDSS treatment, the empirical distributions of LOS and earnings across the three payment mechanisms are statistically different at conventional levels of significance[26] with the bundled P4P mechanism performing best. Observed means of hospital LOS across the three treatments (FFS, BO, BU) are: LOS = (2.96, 2.67, 2.08).[27] Data from the EMR+CDSS treatments reveal higher efficiency (lower LOS as well as lower readmissions) than data from the EMR treatments, but within the EMR+CDSS treatments the BU (bundled) P4P mechanism is the most efficient one.

A decomposition of the performance of payment mechanisms over all EMR and EMR+CDSS treatments shows that the average observed LOS is reduced by the CDSS. A close inspection of figures in the "Patient LOS" row of Table 2 suggests that the use of the P4P mechanisms reduced the average LOS for patients with additional reductions in the LOS resulting from the use of the CDSS. To more rigorously investigate the treatment effects (in the presence of enhanced information) on LOS we ran Censored-Normal regressions on the observed LOS using the same covariate controls as in our analysis of Experiment 1. The results from this regression with robust standard errors are reported in the last three columns of Table 4.

The regression results indicate that with richer information the BU-EMR+CDSS incentive mechanism reduces the LOS relative to the FFS-EMR+CDSS treatment, with the observed reduction being approximately 1 day. However, it is worth noting that the FFS-EMR+CDSS treatment already generated an average reduction in the LOS relative to the FFS-EMR treatment of approximately one day. Therefore, the BO-EMR+CDSS treatment still generated a reduction in the LOS relative to the FFS-EMR treatment.[28]

[26] Kruskal-Wallis test: LOS: chi2(2)=8.15, p=0.017; Earnings: chi2(2)=38.02,p=0.0001).

[27] Observed means of total LOS (sum of the time before a patient was observed by a physician subject and the number of days that the patient was retained in their queue) across the three treatments (FFS, BO, BU) are: LOS = (6.87, 6.55, 5.98).

[28] To validate this we ran another censored regression that pooled the data from both experiments. The treatment coefficients from this regression (with the list of covariates as in the last column of Table 4) were -1.062 (p=0.011) for BU-EMR, -1.099 (p=0.019) for BO-EMR, -1.030 (p=0.028) for FFS-EMR+CDSS, -2.123 (p=0.000) for BU-EMR+CDSS and -1.400 (p=0.001) for BO-EMR+CDSS. The omitted reference group was FFS-EMR and all coefficients were statistically significant different from 0 (at least at the 3% level).

The BU P4P mechanism combined with richer information on daily readmissions appears to generate the largest reductions in LOS relative to the FFS treatment without information. After adjusting for multiple testing, we find that the combination of the CDSS and bundled payments does best in reducing the hospital length of stay in our controlled environment. We conclude that:

Result 4. *Use of P4P compensation, with or without CDSS, reduces hospital length of stay; the effect on LOS is greater with CDSS.*

This final result provides evidence that altering the compensation mechanism provides a cost-effective method to meet the quality standards set forth by CMS if one combines these P4P mechanisms with the information provided by the CDSS.

6. Conclusion

The current healthcare system is in dire need of ways to increase the quality of care it provides while decreasing the cost of providing it. Given the recent pressure by CMS on hospitals to reduce their readmission rates while maintaining the current prospective payment system, hospitals must find creative ways to cost-effectively meet the standards. A potential area where hospitals can improve on their cost-effectiveness is by addressing the extant asymmetry between patients, hospital and physician incentives. Currently, physicians and hospitals do not have the same economic incentives for providing care to a patient. In this research we experimentally tested two alternative pay-for-performance mechanisms that would more closely align these incentives, combined with providing discharge recommendations, to explore the effects of these mechanisms on cost-effectively meeting the quality targets recently enacted by CMS.

Using a three-by-two experimental design defined over three payment mechanisms (fee-for-service, bonus and bundled payments) and two information conditions (EMR and EMR+CDSS), our research indicates that the P4P mechanisms used in our experiment combined with the clinical decision support system (Cox et al. 2014) can be used to make more cost-effective and evidence-based hospital discharge decisions. The largest reduction in average LOS resulted from the bundled payment mechanisms combined with the CDSS information, where the LOS fell by nearly two days. This reduction in LOS did not come with an increase in the probability of readmission. Therefore, the bundled payment mechanism proved to be a very cost-effective mechanism to reduce LOS and meet readmission targets, especially when supported by the CDSS.

