# Management Science

## An Evidence-Based Incentive System for Medicare's End-Stage Renal Disease Program

Donald K. K. Lee, Stefanos A. Zenios,

# An Evidence-Based Incentive System for Medicare's End-Stage Renal Disease Program

## Donald K. K. Lee

Yale School of Management, Yale University, New Haven, Connecticut 06520,
donald.lee@yale.edu

## Stefanos A. Zenios

Graduate School of Business, Stanford University, Stanford, California 94305, stefzen@stanford.edu

Recent legislations directed Medicare to revamp its decades-old system for reimbursing dialysis treatments, with focus on the risk adjustment of payments and on the transition toward a pay-for-compliance system. To design an optimal payment system that incorporates these features, we develop an empirical method to estimate the structural parameters of the principal–agent model underlying Medicare's dialysis payment system. We use the model and parameter estimates to answer the following questions: Can a pay-for-compliance system based only on the intermediate performance measures currently identified by Medicare achieve first-best? How should patient outcomes be risk adjusted, and what welfare gains can be achieved by doing so? Our main findings are as follows: (1) the current set of intermediate measures identified by Medicare are not comprehensive enough for use alone in a pay-for-compliance system; (2) paying for risk-adjusted downstream outcomes instead of raw downstream outcomes can lengthen the hospital-free life of admitted patients by two weeks per patient per year without increasing Medicare expenditures.

## 1. Introduction

This paper is concerned with the design of a payment system for Medicare's End-Stage Renal Disease (ESRD) program. This is a U.S. government program that eventually covers the expenses of all ESRD patients on dialysis therapy (ESRD is chronic failure of the kidneys, and dialysis is the only treatment besides a kidney transplant). Until 2005, dialysis providers were paid by Medicare on a per-treatment basis, so provider revenues were proportional to patient hospital-free life (downstream outcome) and no patient-specific adjustments were applied (case-mix adjustment). In fact, the payment system has remained virtually unchanged since 1983. The passage of legislations in 2003 (Medicare Prescription Drug, Improvement, and Modernization Act (MMA) 2003) and 2008 (Medicare Improvements for Patients and Providers Act (MIPPA) 2008) granted Medicare the authority to revamp it. First, recognizing that the incumbent system provided incentives for providers to selectively admit healthier patients, in 2005 Medicare introduced limited risk adjustment of payment rates based only on patient age, body mass index, and body surface area (*Federal Register* 2004). Second, there is a perception among policy makers that the incumbent pay-per-treatment structure is not performance contingent. Driven by this belief, the ESRD Quality Incentive Program (QIP 2010) was created to develop the "nation's first pay-for-performance" system. With the intention to encourage providers to conform to standardized best practice guidelines, the program proposed paying providers for compliance with specific care processes (intermediate outcomes): On January 1, 2012, provider compensation will be based on two intermediate measures: dialysis dosage adequacy and anemia control (*Federal Register* 2011).

Given Medicare's initiatives, this paper develops an evidence-based procedure for incorporating full risk adjustment and pay for compliance into the design of the dialysis payment system. To do this, we use an empirical principal–agent framework to capture the incentive dynamics between Medicare and a dialysis provider. We estimate the structural parameters of this model from data and use them to answer the following three questions: (i) Is it wise to switch from the current system to a pay-for-compliance one based only on dialysis adequacy and anemia control? (ii) If one is to pay for compliance, how exactly should provider performance be evaluated and rewarded?[1]

---

[1] Section 1881(h)(3)(A) of the Social Security Act mandates the development of an intermediate performance score for determining provider compensation.

(iii) What are the potential benefits of full risk adjustment?

The relevant principal–agent model is as follows: Medicare (the principal) has a stake in the production of patient health outcomes but delegates the control of the treatments to a dialysis provider (the agent). The provider chooses how much effort to put into each task of the treatment process, where effort is costly. Medicare observes each patient's aggregate health outcome (downstream outcome), which is a noisy signal of patient characteristics and the provider's effort in each task. Also observed are several other health outcome measures (typically, process-compliance measures to be referred to as intermediate outcomes) that are noisy signals of patient characteristics and the provider's efforts in a subset of tasks. Medicare can either reward the provider for aggregate downstream outcomes (current practice) or for performance on the intermediate measures (pay for compliance). Under either alternative, the question is to determine the compensation contract that will maximize Medicare's expected payoff.

