THE EFFECTS OF PAY-FOR-PERFORMANCE ON PHYSICIANS’ TREATMENT OF PATIENTS

A Thesis
Submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Master of Public Policy

By

Kristen Michelle Gorzelany, B.A.

Washington, D.C.
April 15, 2010
THE EFFECTS OF PAY-FOR-PERFORMANCE ON PHYSICIANS’ TREATMENT OF PATIENTS

Kristen Michelle Gorzelany, B.A.

Thesis Advisor: John M. Nail

ABSTRACT

As rising healthcare costs continue to garner national attention, pay-for-performance has been one strategy put forth to counteract the increasing costs and to improve the quality and effectiveness of healthcare. One of the concerns that has accompanied the rise of pay-for-performance in the healthcare context is that it could incentivize physicians to avoid the sickest, most complicated patients. Controlling for other measurable variables that affect patient-health outcomes, this thesis examined whether or not pay-for-performance affected physicians’ treatment of patients. There was no comprehensive data set addressing physician participation in pay-for-performance, but the best available data set that I found was the 2007 National Ambulatory Medical Care Survey, which I used to examine the effect of physicians’ participation in pay-for-performance on their referrals of patients to other physicians. Controlling for patient age, tobacco use, and chronic illness, the regressions actually showed that physicians participating in pay-for-performance programs were less likely to refer their patients to other physicians than were physicians not participating in pay-for-performance programs. This unexpected result may be due to the lack of granular data on pay-for-performance programs (i.e., that there was no differentiation in the data between physicians participating in pay-for-performance as measured by adherence to clinical guidelines and physicians participating in pay-for-performance as measured by patient outcomes) and to the lack of information on the reasons for physician referrals to other physicians. This thesis also found that chronically ill patients were less likely to be seeing physicians who participated in pay-for-performance programs.
Additional research, which will necessitate more comprehensive data-collection, is necessary before policymakers can credibly advocate or discredit pay-for-performance as a main, cost-saving component of healthcare reform.
# Table of Contents

Introduction ........................................................................................................................................... 1

Chapter I: Pay-for-performance: the Larger Concept ........................................................................... 3

Chapter II: The U.S. Healthcare Industry: A Brief Overview ................................................................. 5

Chapter III: Pay-for-performance in the Healthcare Industry ............................................................ 10

Chapter IV: Data .................................................................................................................................... 19

Chapter V: Methods of Analysis .......................................................................................................... 24

Chapter VI: Results ............................................................................................................................. 26

Chapter VII: Conclusion ..................................................................................................................... 36

References ........................................................................................................................................... 39
INTRODUCTION

This thesis tested the hypothesis that physicians participating in pay-for-performance programs were more likely to avoid treating sicker, more complicated patients than physicians not participating in such programs.¹ Pay-for-performance programs link physicians’ pay with how well they perform, usually as measured by their adherence to clinical guidelines in treating patients, but sometimes as measured by patient-outcome (i.e., by patients’ mortality and morbidity).

As healthcare-reform advocates often point out, the United States spends more per capita on healthcare than does any other country, yet its citizens are no healthier for it.² Policymakers desperate to stem ballooning healthcare costs, which have not been accompanied by increases in health, have espoused pay-for-performance not because it has been proven to be effective³ but because the philosophy undergirding it is so appealing: increase pay for desired results while withholding pay for treatment that doesn’t empirically improve health commensurate with its cost. Pay-for-performance would seem to improve efficiency and quality, increasing health while decreasing expenditures. And pay-for-performance can also provide cover for policymakers who wish to raise provider-payments in an environment hostile to increasing

¹ It was through my work as a research assistant for Professor Kathryn Zeiler at the Georgetown University Law Center over the summer of 2009 that I became familiar with the existing literature on pay-for-performance, and it was Professor Zeiler who introduced me to the idea that pay-for-performance (as measured by patient-outcome) could have the collateral effect of incentivizing physicians to cherry-pick patients. I resolved to test her hypothesis, as best I could with the data in existence, when I began my thesis in the fall of 2009.

² See World Health Organization, World Health Statistics 2009 at 107 (2009) (“Globally in 2006, expenditure on health was about 8.7% of gross domestic product, with the highest level in the Americas at 12.8%.”); K. Davis et al., The Commonwealth Fund, 59, “Mirror, Mirror on the Wall: An International Update on the Comparative Performance of American Health Care” (Deborah Lorber ed., May 15, 2007) (“[C]omparative analyses consistently show the United States underperforms relative to other countries on most dimensions of performance [including quality, access, efficiency, equity, and healthy lives].”).

healthcare costs: they can easily frame pay-for-performance as a means by which to improve health outcomes, rather than as another naked price increase. As this thesis explains, however, there are complexities to pay-for-performance—both to its design and its implementation—that render its philosophy and its reality incongruent.

Chapter I provides background on pay-for-performance generally. Chapter II provides an overview of some of the problems in the U.S. healthcare industry, including the fee-for-service (FFS) payment structure. Chapter III discusses pay-for-performance in the healthcare context, namely how it has been put forth as a potential solution to the problems inherent in FFS. Chapter IV provides information on the data used to test the hypothesis that physicians participating in pay-for-performance programs are more likely to avoid treating sicker, more complicated patients. Chapter V discusses the methods of analysis used to test this hypothesis. Chapter VI lays out the results of the analyses and discusses some policy implications, and Chapter VII presents conclusions.
CHAPTER I: PAY-FOR-PERFORMANCE: THE LARGER CONCEPT

Pay-for-performance has long appealed to employers and employees alike: “What better way to drive people to work harder and more efficiently . . . than to offer them a special carrot: more money for hitting specific [] targets?” But while performance standards can help make clear the expectations, measurements, and rewards—and while some employees see pay-for-performance or “incentive pay” as an opportunity to increase their earnings—many employers simply don’t consider the implementation costs and downsides of such programs. For example, some employees resent such programs, seeing pay as an entitlement or something based upon tenure, while others’ self-esteem suffers from the implicit gap between their current performance and ideal-employee performance. Although pay-for-performance programs can enhance competition among co-workers, it can do so at the expense of teamwork and creativity. Such programs can also lead employees to perform to the test (i.e., to focus all of their attention on the aspects of performance being rewarded, at the expense of beneficial practices that hadn’t or couldn’t be quantified). Effective implementation of a pay-for-performance program might also result in additional paperwork, redirecting employees’ time and attention from their substantive work.

As Harvard Business School Professor Michael Beer discovered in his analysis of failed pay-for-performance programs, conceiving of, designing, and implementing such programs can

---


5 Id. In addition, managers overseeing pay-for-performance programs felt the programs didn’t “motivate employees to work harder or, perhaps more importantly, to learn.” Id. See also Gregorio Billikopf, “Incentive Pay (pay for performance),” University of California, in Labor Management in AGRICULTURE: CULTIVATING PERSONNEL PRODUCTIVITY (2003) at 94, available at http://www.cnr.berkeley.edu/ucce50/ag-labor/7labor/08.htm#o_1 (“[I]ncentive pay is what [employers] pay when they do not want to pay workers a fair wage”).

6 Lagace, supra n.4 (internal citation omitted).

7 See id.
be “rocky.” To design an effective pay-for-performance program, the problem the employer seeks to address must be amenable to performance-payments; the employer needs to set clear performance goals and decide how much of an employee’s pay will depend upon meeting those goals; the employer needs to have effective adjustment mechanisms in place to protect employees from factors that are beyond their control but can nevertheless affect whether or not they reach their performance goals; and there needs to be a feedback mechanism so that such programs can continually be refined in furtherance of their goals.

8 Id. With regard to designing pay-for-programs in healthcare, the American College of Physicians worries that “the design of pay-for-performance systems will lead to worse care despite measurements that imply good care (‘the patient died, but the electrolytes were in balance’).” L. Snyder & R. L. Neubauer, for the American College of Physicians Ethics, Professionalism and Human Rights Committee, “Pay-for-performance Principles That Promote Patient-Centered Care: An Ethics Manifesto” 147 Ann. Internal Med. 792, 792 (December 4, 2007).