Although not of the same magnitude as those observed under the bundled payment mechanism, the bonus payment mechanism with the CDSS resulted in a reduction in LOS of 1.4 days. Across the treatments the bonus and bundled payment models reduced LOS by at least one day. Therefore, utilizing P4P mechanisms to better align physician, hospital (and indirectly patient) interests is a cost effective way to achieve the quality targets set forth by CMS. In addition, the information provided by the clinical decision support system (CDSS) reduced LOS without increasing readmission rates. With both fee-for-service and bundled payment compensation use of the CDSS reduced LOS by approximately one day whereas the additional reduction in LOS with the bonus mechanism was smaller (approximately 0.4 days).

Our estimates suggest a sizeable cost savings to a hospital if it utilizes the bundled payment mechanism as well as the CDSS. In fact, each provides a cost-effective way to meet Medicare targets in that they have additive effects. This said, there are a few limitations that are worth mentioning. First, the utilization of either of the P4P mechanisms used in our experiment will require buy-in from both hospitals and physicians in order for it to be effective. The latter of these two parties may find these mechanisms unpalatable because they more concretely link their compensation to the cost of care and force them to make cost-benefit calculations that they may argue compromise physician autonomy. Again, piloting of these P4P mechanisms would be required before full-scale utilization. Second, the CDSS requires additional validation and piloting before it can be fully integrated within a hospital electronic medical records system. This is something we are researching at the current time.

In summary, this research suggests that there are a number of ways in which hospitals can cost-effectively reduce their readmission rates and meet the quality metric set forth by CMS. Continued research in this area is needed and we hope our findings stimulate an increased use of experimental economic methods to meet the challenges of the healthcare sector.

References

Allard, M., Jelovac, I. and P.T. Leger. 2011. Treatment and Referral Decisions Under Different Physician Payment Mechanisms. *Journal of Health Economics* 30: 880-893.

Arrow, K.J. 1963. Uncertainty and the Welfare Economics of Medical Care. *American Economic Review* 53(5): 941-973.

Bandiera, O., Barankay, I. and I. Rasul. 2005. Social Preferences and the Response to Incentives: Evidence from Personnel Data. *Quarterly Journal of Economics* 120(3): 917-962.

Bandiera, O., Barankay, I. and I. Rasul. 2009. Social Connections and Incentives in the Workplace: Evidence from Personnel Data. *Econometrica* 77(4): 1047-1094.

Barro, J. and N. Beaulieu. 2003. Selection and Improvement: Physician Responses to Financial Incentives. *NBER Working Paper 10017*.

Beaulieu, N.D. and D.R. Horrigan. 2005. Putting Smart Money to Work for Quality Improvement. *Health Services Research* 40(5): 1318-1334.

Blomqvist, A. 1991. The Doctor as Double Agent: Information Asymmetry, Health Insurance, and Medical Care. *Journal of Health Economics* 10: 411-432.

Centers for Medicare and Medicaid Services (CMS). 2015. Website: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html Last accessed on March 16, 2015.

Chandra, A. and J. Skinner. 2012. Technology Growth and Expenditure Growth in Health Care. *Journal of Economic Literature* 50(3): 645-680.

Clemens, J. and J.D. Gottlieb. 2014. Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health? *American Economic Review* 104(4): 1320-1349.

Cox, J., Sadiraj, V., Schnier, K.E. and J.F. Sweeney. 2014. Higher Quality and Lower Cost from Improving Hospital Discharge Decision Making. *Working Paper 2014-07, Georgia State University Experimental Economics Center.*

Croxson, B., Propper, C. and A. Perkins. 2001. Do Doctors Respond to Financial Incentives? UK Family Doctors and the GP Fundholder Scheme. *Journal of Public Economics* 79: 375-398.

Devlin, R.A. and S. Sarma. 2008. Do Physician Remuneration Schemes Matter? The Case of Canadian Family Physicians. *Journal of Health Economics* 27: 1168-1181.

Doran, T.D., Fullwood, C., Gravelle, H., Reeves, D., Kontopantelis, E., Hireoh, U. and M. Roland. 2006. Pay-for-Performance Programs in Family Practices in the United Kingdom. *New England Journal of Medicine* 355(4): 375-384.