Our first contribution answers the first question of whether Medicare's switch to pay for compliance is advisable. We show under model assumptions that simply paying dialysis providers for downstream outcomes is first-best, but is infeasible under the QIP's policy constraint of paying for compliance. Although in this paper we will not attempt to explain Medicare's motivation for using pay for compliance, we will examine whether the pay-for-compliance system envisioned by Medicare can attain first-best performance by rewarding for dialysis adequacy and anemia control alone. We use our empirical framework to estimate the cost of the efforts induced by the two intermediate measures and to estimate the optimal scoring scheme for provider performance on them (this scoring scheme answers our second question posed earlier). These estimates are then used to determine the optimal pay-for-compliance system. Our numerical results suggest that rewarding for the two intermediate measures alone is far from sufficient to match first-best performance; i.e., there exist important treatment tasks that can only be induced by rewarding for the downstream outcome or for additional intermediate measures. Thus, if Medicare is insistent on paying for compliance instead of downstream outcomes, it should at least identify these additional intermediate measures and reward for them as well.

Our second contribution is to propose an evidence-based method for full risk adjustment and to clarify the benefits of doing so, which addresses the last question posed above. Although an obvious benefit of risk-adjusting patient outcomes is the suppression of provider incentive to screen *prospective* patients for selective admittance, we do not focus on this here. Instead we demonstrate a previously unrecognized benefit of strengthening provider incentive to deliver quality care to patients *already admitted*. In Proposition 1, we show that rewarding for risk-adjusted downstream outcomes leads to a Pareto improvement in the downstream outcomes of admitted patients, Medicare's reward as well as provider earnings. Relative to paying for raw downstream outcomes, our numerical results suggest that the risk-adjusted system can increase an incumbent patient's hospital-free lifespan by two weeks per year without increasing Medicare expenditures or reducing provider profits.

The rest of this paper is organized as follows. Section 2 describes the background behind Medicare's current ESRD payment system and our proposed amendment to reward dialysis providers for risk-adjusted intermediate/downstream outcomes. The necessary data for our problem will be described. Section 3 presents a principal–agent model to capture the dynamics between Medicare and dialysis providers. The structural parameters of this model are estimated in §4. Armed with these estimates, it is then possible to optimally incorporate risk adjustment and process compliance into the payment system. Our estimates and the simulated performances of the optimal payment systems are presented in §5. Concluding remarks appear in §6.

### 1.1. Contributions to Literature

In recent years Medicare has begun collecting process-compliance measures from health providers to use as proxies for care quality. How these measures should be combined into a performance score for a pay-for-performance system remains an open question. Furthermore, there are debates within the medical community over whether providers should be rewarded for intermediate process measures or for patient outcomes: Jha (2006) provides a qualitative discussion of the pros and cons of processes versus outcomes. The article notes that outcomes are aligned with care quality but heavily confounded by patient-specific factors, whereas processes are simpler to measure but may not reflect all relevant aspects of care. For example, Bradley et al. (2006) found that the seven core process measures for acute myocardial infarction explained only 6% of the variation in standardized mortality rates for hospitalized AMI patients; thus, they argued that hospital quality should not be judged on processes alone. At present no rigorous evidence-based framework exists in the medical literature for determining if and how one should pay for intermediate measures, or how outcomes should be risk adjusted. Our work aims to fill this gap.

In the domain of operations management, the most relevant work to our study is Fuloria and

Zenios (2001), which analyzed a theoretical model of dialysis pay for performance using a dynamic principal–agent model. The authors proposed a scheme under which Medicare payments to a dialysis provider are tied to patient hospitalizations and deaths. At the beginning of each time epoch, the provider is paid a fixed fee for each patient it would treat during that epoch. At the end of the epoch, patient hospitalizations and/or deaths are observed and Medicare imposes a retrospective penalty on each adverse outcome. Thus, the scheme rewards for a reweighted version of the downstream outcome implicitly used by Medicare's current system. Because this work predated the MMA, it did not consider the possibility of risk adjustment or pay for compliance.