9 See Lagace, supra n.4. Prof. Beer noted that participants become frustrated with factors beyond their control that nevertheless affected their performance. See also Billikopf, supra n.5 at 97 (“Any time employees are rewarded or punished for that which they cannot control, farm employers are asking for a cynical or disillusioned workforce…. [It’s much better to] break[] down all elements under the control of employees or management that affect profits and reward[] personnel for achieving results.”).

10 See Billikopf, supra n.5 at 95.
CHAPTER II: THE U.S. HEALTHCARE INDUSTRY: A BRIEF OVERVIEW

The traditional payment method in the U.S. healthcare industry has for years been fee-for-service (FFS), in which physicians are paid by the quantity, not the quality or effectiveness, of the healthcare services they provide. Under FFS, “where there is uncertainty about the appropriate level of care, physicians have a financial incentive to overuse care.”11 “[F]ee-for-service payments encourage [physicians] to do as much as possible, whether it helps patients or not.”12 As such, FFS has long been blamed for unnecessary care, which comes at unnecessary cost. Physicians “rack[ ] up charges with extra tests, services, and procedures.”13 Where they need to make judgment calls between conservative treatment on the one hand and surgery on the other, FFS gives them an added incentive to operate.14

Additionally, standards of practice may differ by geographic area, whether because of training or something else. To compare the differences in standards, one researcher compared how physicians in cities known to have high healthcare costs fared against physicians practicing in cities known to have low healthcare costs.15 The researcher asked physicians in both types of cities how they would treat “a seventy-five-year-old woman with typical heartburn symptoms and ‘adequate health insurance to cover tests and medications,’” and she found that although the physicians were equally likely to follow the steps strongly recommended by national guidelines, “when it came to measures of less certain value—and higher cost—the differences were

---

11 D.M. Eddy, “Balancing Cost And Quality In Fee-For-Service Versus Managed Care,” 16 Health Affairs 162 (1997).
12 Doran, supra n.3, at 70.
14 Id. For example, though patients with pain from gallstones infrequently suffer complications and can manage the pain with medication and prevent recurrence with diet-modification, physicians choose to operate: “[B]y operating they happen to make an extra seven hundred dollars.” Id. The blame for such practices lies at the feet of “a few leaders of local institutions [who] took profit growth to be a legitimate ethic in the practice of medicine.” Id.
15 Id.
considerable.” Physicians in high-cost cities were much more likely to refer the patient to a gastroenterologist, order an upper endoscopy, or both, and “typically recommended that patients with well-controlled hypertension see them in the office every one to three months,” while the physicians in low-cost cities were much less likely to refer and order expensive procedures, and typically recommended once every six months. “Physicians from the most expensive cities did the most expensive things.” Interestingly, standards in medical malpractice suits are based upon the prevailing practice in the field. These standards affect and are affected by the “clinical guidelines,” which are in turn used by pay-for-performance programs.

Asymmetric information exacerbates the overconsumption of healthcare. Patients often don’t know which treatments they need and which are unnecessary. They trust their physicians to be their perfect agents, deciding for them the appropriate course of treatment as if the physician were the patient. But as mentioned earlier, physicians being paid by FFS and subject to high medical-malpractice-insurance premiums are being incentivized to provide services, both to maximize their own income and to minimize their exposure to potential lawsuits. Thus, the cost-benefit analysis in which a treating physician engages necessarily puts too little weight on patient benefit (health) and too much on physician benefit (profit).

In addition to physician incentives to over-provide healthcare, health insurance encourages patients to over-consume it. Because patients pay health-insurance premiums up front and then pay only a fraction of the cost of the healthcare they consume thereafter, insured patients consume additional and/or more expensive healthcare than they would if they were paying the full amount out-of-pocket, which results in an inefficient allocation of resources. An uninsured

\[16 \text{ Id.} \]
\[17 \text{ Id.} \]
\[18 \text{ Id.} \]
patient weighs the *true* costs and benefits of healthcare consumption—whether or not to visit the doctor for a cold, or whether or not to purchase a branded prescription versus generic prescription medication—but insurance acts as a sort of thumb on the scale toward consumption. Insurance companies, by paying such a significant a portion of an insured’s healthcare bills, totally skews the insured patient’s calculation of which healthcare services are “worth it.”

<table>
<thead>
<tr>
<th>Patient’s Immediate, Out-of-pocket Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>With insurance:</td>
</tr>
<tr>
<td>Doctor’s Visit</td>
</tr>
<tr>
<td>$20 copay</td>
</tr>
<tr>
<td>Without insurance:</td>
</tr>
<tr>
<td>$250</td>
</tr>
</tbody>
</table>

Since branded and generic medications contain the same active ingredients at the same strength and in the same dosage, the choice between the two (especially considering the extreme price difference) would seem clear, but “insurance executives are the only people who see the full cost of the drugs. Patients don't know or care, because the majority of patients have health insurance,” and thus pay only small copays, regardless of the actual cost of the drug. Insured patients don’t face the same financial incentive to choose the generic drug that uninsured patients do.

---

19 Note that differences in additives and fillers between branded and generic drugs are allowed and that perfect bioequivalency is not required. Some therefore dispute claims that branded and generic drugs are truly interchangeable.


21 An uninsured patient choosing a branded drug pays 79% more ($405 in absolute dollars), while an insured patient pays 33% more ($10 in absolute dollars).
In addition to FFS, asymmetric information, and insurance, the risk of facing medical-malpractice claims has also led to an increase in “defensive medicine,” the overprovision of healthcare that occurs when providers act out of an abundance of caution to avoid potential lawsuits.\textsuperscript{22} All of these things have contributed to the overprovision of healthcare, which ironically exposes patients to additional and unnecessary risk.

One might argue that, if patients pay low costs and incur some benefit, and physicians incur a benefit from the transaction, then both are better off engaging in the transaction, but that disregards the long-term and larger society-wide effects of their actions. The overconsumption and/or consumption of more \textit{expensive} healthcare might seem innocuous—“Americans like to believe that, with most things, more is better”\textsuperscript{23}—but there is an attendant risk of side effects for individuals and of premium increases for the general public. “Research suggests that where medicine is concerned [more] may actually be worse.”\textsuperscript{24} For example, with regard to a patient whose hyperlipidemia could be effectively managed with diet-modification and exercise, the choice to treat the condition with a cholesterol-lowering medication instead introduces a risk of adverse events (e.g., side effects and drug-interactions). With regard to his insurance company, while the patient pays a small copay for the medication, the insurance company pays the rest, and

\textsuperscript{22} See Gawande, supra n.13 (quoting physicians who blame malpractice for increasing healthcare costs: “Doctors order unnecessary tests just to protect themselves.”).

\textsuperscript{23} Id.

\textsuperscript{24} Id. (“[N]othing in medicine is without risks. Complications can arise from hospital stays, medications, procedures, and tests, and when these things are of marginal value the harm can be greater than the benefits.”

A 2003 study of the “treatment received by a million elderly Americans diagnosed with colon or rectal cancer, a hip fracture, or a heart attack…found that patients in higher-spending regions received sixty per cent more care than elsewhere. They got more frequent tests and procedures, more visits with specialists, and more frequent admission to hospitals. Yet they did no better than other patients, whether this was measured in terms of survival, their ability to function, or satisfaction with the care they received. If anything, they seemed to do worse.” Id.
that insurance company will pass its increased costs on to insurance purchasers in the form of premium- and copay-increases.\textsuperscript{25}

The rationale behind the adoption of pay-for-performance programs is that it more directly aligns physicians’ and patients’ incentives with those of society, discouraging the provision and consumption of unnecessary healthcare while assuring that the care that \textit{is} provided is of high quality.