Dumont, E., Fortin, B., Jacquemet, N. and B. Shearer. 2008. Physicians' Multitasking and Incentives: Empirical Evidence from a Natural Experiment. *Journal of Health Economics* 27: 1436-50.

Eggleston, K. 2005. Multitasking and Mixed Systems for Provider Payment. *Journal of Health Economics* 24: 211-223.

Ellis, R.P. and T.G. McGuire. 1986. Provider Behavior Under Prospective Reimbursement. *Journal of Health Economics* 5: 129-151.

Fan, C., Chen, K. and K. Kan. 1998. The Design of Payment Systems for Physicians Under Global Budget – An Experimental Analysis. *Journal of Economic Behavior and Organization* 34: 295-311.

Fuchs, V.R. and Millstein, A. 2011. "The $640 Billion Question — Why Does Cost-Effective Care Diffuse So Slowly?" *New England Journal of Medicine*, 2011, 364, 1855 – 1987.

Gaynor, M. and P. Gertler. 1995. Moral Hazard and Risk Spreading in Partnerships. *RAND Journal of Economics* 26(4): 591-613.

Grytten, J. and R. Sørensen. 2001. Type of Contract and Supplier-Induced Demand for Primary Physicians in Norway. *Journal of Health Economics* 20: 379-393.

Health Affairs. 2013. Medicare Hospital Readmissions Reduction Program. *Health Policy Brief* November 12, 2013. Available online: http://www.healthaffairs.org/healthpolicybriefs/brief.php?brief_id=102
Last accessed on March 16, 2015

Hemenway, D., Killen, A., Cashman, S.B., Parks, C.L. and W.J. Bicknell. 1990. Physicians' Responses to Financial Incentives. *New England Journal of Medicine* 322(15): 1059-1063.

Hennig-Schmidt, H., Selten, R. and D. Wiersen. 2011. How Payment Systems Affect Physicians' Provision Behavior – An Experimental Investigation. *Journal of Health Economics* 30: 637-46.

Hickson, G.B., Altemeier, W.A. and J.M. Perrin. 1987. Physician Reimbursement by Salary or Fee-for-Service: Effect on Physician Practice Behavior in a Randomized Prospective Study. *Pediatrics* 80(3): 344-350.

Iversen, T. and H. Lurås. 2000. The Effect of Capitation on GP's Referral Decisions. *Health Economics* 9: 199-210.

Institute of Medicine, National Academy of Sciences, *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America.* Washington, DC: The National Academies Press, 2012.

Li, J., Hurley, J., DeCicca, P. and G. Buckley. 2014. Physician Response to Pay-for-Performance – Evidence from a Natural Experiment. *Health Economics* 23(8): 962-978.

Lindenauer, P.K., Remus, D., Roman, S., Rothberg, M.B., Benjamin, E.M., Ma, A. and D.W. Bratzler. 2007. Public Reporting and Pay for Performance in Hospital Quality Improvement. *New England Journal of Medicine* 356(6): 486-496.

Matsaganis, M. and H. Glennerster. 1994. The Threat of 'Cream Skimming' in the Post-Reform NHS. *Journal of Health Economics* 13: 31-60.

Melichar, L. 2009. The Effect of Reimbursement on Medical Decision Making: Do Physicians Alter Treatment in Response to a Managed Care Incentive? *Journal of Health Economics*.

Moreno-Serra, R. and A. Wagstaff. 2010. System-wide Impacts of Hospital Payment Reforms: Evidence from Central and Eastern Europe and Central Asia. *Journal of Health Economics* 29: 585-602.

Mullen, K., Frank, R.G. and M.B. Rosenthal. 2010. Can You Get What You Pay For? Pay-For-Performance and the Quality of Healthcare Providers. *RAND Journal of Economics* 41(1): 64-91.

Pham, H.H., Schrag, D., O'Malley, A.S., Wu, B. and P.B. Bach. 2007. Care Patterns in Medicare and Their Implications for Pay for Performance. *New England Journal of Medicine* 356(11): 1130-1139.