## 2. Problem Context and Data

The ESRD program is eventually responsible for the healthcare costs of all ESRD patients in the United States ($16 billion in 2003 and growing at a 10% annual rate (USRDS 2005)). Patients require dialysis therapy three times a week at a dialysis center, and the dialysis provider is reimbursed by Medicare. Our data comes from the U.S. Renal Data System (USRDS 2005), and we focus on Medicare patients who received in-center hemodialysis in 2003 from a particular chain of outpatient (non-hospital-based) dialysis providers situated in one state. To compare facilities of similar sizes, we excluded facilities of the chain that were geographically isolated from the rest and have far fewer patients (five facilities, 129 patients total). This resulted in a study cohort of $M = 12$ dialysis facilities, all within a 20-mile radius of each other. In 2003, $N = 842$ Medicare patients were treated at these facilities. For confidentiality, further provider information will be aggregated in a way to

prevent the chain and state from being identified. The available patient-level and provider-level data will be introduced as we progress through this section, and they are also gathered in Tables 1 and 2 for ease of review.

### 2.1. The Dialysis Process, Intermediate, and Downstream Outcomes

During dialysis, the patient's blood flow is connected to the dialysis machine through an access point on his or her arm (referred to as *vascular access*). The patient's blood is routed through a filter (dialyzer) to remove toxins that accumulate as a result of kidney failure. The amount of toxins removed during each treatment is measured by the so-called urea reduction ratio (URR); a treatment resulting in a URR of at least 65% is considered to have delivered *adequate* dialysis dosage and typically takes three to four hours. During each dialysis treatment, the provider also administers the drug erythropoietin alpha (EPO) to patients. EPO is used to treat the anemia caused by the failing kidneys' inability to stimulate red blood cell production. A patient is said to have received *sufficient* anemia control if the patient's hematocrit level (a measure of red blood cell count) lies within the range of 33% to 36%. By dedicating effort to treatment tasks that control dialysis dosage and anemia, each provider can achieve outcomes consistent with clinical guidelines. Examples of tasks that enhance these outcomes include frequent measurement of patient weight, body size, and hematocrit levels, and quality control systems for dialysis machines. Although these intermediate tasks are costly to the provider, they enhance a patient's URR and hematocrit level. For each dialysis treatment provided to patient $i$ by facility $j$, the database included information on the

**Table 1** Summary of Patient-Level and Provider-Level Data Used in the Analysis

| Variable name | Description (if applicable) | Data summary |
|---|---|---|
| $ANEMIA_{ij}$ | Control of patient $i$'s anemia (2): % of one year's treatments with hematocrit levels between 33% and 36% | Quartiles: [16.0%, 30.7%, 48.1%] |
| $BASEPAY_j$ | 2003 base composite rate (5) paid to provider $j$ for one full year of dialysis therapy for one patient | Quartiles: [$20,760, $21,200, $21,295] |
| $CPAT_{ij}$ | Transformed version of patient $i$'s attributes $PAT_{ij}$ for risk-adjustment purposes (20) | |
| $DOSE_{ij}$ | Patient $i$'s dialysis adequacy (1): % of one year's treatments with a Urea Reduction Ratio of $\geq 65\%$ | Quartiles: [83.7%, 94.0%, 100%] |
| $DSOUT_{ij}$ | Patient $i$'s downstream outcome (3): % of the year that a patient was alive and hospital-free | Quartiles: [0.419, 0.900, 0.995] years |
| $EPOCOST$ | The chain's acquisition cost for 1,000 units of EPO drug (4) | $8.93 per 1,000 units |
| $EPOPAY$ | Medicare's nationwide EPO reimbursement rate for 1,000 units (4) | $10 per 1,000 units |
| $PAT_{ij}$ | A comprehensive set of patient $i$'s demographic information and medical history (6) | See Table 2 |
| $WAGEIDX_j$ | Labor wage adjuster used to derive provider $j$'s composite rate $BASEPAY_j$ (5) | |