\textsuperscript{25} Joffe-Walt, \textit{supra} n. 20. With regard to passing on costs, a vice president of an insurance company confirmed that, as its costs rise, it has to raise premiums. \textit{See id.}
CHAPTER III: PAY-FOR-PERFORMANCE IN THE HEALTHCARE CONTEXT

“When performance measurement is combined with financial incentives to bring about clinician and systems change, the result is pay-for-performance programs.” Pay-for-performance somewhat recently arose as a “solution” to excessive, unnecessary care. There are now over one hundred active pay-for-performance programs in the U.S. More than half of the private health plan contracts in the U.S. include pay-for-performance, and the government has also moved toward adoption of pay-for-performance in Medicare and Medicaid. Yet, claims regarding pay-for-performance are varied and have varying levels of empirical support, which is not surprising given the variation in pay-for-performance programs themselves.

There are two main types of pay-for-performance programs in the United States healthcare market. One focuses on physician adherence to established clinical guidelines in treating patients. The other focuses on patients’ ultimate health-outcomes (i.e., morbidity and mortality). There are significant problems with designing and implementing both types of pay-for-performance. As noted in the INTRODUCTION, supra page 4, the critical first step in designing a successful pay-for-performance program is that the problem one seeks to address must actually be amenable to incentive pay. This means that physicians have to be able to meet the

26 Snyder & Neubauer, supra n.8 at 792.
27 M. B. Rosenthal, Testimony before the House Subcomm. on Employer-Employee Relations, Hearing on Examining Pay-for-performance Measures and Other Trends in Employer-Sponsored Health Care (May 17, 2005) (“The most recent estimates suggest that there are more than 100 individual pay-for-performance efforts underway in the U.S. health care sector.”).
28 See Rand, Compare, Policy Options, Pay-for-performance Overview, http://www.randcompare.org/policy-options/physician-pay-for-performance (internal citations omitted). See also M.B. Rosenthal et al., “Pay for Performance in Commercial HMOs,” 335 New England J. Med. 1895 (2006) (Surveying 252 HMOs across the country and finding that in 2005, 52.1% of health plans (81.3% of persons enrolled in HMOs) were participating in pay-for-performance.) Rosenthal et al. found that participation in pay-for-performance was statistically significantly associated with geographic region (programs less likely to be in existence in the South), use of primary care providers as gatekeepers (positively associated with pay for performance), use of capitation to pay providers, (positively associated with pay for performance) and whether the plans received bonuses or penalties according to performance (also positively associated with pay for performance).
performance goals. While physicians should be capable of adhering to clinical guidelines, though the relationship between guideline-adherence and quality of care is putative. With regard to physicians’ capacity to bring about positive patient-health outcomes, physicians cannot control patient outcomes to the extent that pay-for-performance programs by patient outcome would seem to suggest. There are too many other factors affecting patient health that are beyond the physician’s control. Thus, the validity of the quality measures and the centrality of risk-adjustment have been two main sticking points in this second type of pay-for-performance.

One of the main problems with designing and implementing a pay-for-performance program by patient outcome is that patients’ poor outcomes may be due mostly to the patient’s baseline health, access to care, and/or noncompliance with physicians’ instructions, none of which is within the physician’s control. Yet physicians participating in pay-for-performance programs will suffer for those outcomes even when they have done everything possible to help the patient reach a positive outcome.

Thus, as it turns out, physicians’ interest in their own financial wellbeing stands to affect their treatment decisions under both FFS and pay-for-performance by patient outcome. Under pay-for-performance by patient outcome, there is just as strong of an incentive for physicians to avoid sick patients as there was for them to over-treat patients under FFS. Under FFS, the risk is that physicians might over-treat to earn more money, while under pay-for-performance, the risk is that they might avoid sicker, more complicated patients to earn more money or to avoid being deprived of additional income.

This is why there is a risk that, unless physicians are “firmly convinced” that risk adjustment is sufficient, they could decide that the easiest way to achieve high scores is to avoid
sick or challenging patients.\textsuperscript{29} This is also why risk-adjustment is essential to appropriately implementing pay-for-performance programs by patient outcome. Without accurate risk-adjustment, two identically qualified physicians exercising identically sound judgment would nevertheless be found to have “performed” at much different levels under a pay-for-performance program by patient outcome. For example, a physician serving patients who were predominantly older, sedentary, poor, and non-compliant would be at a stark disadvantage relative to a physician serving predominantly young, active, affluent, and compliant patients. Accurate risk-adjustment would eliminate the incentive for physicians to avoid sick patients by leveling the playing field so that outcomes would be considered \textit{in light of} the other factors affecting patient outcomes that are out of the physician’s control.

Without adequate risk-adjustment, “P4P and public reporting might induce individual physicians and medical groups to avoid patients whom they perceive as being likely to lower their quality scores, particularly if quality measures are not adequately adjusted for the patients’ overall health status . . . .”\textsuperscript{30} The United Kingdom, which pays its physicians in accordance with their performance, has made allowances to counteract such perverse incentives: physicians can “flag very sick or noncompliant patients to be removed from their patient base for reporting purposes,” so “those patients remain in the practice...but don’t affect the physician’s performance profile.”\textsuperscript{31} And in fact, analyses of the UK’s national pay-for-performance program have found that physicians’ ability to exclude some patients from their reporting “was the strongest predictor of physician performance.”\textsuperscript{32} Thus, the concern that “physicians, hospitals and other providers

\begin{itemize}
\item \textsuperscript{30}L.P. Casalino et al., “Will Pay-For-Performance And Quality Reporting Affect Health Care Disparities?” 26 Health Affairs 405 (2007).
\item \textsuperscript{31}Id. \textit{See also} T. Doran et al., “Exclusion of Patients from Pay-for-performance Targets by English Physicians;” 359 New Eng. J. Med. 274 (2008).
\item \textsuperscript{32}S.J. Tanenbaum \textit{supra} n.3.
\end{itemize}
may ‘cherry pick’ healthy patients and avoid the sickest and most complex patients in order to improve quality scores” is well grounded.

Others, including the American College of Physicians Ethics, have also expressed the concern that physicians participating in pay-for-performance programs by patient outcome will avoid sicker, more complicated patients. Pay-for-performance programs that pay physicians for positive patient outcomes can “result in the deselection of patients, ‘playing to the measures.’” Pay-for-performance programs by patient outcome “encourage physicians to improve their performance scores by dropping (or refusing to accept) difficult patients whose outcome measures . . . do not meet the quality standard and therefore worsen the physician’s profile.”

Nine out of the thirty-five physician organizations involved in statewide pay-for-performance program in California who were interviewed “gave examples of negative consequences associated with the pay-for-performance program, such as dropping noncompliant patients and diverting resources to activities with no health benefit.”

But it is not just direct financial incentives that affect physicians’ treatment of patients. “[P]ublic reporting of providers’ performance also appears to yield less desirable supply responses, including avoidance of patients perceived to be high-risk.” Cardiac surgeons whose performance measures were made public altered their practices in response to that reporting. Specifically, sixty-two percent reported that they’d refused to operate on at least one high-risk patient.

34 See Snyder & Neubauer, supra n.8.
37 Rosenthal, supra n.35.
coronary-artery bypass patient primarily because of public reporting, which indicates that physicians might avoid treating sicker patients even when the effect on physicians is not directly financial.\(^{38}\)

Beyond the type of performance incentivized (i.e., adherence to guidelines or patient outcomes) pay-for-performance programs differ by whether they aggregate performance data from all physicians within a certain institution (e.g., a hospital or practice) or operate at an individual-physician level. While individual physicians’ performance scores are more vulnerable to the variations in patient populations, an aggregate score for a physician group diffuses the risk of treating baseline-unhealthy and noncompliant patients. The flipside is that the financial incentive to meet the performance targets is also more potent for individually scored physicians. In a physician-group-scoring setting, individual physicians may be more inclined to free-ride on their colleagues’ efforts. Regardless, the adverse effects of pay-for-performance by patient outcome seem to apply to both hospital- and physician-level pay-for-performance programs: “[B]oth physicians and hospitals have been found to attempt to select healthier patients under prospective payment to maximize net revenues.”\(^{39}\) And nonprofit providers of substance-abuse treatment also selected less-severely-ill clients where a “performance-based contract” system was in place.\(^ {40}\)

Increased referrals to specialists may be another unanticipated effect of quality-measurement approaches like pay-for-performance,\(^ {41}\) and referrals to other physicians or other

\(^{38}\) See Burack et al., “Public Reporting of Surgical Morality: A Survey of New York State Cardiothoracic Surgeons,” 68 Annals Thoracic Surgery 1195-1202 (1999) (finding that most of the cardiac surgeons who responded to an anonymous mail survey surgeons (62%) refused to operate on at least one high-risk CAB patient over the prior year, primarily because of public reporting.).