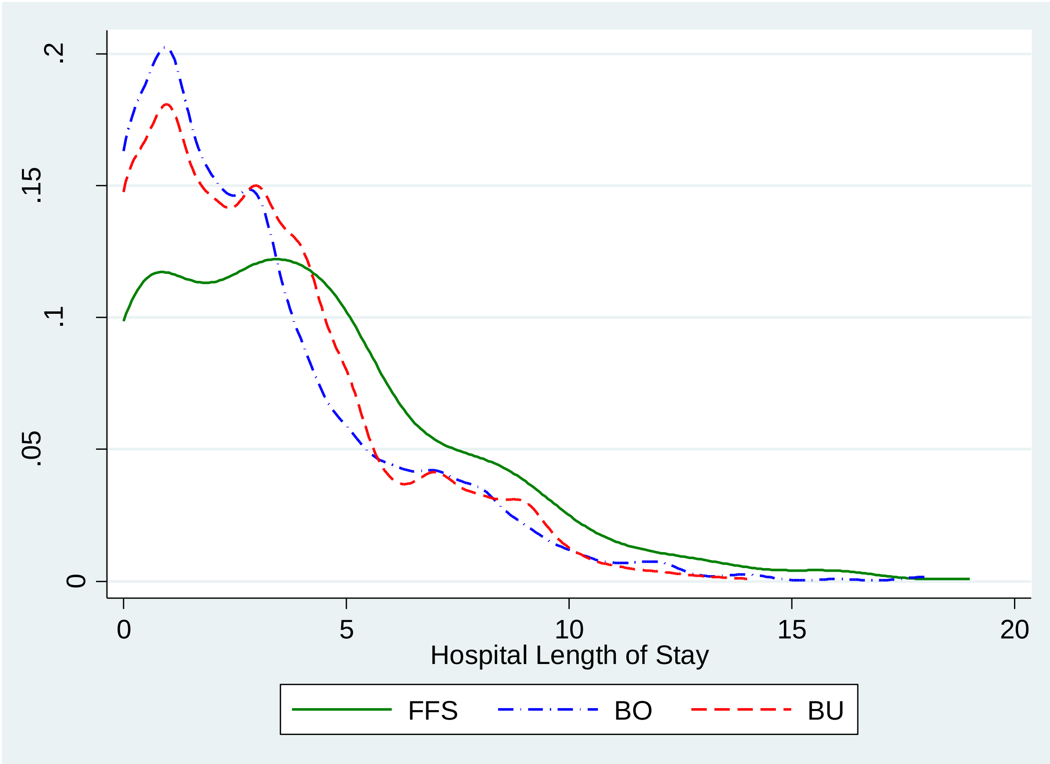
Rosenthal, M.B. and R.G. Frank. What Is The Empirical Basis for Paying for Quality in Health Care? *Medical Care Research and Review* 63: 135-157.

Sarma, S., Devlin, R.A., Belhadji, B. and A. Thind. 2010. Does the Way Physicians are Paid Influence the Way They Practice? The Case of Canadian Family Physicians' Work Activity. *Health Policy* 98: 203-217.

Woodward, R.S. and F. Warren-Boulton. 1984. Considering the Effects of Financial Incentives and Professional Ethics on 'Appropriate' Medical Care. *Health Economics* 3: 223-237.

FIGURES

Figure 1: Hospital Length of Stay (LOS in the Experiment) across P4P Mechanisms. Each line represents the Gaussian kernel density of the LOS by payment mechanism in Experiment 1. FFS: fee-for-service; BO: bonus payment; BU: bundled payment.



TABLES

Table 1. Treatments and Number of Subjects. Experiment 1 combines the electronic medical treatment (EMR) with the fee-for-service (FFS), bonus (BO) and bundled (BU) payment mechanisms. Experiment 2 adds the clinical decision support system (CDSS) to the EMR and combines them with the fee-for-service (FFS), bonus (BO) and bundled (BU) payment mechanisms.

	Fee-for-Service (FFS)	Bonus Payment (BO)	Bundled Payment (BU)
Experiment 1: EMR	24	36	40
Experiment 2: EMR+CDSS	23	43	43

Table 2: Descriptive Statistics by Treatment: Experiment 1 uses the electronic medical record (EMR) information combined with the three payment mechanisms: fee-for-service (FFS), bonus (BO) and bundled (BU). Experiment 2 uses the EMR and clinical decision support software (CDSS) combined with the three payment mechanisms: FSS, BO and BU. Demographic information includes the percentage female, age and both their undergraduate grade point average (GPA) as well as their medical GPA, both on a 4.0 scale. Earnings are total earnings in U.S. dollars. Performance measures include the rate of readmission following a patient's discharge in the experiment and the patient's length of stay (LOS). Means for each variable are reported in the table with standard deviations reported in parentheses.

	Experiment 1			Experiment 2		
Treatment	FFS	BO	BU	FFS	BO	BU
# of Subjects	24	36	40	23	43	43
Demographics						
Percentage Female	0.458	0.389	0.575	0.652	0.581	0.488
	(0.509)	(0.494)	(0.506)	(0.487)	(0.499)	(0.506)
Age	26.917	26.278	26.200	26.739	26.186	33.279
	(2.225)	(1.717)	(1.488)	(3.828)	(2.510)	(4.000)
Undergraduate GPA	3.660	3.704	3.690	3.701	3.713	3.701
	(0.219)	(0.200)	(0.192)	(0.195)	(0.174)	(0.256)
Medical GPA	3.575	3.571	3.546	3.545	3.579	3.653
	(0.219)	(0.260)	(0.284)	(0.241)	(0.255)	(0.298)
Subjects' Earnings						
	137.500	149.194	150.588	139.130	160.895	167.523
	(8.341)	(24.425)	(19.352)	(7.635)	(17.315)	(16.230)
Performance						
Readmission Rate	0.087	0.120	0.110	0.077	0.093	0.099
	(0.283)	(0.326)	(0.313)	(0.267)	(0.290)	(0.299)
Patient LOS	4.111	2.993	3.099	2.960	2.667	2.078
	(3.403)	(2.885)	(2.671)	(2.447)	(1.973)	(1.545)

Table 3: Probit Regressions for Readmissions: Left three columns contain regression results (marginal effects) for Experiment 1. The right four columns contain regression results for Experiment 2. Additional variables in the probit regressions include: Female: binary indicator for female gender; Athlete: binary indicator for whether or not they were a student-athlete in college; Musical: binary indicator for playing a musical instrument; GPA: grade point average in medical school and in undergrad; Risk Attitudes Index: subjects risk index from post-experiment survey; Recommended LOS: The CDSS recommended LOS for the patient; Understay: number of days discharged before the CDSS recommended LOS; Overstay: number of days discharged after the CDSS recommended LOS; Target Probability: the target probability calculated by the CDSS. Robust p-value in parentheses*: *** p<0.01, ** p<0.05, * p<0.1*.

Treatments	Experiment 1 (EMR)			Experiment 2 (EMR+CDSS)			
BU	0.024 (0.173)	0.020 (0.235)	0.020 (0.228)	0.023 (0.144)	0.035** (0.033)	0.034** (0.035)	0.024 (0.120)
BO	0.035* (0.062)	0.028* (0.098)	0.030* (0.077)	0.017 (0.295)	0.025 (0.161)	0.024 (0.168)	0.022 (0.159)
Demographics							
Female		-0.010 (0.419)	-0.010 (0.396)		0.006 (0.557)	0.006 (0.589)	0.010 (0.323)
Athlete		-0.031** (0.020)	-0.031** (0.020)		-0.013 (0.326)	-0.013 (0.309)	-0.007 (0.560)
Musical		-0.027** (0.035)	-0.026** (0.040)		-0.007 (0.497)	-0.007 (0.519)	-0.004 (0.692)
Medical GPA		0.064*** (0.004)	0.065*** (0.003)		0.012 (0.485)	0.010 (0.575)	0.007 (0.697)
Undergraduate GPA		-0.002 (0.962)	-0.000 (0.993)		-0.033 (0.109)	-0.030 (0.127)	-0.027 (0.147)
Risk Attitude Index		-0.000 (0.975)	-0.000 (0.971)		-0.003 (0.389)	-0.002 (0.422)	-0.003 (0.360)
Other							
Recommended LOS							-0.009*** (0.000)
Understay							0.019*** (0.000)
Overstay							-0.011** (0.036)
Target Probability			0658*** (0.000)			0.662*** (0.000)	0.762*** (0.000)
Observation	2,742	2,742	2,742	3,039	2,985[a]	2,985[a]	2,985[a]
Log-Likelihood	-936.0	-928.9	-908.7	-930.8	-904.4	-878.9	-862.4

[a]In experiment 2, two subjects have missing information on (medical) GPA; a total of 54 discharge decisions on regular patients made by them were deleted in the last two models with controls of demographics.