\(^{39}\) Rosenthal Testimony, supra n.27.


healthcare establishments could be a legitimate measure of performance and a means of quality improvement. One article, however, voiced more specific concerns about patient transfers: “NICUs might transfer their highest-risk patients to other hospitals.” Thus, the concern that physicians participating in pay-for-performance by patient outcome might refer sicker, more complicated patients out to other physicians is not a new one. However, considering that a 2007 study showed each Medicare beneficiary had, on average, two primary care physicians and five specialists, the likelihood of pay-for-performance systems “accurately attribut[ing] responsibility for the outcome of care for such patients” is low. In such an environment, passing off problem patients via referrals may be confused with the simple realities of the modern practice of medicine, wherein referrals may reflect nothing more than common—perhaps even appropriate—practice.

Beckman et al. (2006) studied how physicians reacted to the introduction and evolution of a pay-for-performance program, taping thirty primary-care physicians in focus groups at thirteen and twenty-six months after beginning the program and analyzing their comments. The physicians did express concern about patients’ effects on their pay-for-performance scores and about the influence pay-for-performance programs might have on “professionalism,” which reflect the same concerns Professor Beer identified in his study of failed pay-for-performance

42 See J. Carter et al., “Adherence to Burn Center Referral Criteria: Are Patients Appropriately Being Referred?” 31 J. Burn Care & Res. 26 (Jan./Feb. 2010).
44 See Snyder & Neubauer, supra n.8, at 793 (internal citation omitted).
45 See id.
47 Id.
programs. Other researchers have also found that physicians tend to believe that many factors affect patient health and the quality of healthcare, including insurance, patient noncompliance, lack of time, and improper physician oversight. Nevertheless, physicians largely support increased pay for delivering “high-quality” care, despite that they questioned the measurements’ accuracy.  

Whether the magnitude of the financial incentives offered by pay-for-performance programs has an effect on physicians’ behavior is unclear. Studies have shown that even a small change in payment structure can have a significant effect on patient behavior, but it’s not clear that a change in physicians’ compensation structure would have an analogous significant effect on physician behavior. Some assert that “as performance incentives constitute a larger fraction of income…this will increase the likelihood for negative consequences such as gaming the data, ignoring other aspects of quality, or refusing to treat complex patients.” But in recent years, U.S. pay-for-performance programs tied an average of only five percent of a physician’s salary to performance.  

---

48 See supra pp. 3–4.  
50 Id.  
51 One study found that “incentives equivalent to 5% of total income would be required in order to influence physician’s activities,” though “some of the more successful schemes have rewards of up to [$7,500] per physician per year, [while] others have been successful offering incentives as modest as [$1.60] per patient for influenza immunizations.” Doran, supra n.3, at 73 (internal citations omitted). “Probably the most extravagant financial incentives are to be found in the UK’s QOF scheme, which offered maximum rewards in excess of £25000 [$50,000] per physician in its first year.” Id. (internal citations omitted))  
52 See Joffe-Walt, supra n.20, citing WSJ article by Rockoff. Copay increases of ten percent have been followed by drug-spending decreases of six percent.  
53 Damberg, supra n.36.  
54 Rosenthal Testimony, supra n.27 (“It is also noteworthy that among pay-for-performance programs in the U.S., few payers put at risk more than 5 percent of payments. Moreover, because of the small market shares of some pay-for-performance program sponsors, the percent of a physician’s overall revenue that is at stake can be much less than 5 percent.”).
Thus, while there has been much discussion about the theoretical effects of pay-for-performance programs, there has been very little research into what physicians actually do with sicker, more complicated patients when they participate in pay-for-performance programs. This thesis looked at referrals to other physicians as a function of other patient-health indicators and whether or not the physician is participating in a pay-for-performance program.

Previous research into pay-for-performance programs has faced severe data limitations, and I faced the same limitations in my research. Collecting granular data on patients and physicians are prohibitive. Data-collection itself is time-consuming, electronic record-keeping is not yet universal, and concerns about patient privacy and physician autonomy stifle the establishment of an open and interconnected database of physician-patient information. Patients move, change health-insurers, and see new physicians; there is no central repository of any one patient’s health records; and it is virtually impossible to tease apart the effects any one physician or visit has on a patient’s health. Physicians move, start solo practices, join hospitals, and decide which health-insurance they will accept. There is no central system through which a researcher could follow a physician from one practice to another, analyzing his patient interactions to see why patients had come in, what the disposition of the visit was and why, and whether or not the physician’s participation in pay-for-performance programs affected his or her treatment decisions. Following a group of patients and their physicians over a period of time during which pay-for-performance was implemented so as to monitor the changes the program brought about would be ideal, but to my knowledge, there has never been such an occurrence that was monitored by an agency willing to make its data publicly available.\footnote{The Veterans’ Administration has implemented pay-for-performance and has a relatively closed set of patients and physicians, but its data has not been made publicly available.}
Thus, given the data that was publicly available, I attempted to test the effects of physician participation in pay-for-performance on physicians’ treatment of patients, as measured by referrals to other physicians. The major data-limitations I faced included a large number of “blank” and “missing” values on variables important to my analysis, a lack of information on the type of pay-for-performance program in which physicians were participating as well as whether or not physicians has a choice regarding their participation in the program, and a lack of reasons for physicians’ referrals. To combat these limitations, I dropped the observations that lacked information on the variables that were important to the regressions I ran, and I created an indicator variable for participation in pay-for-performance programs so that if I found any statistically significant positive correlation between physician participation in pay-for-performance programs and referrals to other physicians, I could infer that the actual effect of physician participation in a pay-for-performance program by patient outcome would have been even more significant.
CHAPTER IV: DATA

The data set with which I worked was the 2007 National Ambulatory Medical Care Survey (NAMCS). NAMCS was conducted by the Division of Health Care Statistics, National Center for Health Statistics, Centers for Disease Control and Prevention. The data-collection instrument was a survey, and the population sample included 3,540 randomly selected, non-federally employed office-based\(^{56}\) physicians\(^{57}\) who provided a total of 32,778 Patient Record Forms (PRFs)—data on approximately thirty patient visits per physician, over a randomly selected one-week period.

However, 1,141 of the physicians sampled did not meet all of the criteria and were thus ruled ineligible for the study. Of the 2,399 eligible physicians, only 1,568 participated. Only 1,357 completed PRFs, and only 1,266 participated “fully or adequately,”\(^{58}\) while ninety-one participated minimally.\(^{59}\)

There were 431 variables included in the data set, including the following variables, which were of particular relevance to testing my hypothesis and which I’ve therefore included in my regressions:

- **age (patient age):** This variable ascertains a patient’s current age and is relevant because I assume that as patients age, they suffer from more medical problems, which would be an incentive for a physician whose income is tied to his or her performance to avoid the patient, or refer the patient out to another physician.

---

\(^{56}\) Classified by the American Medical Association (AMA) or the American Osteopathic Association (AOA) as "office-based, patient care" - not specialties of anesthesiology, pathology, and radiology. Community health centers included.

\(^{57}\) 3,301 Medical Doctors; 239 Doctors of Osteopathy.

\(^{58}\) Full/adequate participation = submitting at least half of the expected PRFs, based on the total number of visits during the reporting week.