Table 4: Censored-Normal Regressions for Patient Length of Stay: Left three column contains regression results for Experiment 1. The right three columns contain regression results aggregating the data from Experiment 2. Dependent variable is the LOS within the experiment. Additional variables in the regressions include: Female: binary indicator for female gender; Athlete: binary indicator for whether or not they were a student-athlete in college; Musical: binary indicator for playing a musical instrument; GPA: grade point average in medical school and in undergrad; Risk Attitudes Index: subjects risk index from post-experiment survey; Recommended LOS: The CDSS recommended LOS for the patient; Target Probability: the target probability calculated by the CDSS. Robust p-value in parentheses*: *** p<0.01, ** p<0.05, * p<0.1*.

Treatments		Experiment 1 (EMR)			Experiment 2 (EMR+CDSS)	
BU	-1.075**	-1.064**	-1.089**	-0.960***	-0.962***	-0.976***
	(0.014)	(0.013)	(0.011)	(0.003)	(0.005)	(0.004)
BO	-1.221**	-1.056**	-1.066**	-0.280	-0.234	-0.244
	(0.013)	(0.029)	(0.028)	(0.418)	(0.516)	(0.496)
Demographics						
Female		0.322	0.314		0.326	0.338*
		(0.354)	(0.365)		(0.110)	(0.100)
Athlete		-0.157	-0.161		0.341	0.351
		(0.696)	(0.690)		(0.117)	(0.109)
Musical		1.062***	1.068***		0.256	0.249
		(0.002)	(0.002)		(0.243)	(0.259)
Medical GPA		-0.804	-0.814		-0.126	-0.122
		(0.187)	(0.184)		(0.759)	(0.766)
Undergraduate GPA		-0.178	-0.217		0.275	0.274
		(0.832)	(0.798)		(0.601)	(0.601)
Risk Attitude Index		0.011	0.010		-0.006	-0.009
		(0.926)	(0.929)		(0.930)	(0.899)
Other						
Target Probability			7.235***			3.235***
			(0.000)			(0.000)
Start Date			-0.439***			-0.316***
			(0.000)			(0.000)
Constant	3.871***	6.674*	7.534**	2.821***	1.864	2.624
	(0.000)	(0.069)	(0.043)	(0.000)	(0.389)	(0.223)
Observations	2,742	2,742	2,742	3,093	2,985[a]	2,985[a]

[a]In experiment 2, two subjects have missing information on (medical) GPA; a total of 54 discharge decisions on regular patients made by them were deleted in the last two models with controls of demographics.

Appendix 1. EMR Facsimile Screens

Experiment Day Number: 1 | **Number of Possible Discharges Left: 30** | **Earnings: $0**

Patient ID: 18943412 | **Sex: Female** | Previous | Ready | Next
Name: Doe, Susan | **Age: 76**

Inpatient Summary

Visit Reason: PANCREATECTOMY

Patient Information

Ethnicity: White
Admission Type Urgent
Hospital: Emory
Readmit: No
Patient Death: No

Description

This is a 76 year old female admitted with obstructive jaundice secondary to chronic pancreatitis. The patient has undergone a Whipple procedure and is now 12-days postop.

Problems

Diabetes: No
Cancer: No
Hypertension: Yes
Alcohol: 0.0

Input & Output

	Stool Count
Day 12	0.0
Day 11	1.0
Day 10	0.0
Day 9	0.0
Day 8	0.0
Day 7	1.0
Day 6	1.0
Day 5	2.0
Day 4	2.0
Day 3	0.0
Day 2	0.0
Day 1	0.0

Figure A1: Inpatient Summary

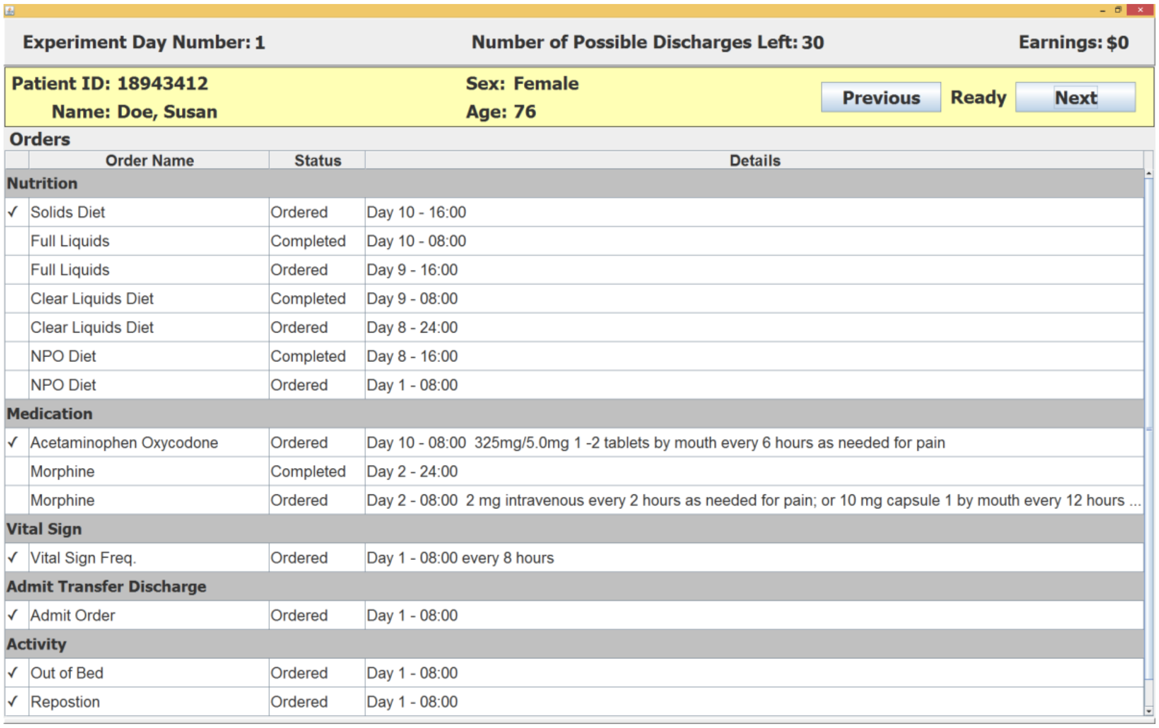
Experiment Day Number: 1 | **Number of Possible Discharges Left: 30** | **Earnings: $0**

Patient ID: 18943412 | **Sex: Female** | Previous | Ready | Next
Name: Doe, Susan | **Age: 76**

Lab

All Lab Results	Day 12 - 24...	Day 12 - 16...	Day 12 - 08...	Day 11 - 24...	Day 11 - 16...	Day 11 - 08...	Day 10 - 24...	Day 10 - 16...	Day 10 - 08...	Day 9 - 24:00
Sodium Level			129 mmol/L			133 mmol/L			131 mmol/L	
Potassium			4.3 mmol/L			4.4 mmol/L			4.2 mmol/L	
CO2			26 mmol/L			27 mmol/L			27 mmol/L	
BUN/Creat Ratio			11			16			13	
Creatinine			0.6 mg/dL			0.6 mg/dL			0.5 mg/dL	
Bilirubin Total			2.7			2.5			2.8	
AST (SGPT)			41 IU/L			43 IU/L			59 IU/L	
ALT (SGPT)			31 IU/L			29 IU/L			38 IU/L	
Alkaline Phosphatase			238 IU/L			208 IU/L			237 IU/L	
Albumin Level			1.9 mg/dL			1.7 mg/dL			1.5 mg/dL	
Calcium			8.2 mmol/L			8.2 mmol/L			7.7 mmol/L	
White Blood Count			27.1 10E 3/...			25.1 10E 3/...			30.8 10E 3/...	
Hematocrit			25			20.7			21.7	
Plateletes			525			455			473	
PT/INR										
PTT										