\(^{59}\) Minimal participation = submitting fewer than half of the expected PRFs.
- **usetobac (tobacco use):** This variable ascertains patients’ current tobacco use and is relevant because I assume that current tobacco use is associated with more medical problems, which would be an incentive for a physician whose income is tied to his or her performance to avoid the patient, or refer the patient out to another physician.

- **totchron (total number of chronic conditions (possible range of 0-14)):** This variable ascertains the number of chronic diseases from which a patient suffers. Chronic diseases in the context of this thesis include asthma, arthritis, cancer, cerebrovascular disease, congestive heart failure, chronic renal failure, chronic obstructive pulmonary disease, depression, diabetes, hyperlipidemia, hypertension, ischemic heart disease, obesity, and osteoporosis. I assume that, as the number of chronic diseases from which a patient suffers increases, the patient has more medical problems, which would be an incentive for a physician whose income is tied to his or her performance to avoid the patient, or refer the patient out to another physician.

- **chronill (chronic illness):** I created an indicator variable based on PRF responses to questions regarding the chronic diseases from which a patient suffered. I coded PRFs noting zero chronic diseases as “0” for `chronill` and coded PRFs noting a patient was suffering from one, two, three, four, five, six, seven, eight, or nine chronic diseases—there were no PRFs indicating any one patient suffered from more than nine chronic diseases—as “1” for `chronill`. I assume that being chronically ill is associated with more medical problems, which would be an incentive for a physician whose income is tied to his or her performance to avoid the patient, or refer the patient out to another physician.

- **refothmd (referral to another physician):** This variable indicates that the physician who saw the patient referred him or her to another physician.
Participation in pay-for-performance programs:

- **nop4p** (non-participation in pay-for-performance): I created this indicator variable based on PRF responses to the question, “Are performance measures on practice available to the public?” Because there were three possible answers to the question—“yes,” “no,” or “not applicable”—I assumed that an answer of “not applicable” indicated that there were no performance measures to be reported (as opposed to an answer of “yes,” which would indicate that there were performance measures and they were available to the public, and as opposed to an answer of “no,” which would indicate that there were performance measures but they were not available to the public). I coded PRFs with values of “not applicable” as “1” for *nop4p*, and I coded PRFs with values of either “yes” or “no” as “0” for *nop4p*. Note that a value of “0” for *nop4p* indicates that a physician was participating in pay-for-performance.

- **hip4p** (high physician participation in pay-for-performance program): I created this indicator variable based on PRF responses to the question, “What percent of your practice care revenue is based on bonuses, returned withholds or other performance-based payments?” I coded PRFs with values of “1” on *prp4pr*—indicating that zero to twenty-five percent of the physician’s practice-care revenue was based on bonuses, returned withholds or other performance-based payments—as “0” for *hip4p*, and I coded PRFs with values of two, three, or four on *prp4pr*—indicating that twenty-six to one-hundred percent of the physician’s practice-care revenue was based on bonuses, returned withholds or other performance-based payments—as “1” for *hip4p*. Depending upon the type of performance program in place, physicians with more than twenty-five percent of their patient-care compensation linked to

---

60 NAMCS variable name associated with this question is *measpub*.
61 NAMCS variable name associated with this question is *prp4pr*. 
performance-based payments would be more strongly incentivized to avoid or refer out to other physicians their sicker, more complicated patients.

**Data not available through NAMCS:**

NAMCS did not have data on the breakdown of pay-for-performance programs by patient outcome, versus pay-for-performance programs by adherence to clinical guidelines. Where physicians participate in pay-for-performance by adherence to clinical guidelines, I would not expect to see the same avoidance of sicker patients, because adherence to clinical guidelines is within the physicians’ control and would not be undermined by a noncompliant or particularly unhealthy patient as long as accurate risk-adjustment mechanisms were in place. NAMCS also does not include information on whether physicians had a choice in participating in pay-for-performance. If some physicians chose to participate in pay-for-performance programs, I would assume that those who opted in would more likely have healthier patients.

NAMCS does not include longitudinal data. The physicians surveyed one year are not necessarily the same group of physicians surveyed in the previous year. Had NAMCS conducted surveys of the same physicians over time, I might have been able to use to data sets from different years to conduct a difference-in-difference model, comparing the data on physicians’ referral behavior before pay-for-performance to the data on those same physicians’ referral behavior after pay-for-performance. By looking at how the average patient-risk levels in two similarly situated groups of doctors (one of which implemented a pay-for-performance program, and the other of which continued under a traditional payment structure) changed, I would be able to see whether physicians in pay-for-performance practices were deselecting sicker or more complicated patients.
NAMCS also does not include data that directly speaks to physician cherry-picking. Physicians may refer patients to other physicians for medically legitimate reasons (e.g., so that the patient could receive cancer treatment) or for illegitimate purposes (e.g., so that the physician could avoid continuing to deal with a sicker, more complicated patient who stood to negatively impact a physician’s performance rating). NAMCS does not provide a mechanism by which to determine whether the referrals are self-serving or medically legitimate.

Additionally, NAMCS’s most direct measure of whether a physician was participating in pay-for-performance was its prp4pr variable, which was broken into four quartiles indicating the percentage of a physician’s practice-care revenue that was based on bonuses, returned withholds, or other performance-based payments. NAMCS’s variable, prp4pr, did not distinguish between physicians whose pay was not at all tied to performance and physicians whose pay (up to twenty-five percent) was tied to performance.\(^62\) I therefore had to create nop4p and hip4p to capture whether or not a physician was participating in pay-for-performance at all, and whether or not that physician had a significant (twenty-six percent or more) stake in a pay-for-performance program.

\(^62\) The quartiles were defined as follows: 1) less than or equal to twenty-five percent, 2) twenty-six percent to fifty percent, 3) fifty-one to seventy-four percent, and 4) more than seventy-five percent.
CHAPTER V: METHODS OF ANALYSIS

Based on the available data, I focused on patient-records with reported values for the variables I included in my regression. Therefore, I dropped observations whose values for the variables that were of interest to me were either missing or blank.

My hypothesis was that the sicker a patient was (as measured by patient age, tobacco use, and chronic illness), the more likely a physician participating in a pay-for-performance program would be to refer that patient out to another physician. Accordingly, I also hypothesized that the higher the percentage of the physician’s practice-care revenue that was based on bonuses, returned withholds, or other performance-based payments, the more likely a physician would be to refer that patient out to another physician.

I ran four probit regressions to analyze the relationships.

Model #1: \[ \text{Pr}(\text{refothmd}) = \beta_0 + \beta_1(\text{age}) + \beta_2(\text{usetobac}) + \beta_3(\text{totchron}) + \beta_4(\text{nop4p}) \]

Model #2: \[ \text{Pr}(\text{refothmd}) = \beta_0 + \beta_1(\text{age}) + \beta_2(\text{usetobac}) + \beta_3(\text{totchron}) + \beta_4(\text{hip4p}) \]

Model #3: \[ \text{Pr}(\text{refothmd}) = \beta_0 + \beta_1(\text{age}) + \beta_2(\text{usetobac}) + \beta_3(\text{chronill}) + \beta_4(\text{nop4p}) \]

Model #4: \[ \text{Pr}(\text{chronill}) = \beta_0 + \beta_1(\text{age}) + \beta_2(\text{usetobac}) + \beta_3(\text{nop4p}) \]

Where:

- **refothmd**: referral to other physician
- **age**: patient age in years
- **usetobac**: whether or not patient currently uses tobacco
- **totchron**: total number of chronic diseases patient has
- **nop4p**: whether or not physician participates in pay-for-performance program
- **hip4p**: whether or not >25% of physician’s patient-care revenue is linked to performance

In each of the first three regressions, I included controls for patient age, tobacco use, and chronic illness, because each of these variables indicates a potentially sicker, more complicated patient and could therefore lead a physician to refer that patient to another physician where the
physician’s compensation was dependent upon patient-health outcomes. In the fourth regression, I included age, tobacco use, and physician participation in pay-for-performance so that I could see the correlation between those variables and the likelihood of a patient being chronically ill. Age and tobacco use are often associated with increased chronic illness, but I ran this regression to test whether or not the physicians participating in p4p programs were likelier to have sicker patients.
CHAPTER VI: RESULTS