Figure A2: Laboratory Data



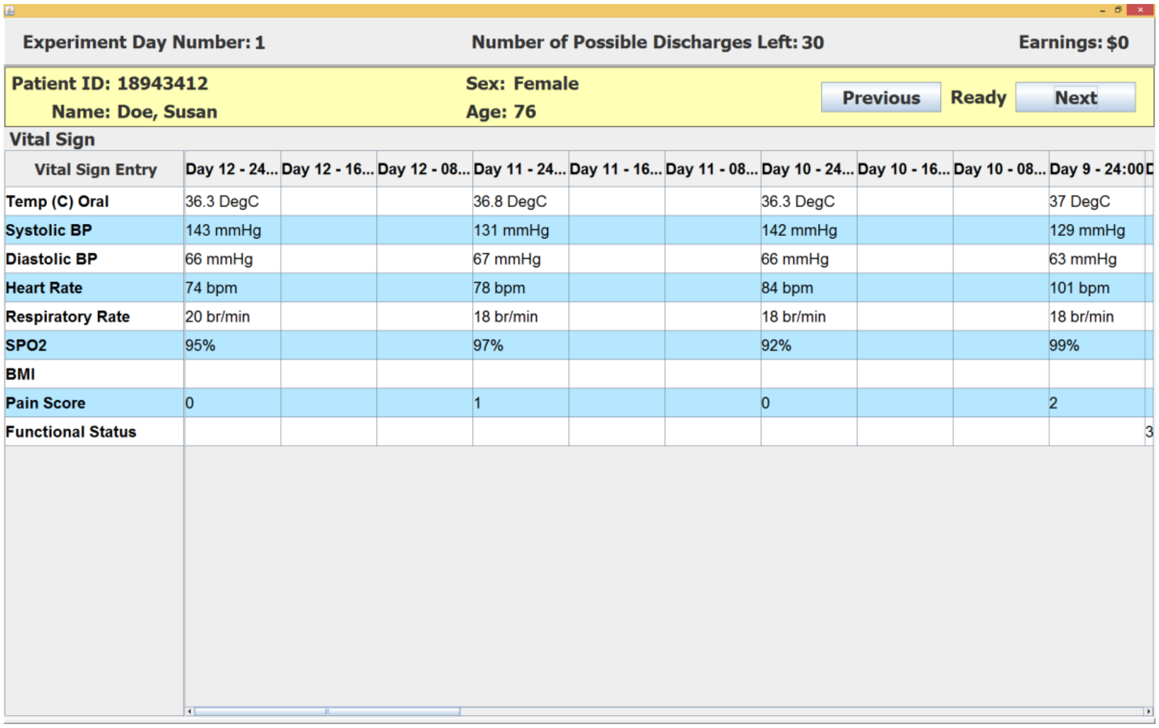
Experiment Day Number: 1 | **Number of Possible Discharges Left: 30** | **Earnings: $0**

Patient ID: 18943412 **Sex: Female**
Name: Doe, Susan **Age: 76**
Previous Ready Next

Orders

Order Name	Status	Details
Nutrition		
✓ Solids Diet	Ordered	Day 10 - 16:00
Full Liquids	Completed	Day 10 - 08:00
Full Liquids	Ordered	Day 9 - 16:00
Clear Liquids Diet	Completed	Day 9 - 08:00
Clear Liquids Diet	Ordered	Day 8 - 24:00
NPO Diet	Completed	Day 8 - 16:00
NPO Diet	Ordered	Day 1 - 08:00
Medication		
✓ Acetaminophen Oxycodone	Ordered	Day 10 - 08:00 325mg/5.0mg 1 -2 tablets by mouth every 6 hours as needed for pain
Morphine	Completed	Day 2 - 24:00
Morphine	Ordered	Day 2 - 08:00 2 mg intravenous every 2 hours as needed for pain; or 10 mg capsule 1 by mouth every 12 hours ...
Vital Sign		
✓ Vital Sign Freq.	Ordered	Day 1 - 08:00 every 8 hours
Admit Transfer Discharge		
✓ Admit Order	Ordered	Day 1 - 08:00
Activity		
✓ Out of Bed	Ordered	Day 1 - 08:00
✓ Repostion	Ordered	Day 1 - 08:00

Figure A3: Orders



Experiment Day Number: 1 | **Number of Possible Discharges Left: 30** | **Earnings: $0**

Patient ID: 18943412 **Sex: Female**
Name: Doe, Susan **Age: 76**
Previous Ready Next

Vital Sign

Vital Sign Entry	Day 12 - 24...	Day 12 - 16...	Day 12 - 08...	Day 11 - 24...	Day 11 - 16...	Day 11 - 08...	Day 10 - 24...	Day 10 - 16...	Day 10 - 08...	Day 9 - 24:00	D
Temp (C) Oral	36.3 DegC			36.8 DegC			36.3 DegC			37 DegC	
Systolic BP	143 mmHg			131 mmHg			142 mmHg			129 mmHg	
Diastolic BP	66 mmHg			67 mmHg			66 mmHg			63 mmHg	
Heart Rate	74 bpm			78 bpm			84 bpm			101 bpm	
Respiratory Rate	20 br/min			18 br/min			18 br/min			18 br/min	
SPO2	95%			97%			92%			99%	
BMI											
Pain Score	0			1			0			2	
Functional Status											3

Figure A4: Vital Signs