This section provides an overview of the data via a table of summary statistics and then discusses the results of the four probit regressions I ran to test the hypothesis that physicians participating in pay-for-performance would be more likely to avoid sicker, more complicated patients. The following table summarizes the variables in which I was interested for the purposes of testing my hypothesis:

### Descriptive Statistics for Physician Record Forms

(Number of observations: 6,318)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Maximum</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>age</strong> patient age in years</td>
<td>47.04</td>
<td>100+</td>
<td>2</td>
</tr>
<tr>
<td><strong>totchron</strong> number of chronic conditions</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2,186 1,690 1,103 728 366 164 58 19 3 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>usetobac</strong> currently use tobacco</td>
<td>Yes No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,149 5,169</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>chronill</strong> chronically ill</td>
<td>Yes No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4,132 2,186</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>refothmd</strong> referral to another physician</td>
<td>Yes No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>543 5,775</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>nop4p</strong> physician non-participation in pay-for-performance program</td>
<td>Yes No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,158 3,160</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>hip4p</strong> &gt;25% of physician compensation is linked to performance</td>
<td>Yes No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>317 6,001</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Detail on physician performance measures

<table>
<thead>
<tr>
<th></th>
<th>≤ 25%</th>
<th>26–50%</th>
<th>51–75%</th>
<th>&gt; 75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>pccprod</td>
<td>6,001</td>
<td>173</td>
<td>16</td>
<td>128</td>
</tr>
<tr>
<td>pccsat</td>
<td>2,884</td>
<td>3,434</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pccqoc</td>
<td>1,458</td>
<td>4,860</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pccprof</td>
<td>743</td>
<td>5,575</td>
<td></td>
<td></td>
</tr>
<tr>
<td>measpub</td>
<td>824</td>
<td>5,494</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Regression Analysis

Throughout my regression analysis, I consistently found statistically significant correlations between physician participation in pay-for-performance and referrals to other physicians, but the relationship ran in the opposite direction than I’d hypothesized. I also found a statistically significant positive correlation between patient age and referrals to other physicians, such that advanced age was associated with an increased likelihood of referral to another physician.

---

63 These variables—pccprod, pccsat, pccqoc, pccprof, and measpub—were not included as variables in the regressions I ran, but they are included in this table because they reflect the distribution of the types of performance measures in the PRFs I analyzed.
physician. In my fourth regression, which tested the relationship between physicians’ participation in pay-for-performance programs and patients’ likelihood of being chronically ill, I found a statistically positive correlation such that physician participation in pay-for-performance was associated with an increased likelihood of a patient’s being chronically ill.

The probit model gives the probability of that a physician will refer a patient to another physician. In the first model, age has a statistically significant negative association with the likelihood that a patient will be referred to another physician. A physician’s participation in pay-for-performance has the largest effect of the variables studied and also has a statistically significant negative association with the likelihood that a physician will refer the patient to another physician.

An average patient\textsuperscript{64} whose physician is not participating in pay-for-performance has an 8.2\% chance of being referred to another physician, while a patient whose physician is participating in a pay-for-performance program has a 5.8\% chance of being referred to another physician.\textsuperscript{65} A physician’s participation in pay-for-performance is associated with a twenty-nine percentage-point decrease in the likelihood of referral to another physician.\textsuperscript{66} A physician’s participation in pay-for-performance is associated with a statistically significantly decrease in likelihood that the physician will refer his/her patient to another physician.

\begin{itemize}
\item \( Y = -1.466743 \cdot 0.0024626(47.04) \cdot \text{age} + 0.0899213(0) \cdot \text{usetobac} + 0.0093329(1.4) \cdot \text{totchron} + 0.1778109(1) \cdot \text{Nop4p} \)
\item \( Y = -1.466743 \cdot 0.116 \cdot \text{age} + 0.0131 \cdot \text{totchron} + 0.1778109 \cdot \text{Nop4p} \)
\item \( Y = -1.39 \) = z-score. Z table under -1.39 \( \Rightarrow \) 0.0823
\end{itemize}

Average patient whose physician is not participating in pay-for-performance:

\begin{itemize}
\item \( Y = -1.466743 \cdot 0.116 \cdot \text{age} + 0.0131 \cdot \text{totchron} \)
\item \( Y = -1.57 \) = z-score. Z table under -1.57 = 0.0582
\end{itemize}

\textsuperscript{64} I created an “average” patient based upon the 6,318 PRFs in my sample, in which the average patient age is 47 years old, does not currently use tobacco (only 18.2\% of the patients currently use tobacco), and suffers from 1.4 chronic illnesses.

\textsuperscript{65} Assuming the “average” patient whose physician is not participating in pay-for-performance, my calculation of the likelihood of physician referral is as follows:

\textsuperscript{66} (0.0823-0.0582)/0.0823=.29283111
In the second model, a patient’s age has the largest effect of the variables studied and has a statistically significant negative association with the likelihood that a patient will be referred to another physician. High participation in pay-for-performance programs (where a physician’s compensation is more than twenty-five percent reliant upon performance) also has a statistically significant negative association with the likelihood that a patient will be referred to another physician. Current use of tobacco has statistically significant positive association with the likelihood that a patient will be referred to another physician.

An average patient\(^6^\) whose physician’s compensation is more than twenty-five percent reliant upon performance has a 5.5% chance of being referred to another physician, while a patient whose physician’s compensation is less than twenty-five percent reliant upon performance has an 8.7% chance of being referred to another physician.\(^6^\) A physician’s

\(^{6}\) 47 yrs old, not currently using tobacco, suffering from 1.4 chronic illnesses. I created this “average” profile based upon the patients in the remaining sample (of 6,318). The average patient age is 47 years old (mean; median was 50 years old), does not currently use tobacco (only 18.2% of the patients currently use tobacco), and suffers from 1.4 chronic illnesses.

\(^{6}\) Average patient whose physician’s compensation is more than 25% tied to performance:

- \(Y = -1.369222 -.0023927(47.04)age + .0935298(0)usetobac + .008866(1.4)totchron -.2386411(1)hip4p\)
  - \(Y = -1.369222 -.113age + .124totchron -.2386411hip4p\)
  - \(Y = -1.60 = z\)-score. Z table under -1.60 \(\rightarrow 0.0548\)

Average patient whose physician’s compensation is \(\leq\) 25% tied to performance:

- \(Y = -1.369222 -.0023927(47.04)age + .0935298(0)usetobac + .008866(1.4)totchron\)
compensation being more than twenty-five percent reliant upon performance is associated with a fifty-nine percentage-point decrease in the likelihood of being referred to another physician. As with the first model, which showed that a physician’s participation in pay-for-performance is associated with a lower likelihood that the physician will refer his or her patient out to another physician, this model shows that a physician’s high participation in pay-for-performance (where the physician’s compensation is more than twenty-five percent reliant upon performance) is also statistically significantly associated with a lower likelihood that a physician will refer the patient to another physician.

Model #2

<table>
<thead>
<tr>
<th>refothmd</th>
<th>Coefficient</th>
<th>P value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>-.0023927**</td>
<td>0.037</td>
<td>-.0046443</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-.0001411</td>
</tr>
<tr>
<td>usetobac</td>
<td>.0935298*</td>
<td>0.099</td>
<td>-.0176041</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.2046637</td>
</tr>
<tr>
<td>totchron</td>
<td>.008866</td>
<td>0.632</td>
<td>-.0273663</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.0450982</td>
</tr>
<tr>
<td>hip4p</td>
<td>-.2386411**</td>
<td>0.044</td>
<td>-.4714034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-.0058788</td>
</tr>
<tr>
<td>constant</td>
<td>-1.369222</td>
<td>0.000</td>
<td>-1.532109</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-1.206335</td>
</tr>
</tbody>
</table>

***p < 0.01  **p < 0.05  *p < 0.10

In the third model, whether a physician participates in pay-for-performance has the largest effect of the variables studied and has a statistically significant negative association with the likelihood that a patient will be referred to another physician. A patient’s age also has a statistically significant negative association with the likelihood that a patient will be referred to another physician.

\[ Y = -1.369222 - .113\text{age} + .124\text{totchron} \]

\[ Y = -1.36 = z\text{-score} \]

\( Z \text{ table under } -1.36 \rightarrow 0.0869 \)

\( (0.0548-0.0869)/0.0548=\approx-0.58576642 \)
An average patient\textsuperscript{70} whose physician is \textit{not} participating in pay-for-performance has an 8.3\% chance of being referred to another physician, while a patient whose physician \textit{is} participating in a pay-for-performance program has a 5.6\% chance of referral.\textsuperscript{71} Physician participation in pay-for-performance is associated with a twenty-nine percentage-point decrease in the likelihood of referral to another physician.\textsuperscript{72}

\textbf{Model #3}

\begin{tabular}{lrrr}
\hline
refothmd & Coefficient & P value & 95\% Confidence Interval \\
\hline
\textit{age} & -0.0023458** & 0.039 & -0.0045757 -0.0001159 \\
\textit{usetobac} & 0.0897217 & 0.115 & -0.0219895 0.2014329 \\
\textit{chronill} & 0.0179653 & 0.747 & -0.0912106 0.1271412 \\
\textit{nop4p} & 0.1778917*** & 0.000 & 0.088991 0.2667925 \\
\textit{constant} & -1.470637 & 0.000 & -1.640041 -1.301232 \\
\hline
\end{tabular}

***\textit{p} < 0.01  **\textit{p} < 0.05  *\textit{p} < 0.10

\textsuperscript{70} 47 yrs old, not currently using tobacco, suffering from chronic illness. I created this “average” profile based upon the patients in the remaining sample (of 6,318). The average patient age is 47 years old (mean; median was 50 years old), does not currently use tobacco (only 18.2\% of the patients currently use tobacco), and suffers from at least one chronic illness (65\% of the patients are chronically ill).

\textsuperscript{71} Average patient whose physician is not participating in pay-for-performance:

\begin{itemize}
\item Y = -1.470637 -0.0023458(47.04)\textit{age} + 0.0897217(0)\textit{usetobac} + 0.0179653(1)\textit{chronill} + 0.1778917(1)\textit{nop4p}
\item Y = -1.470637 -0.0110\textit{age} + 0.0179653\textit{chronill} + 0.1778917\textit{nop4p}
\item Y = -1.38478 = Z-score. Z table under 1.38 \rightarrow 0.0838
\end{itemize}

Average patient whose physician \textit{is} participating in pay-for-performance:

\begin{itemize}
\item Y = -1.470637 -0.0023458(47.04)\textit{age} + 0.0897217(0)\textit{usetobac} + 0.0179653(1)\textit{chronill} + 0.1778917(0)\textit{nop4p}
\item Y = -1.470637 -0.110\textit{age} + 0.0179653\textit{chronill}
\item Y = -1.5626717 = Z-score. Z table under 1.56 \rightarrow 0.0594
\end{itemize}

\textsuperscript{72} (0.0838 -0.0594)/0.0838 = .29116945

31
<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>-1.466743</td>
<td>-1.369222</td>
<td>-1.470637</td>
</tr>
<tr>
<td>Age</td>
<td>-.0024626**</td>
<td>-.0023927**</td>
<td>-.0023458**</td>
</tr>
<tr>
<td>usetobac</td>
<td>.0899213</td>
<td>.0935298*</td>
<td>.0897217</td>
</tr>
<tr>
<td>tothron</td>
<td>.0093329</td>
<td>.008866</td>
<td>N/A</td>
</tr>
<tr>
<td>chronill</td>
<td>N/A</td>
<td>N/A</td>
<td>.0179653</td>
</tr>
<tr>
<td>nop4p</td>
<td>.1778109***</td>
<td>N/A</td>
<td>.1778917***</td>
</tr>
<tr>
<td>hip4p</td>
<td>N/A</td>
<td>-.2386411**</td>
<td>N/A</td>
</tr>
</tbody>
</table>

***p < 0.01  **p < 0.05  *p < 0.10

In the fourth model—which I ran to test whether physicians participating in pay-for-performance programs are likelier to have sicker patients—patient age, whether or not a patient currently uses tobacco, and whether or not a physician participates in pay-for-performance all had a statistically significant positive association with the likelihood that a patient was chronically ill. An average patient\(^3\) whose physician is not participating in pay-for-performance has 61.4% chance of being chronically ill, while a patient whose physician is participating in a pay-for-performance program has a 58% chance of being chronically ill.\(^4\) It thus seems that

\(^3\) 47 yrs old, not currently using tobacco. I created this “average” profile based upon the patients in the remaining sample (of 6,318). The average patient age is 47 years old (mean; median was 50 years old) and does not currently use tobacco (only 18.2% of the patients currently use tobacco).

\(^4\) Average patient whose physician is *not* participating in pay-for-performance:
- \(Y = -1.382688 + .033712(47.04) + .2063939(0) + .0892971(1) + 1.586age + .0892971nop4p\)
  - \(Y = -1.382688 + 1.586age + .0892971nop4p\)
  - \(Y = .2926091 = z\)-score. \(Z\) table under .2926091 → 0.6141

Average patient whose physician is *is* participating in pay-for-performance:
- \(Y = -1.382688 + .033712(47.04) + .2063939(0) + .0892971(0) + 1.586age\)
- \(Y = -1.382688 + 1.586age\)
physician participation in pay-for-performance is associated with a six percentage-point decrease in the likelihood of a patient’s being chronically ill.\textsuperscript{75}

### Model #4

<table>
<thead>
<tr>
<th>Chronil</th>
<th>Coefficient</th>
<th>P value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.033712</td>
<td>0.000***</td>
<td>0.0320369 - 0.035387</td>
</tr>
<tr>
<td>Usetobac</td>
<td>0.2063939</td>
<td>0.000***</td>
<td>0.1152938 - 0.2974939</td>
</tr>
<tr>
<td>nop4p</td>
<td>0.0892971</td>
<td>0.014**</td>
<td>0.0180377 - 0.1605565</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.382688</td>
<td>0.000</td>
<td>-1.520711 - 1.244665</td>
</tr>
</tbody>
</table>

***$p < 0.01$  **$p < 0.05$  *$p < 0.10$

### Policy Implications

The policy implications of my findings are unclear. I found that physicians’ participation in pay-for-performance had the opposite effect than I’d hypothesized, such that participation in pay-for-performance was actually associated with a lower likelihood of referral to another physician. My hypothesis predicted that physicians participating in pay-for-performance would be more likely to refer their patients to other physicians.

However, my hypothesis was actually specific to pay-for-performance by patient outcome, and NAMCS did not provide a means by which to distinguish pay-for-performance by patient outcome from pay-for-performance by adherence to guidelines. NAMCS also did not indicate whether participation in pay-for-performance was voluntary or mandatory, nor did it provide reasons for physician referrals.

Because I could not control for differences in the type or voluntariness of pay-for-performance, my analysis could be biased. If, as I hypothesize, pay-for-performance by patient-outcome encourages physicians to avoid unhealthy patients, then not controlling for the type of

\textsuperscript{75} Y = .203312 = z-score. Z table under .203312 \rightarrow 0.5793

\textsuperscript{75} \((0.6141-0.5793)/0.6141 = 0.0566683\)
pay-for-performance program would make it seem that pay-for-performance has less of an effect on physicians’ avoidance of sicker, more complicated patients than it actually does. If participation in pay-for-performance is voluntary, then physicians who typically had sicker, more complicated patients would be less likely to participate in pay-for-performance programs. While physicians who do not participate in pay-for-performance by patient outcome might make fewer self-serving referrals of sick patients than physicians who do participate in pay-for-performance, the lack of information on the reason for the referral makes it impossible to tease apart the effects of pay-for-performance on physician referrals.

Much more research analyzing the effects of pay-for-performance programs needs to be done, but for that research to be meaningful, researchers will need more data, and more detailed data with which to work. Given the relative dearth of detailed data available on pay-for-performance programs, policymakers are at a disadvantage in advocating for implementation, modification, or repeal of such programs. Without data specifying the type of pay-for-performance programs that are in effect, whether or not and to what extent physicians’ pay is affected by performance measures, whether or not participation in pay-for-performance is mandatory, and what the reasons for physician referrals are, policymakers have little evidence upon which to base decisions regarding pay-for-performance programs; and those decisions have significant impacts on patient health, physicians’ bottom line, and the effective operation of the healthcare system as a whole.

Other researchers have bemoaned the lack of empirical trials and of baseline comparisons in pay-for-performance programs, which make conclusions regarding the effects of such programs extremely difficult. One author went so far as to conclude, “[O]nly well-designed

---

76 See Doran, supra n.3 at 70; S.J. Tanenbaum, Pay for Performance in Medicare: Evidentiary Irony and the Politics of Value, 34 J. Health Pol., Pol’y & Law 717 (2009).
randomized controlled trials [] could establish effectiveness.” 77 In the absence of such trials, however, NAMCS could modify its questionnaire to obtain additional information. In the case of the percentage of a physician’s pay that is tied to performance measures, for example, it would be as simple as changing the first quartile to “zero” and setting up the second as “one to twenty-five percent,” the third as “twenty-six to fifty percent,” and the fourth as “fifty-one percent or more.” Better yet, for a more accurate picture, NAMCS could ask physicians to write in the actual percentage of their pay that is tied to performance measures. Because the vast majority of physicians participating in pay-for-performance programs have less than twenty-five of their pay tied to performance measures, a questionnaire designed to capture the detail and variation within what is currently the first quartile (zero to twenty-five percent) would greatly assist analysts in the field.

Similarly, NAMCS could collect data on the voluntariness of participation in pay-for-performance, and on the reason for referrals to other physicians. If researchers had access to data regarding whether the patient was referred to a specialist or a primary care physician, that would also be informative, as would data on the kind of specialist to whom a patient was referred. With more granular data, policy analysts, researchers, and healthcare administrators would be able to more effectively develop, implement, and evaluate pay-for-performance programs.

77 Tanenbaum, supra n.2 (internal citation omitted).
CHAPTER VII: CONCLUSION

Though I found a statistically significant negative effect of participation in pay-for-performance programs on physician referrals, pay-for-performance by patient outcome may well be resulting in physicians avoiding sicker and more complicated patients; it may just be that the vast majority of physicians participating in pay-for-performance were participating in programs based upon adherence to guidelines and that the negative effect of participation in pay-for-performance by patient outcome was swamped by the positive effect of participation in pay-for-performance by adherence to guidelines. After all, one study in the context of pay-for-performance by adherence to guidelines did find, that, “contrary to expectations, patients with greater complexity had higher odds of receiving high-quality care for hypertension.”

It could also be that the analyses I attempted using this dataset were unsuccessful. The NAMCS data was neither gathered nor broken down in a way that made my ideal analysis possible, but there may also have been ways to analyze the data that I did have that for some reason I did not employ. For example, physician referrals might not have been the best or the only means by which to measure physician avoidance of sicker, more complicated patients. And there might have been additional or better patient-health indicators that would better predict physician avoidance (if physicians are in fact avoiding sicker, more complicated patients). Perhaps particular chronic illnesses are more likely to result in adverse outcomes and would thus more strongly incentivize a physician whose compensation was tied to performance to avoid or refer the patient out to another physician. Additionally, my creation of an indicator for physician participation in pay-for-performance, nop4p, relied upon the assumption that an answer of “not applicable” to the question, “Were performance measures made available to the public?”

78 Petersen et al., “Will Hypertension Performance Measures Used for Pay-for-performance Programs Penalize Those Who Care for Medically Complex Patients?”, 119 Circulation 2978-85 (2009). Quality in this case was measured by adherence to guidelines and by patients’ perceptions of care.
As other articles have pointed out, there is much more research that needs to be done regarding pay-for-performance in general, and regarding pay-for-performance in the healthcare context in particular. The results of this thesis do not clearly point toward the success or failure of pay-for-performance in healthcare. While politicians continue to tout the potential of pay-for-performance, there are exemplars of success in healthcare like the Mayo Clinic—often touted as a model of high-quality, low-cost healthcare—, which achieved success by eliminating financial incentives for physicians. 79 Making a Mayo Clinic out of the entire U.S. healthcare system isn’t feasible, but its means of saving costs and improving quality should remind policymakers that their assumptions about what works are not without flaws.

In attempting to align physicians’ incentives with patient health, pay-for-performance must avoid creating perverse incentives that incentivize doctors to avoid sicker, more complicated patients because those patients would have a negative impact on their performance score. Pay-for-performance programs must do more than define “quality care”; they must ensure that problem is amenable to performance-payments; they must set clear performance goals, measure performance with robust quality indicators, reward achievement while providing effective risk-adjustment to protect physicians from factors beyond their control that can nevertheless affect their performance; deal with unintended effects; include feedback mechanisms; and be open to revision based upon that feedback so that such programs can continually be refined in furtherance of their goals.

Professor Beer suggested that pay-for-performance be viewed as “a larger exercise in fairness and justice” and warned managers not to proceed with such programs “until both sides

79 Gawande (The Mayo Clinic “pooled all the money the doctors and the hospital system received and began paying everyone a salary, so that the doctors’ goal in patient care couldn’t be increasing their income.” It then “promoted leaders who focused first on what was best for patients, and then on how to make this financially possible.”)
understand what they are getting into.**80** And to recap the American College of Physicians Center for Ethics and Professionalism,

[P]ay-for-performance programs should not use incentives that encourage physicians to discriminate against a class or category of patients...incentives should encourage physicians to care for the sickest and most vulnerable patients[, and] society should insist that health care systems do the most—not the least—for patients who need care the most. It should hold health care systems accountable for solving such problems as language barriers or poor health literacy.81

---

80 Id. (Lagace)
81 Snyder & Neubauer, supra n.8 at 793.
REFERENCES

American College of Physicians, “Market Forces Now Pushing Pay-for-performance,”
http://www.acponline.org/journals/news/may05/pm.htm.

H. Beckman et al., “Physician Reactions to Quantitative Individual Performance Reports,” 21

G. Billikopf, “Incentive Pay (pay for performance),” University of California, in Labor
Management, in AGRICULTURE: CULTIVATING PERSONNEL PRODUCTIVITY
(2003), available at http://www.cnr.berkeley.edu/ucce50/ag-labor/7labor/08.htm#o_1 J.H.

Burack et al., “Public Reporting of Surgical Mortality: A Survey of New York State

J. Carter et al., “Adherence to Burn Center Referral Criteria: Are Patients Appropriately Being
Referred?” 31 J. Burn Care & Res. 26 (Jan./Feb. 2010).

L.P. Casalino et al., “Will Pay-For-Performance and Quality Reporting Affect Health Care
Disparities?” 26 Health Affairs 405 (2007).

C.L. Damberg et al., “Taking Stock Of Pay-For-Performance: A Candid Assessment From The
Front Lines: Data From the Nation’s Largest P4P Program Show Some Positive Changes
But No Breakthrough Improvements in the Quality of Care,” 28 Health Affairs 517
(2009).

K. Davis et al., The Commonwealth Fund, 59, “Mirror, Mirror on the Wall: An International
Update on the Comparative Performance of American Health Care,” (Deborah Lorber ed.,
May 15, 2007).

T. Doran et al., “Exclusion of Patients from Pay-for-performance Targets by English Physicians,”

D.M. Eddy, “Balancing Cost And Quality In Fee-For-Service Versus Managed Care,” 16 Health Affairs 162 (1997).