**Table 2    Patient Attributes That Constitute $PAT_{ij}$**

| Attribute | Description (if applicable) | Data summary |
|---|---|---|
| Age | Age on January 1, 2003 | Quartiles: [55.9, 66.7, 76.2] years |
| Gender | | 54.3% male |
| Race | | 4.74% Asian/Pacific Islander, 41.1% African-American, 54.2% Causasian/Other |
| Time since dialysis initiation | Time spent on dialysis therapy as of January 1, 2003 | Quartiles: [0.00, 1.25, 3.00] years |
| Body mass index | | Quartiles: [22.5, 26.1, 30.7] kg/m² |
| Body surface area | | Quartiles: [1.65, 1.81, 1.97] m² |
| Serum albumin level | Measure of nutrition levels; low albumin levels indicate malnourishment (≤3.5 g/dl) | Quartiles: [2.80, 3.30, 3.65] g/dl |
| Serum creatinine level | Indicator of kidney function | Quartiles: [5.4, 7.1, 9.1] mg/dl |
| Blood urea nitrogen level | Indicator of kidney function | Quartiles: [65, 85, 107] mg/dl |
| Hemoglobin level | Indicator of kidney function | Quartiles: [8.2, 9.5, 10.7] g/dl |
| Glomerular filtration rate | Indicator of kidney function | Quartiles: [5.95, 8.02, 10.4] ml/min |
| Ability to move | | 98.2% of patients |
| Tobacco use | | 2.78% of patients |
| Alcohol use | | 0.927% of patients |
| Drug dependence | | 1.65% of patients |
| Cancer | | 3.91% of patients |
| Diabetes | | 14.1% of patients |
| History of hypertension | | 76.7% of patients |
| Pulmonary disease | | 4.84% of patients |
| Congestive heart failure | | 37.9% of patients |
| Ischemic heart disease | | 25.9% of patients |
| Myocardial infarction | | 5.87% of patients |
| Cardiac dysrhythmia | | 5.66% of patients |
| Pericarditis | | 0.927% of patients |
| Cerebrovascular disease | | 8.34% of patients |
| Peripheral vascular disease | | 13.9% of patients |

URR and hematocrit level. These were used to define patient $i$'s intermediate measures as

$$DOSE_{ij} = \text{\% of a full year's treatments that resulted in adequate dialysis dosage,} \quad (1)$$

$$ANEMIA_{ij} = \text{\% of a full year's treatments that achieved sufficient anemia control.} \quad (2)$$

In 2003 the median *DOSE* was 94% and the median *ANEMIA* was 31%.

The complications that arise from the dialysis treatment itself may cause patients to be hospitalized or even die. Providers influence hospital admissions and mortality rates by ensuring adequate dialysis dosage and sufficient anemia control, and also by engaging in other tasks. For example, a preventable cause of hospital admissions is excessive fluid gain in patients with congestive heart failure. Specifically, because their kidneys do not function, patients retain substantial amounts of fluid between dialysis sessions. Patients with heart failure who gain more than four pounds in fluids between treatments have higher hospital admission rates. The dialysis clinics, however, can reduce these admissions rates by implementing various measures to manage excessive weight gain between treatments: dietary counseling and/or equipping patients with an Internet-enabled scale that allows remote weight monitoring. These downstream tasks are costly to the provider but have a significant impact on patient hospital-free lifespan. We used hospitalization and death records from the data set to define patient $i$'s downstream outcome as

$$DSOUT_{ij} = \text{fraction of the year for which patient } i \text{ is alive and hospital free.} \quad (3)$$

The median patient in our study population experienced 0.90 years of hospital-free life in 2003. Note that at any point in time during 2003, a patient belonged to one of three mutually exclusive states: alive and hospital free, hospitalized, or deceased. Our definition of the downstream outcome places no weight on the latter two states; thus, it is the most unforgiving metric of care quality. Hence, analyses based on this provide a conservative estimate for the benefits that can be had from optimizing the payment system.

## 2.2.    Costs and Incentives
The expenses stemming from patient hospital admissions account for a third of the Medicare program's expenditures in 2003 (USRDS 2005), and Medicare

is interested in maximizing patient downstream outcomes. Since 1983 Medicare has paid dialysis providers under a composite rate system that was still used in 2003 (U.S. Government Accounting Office (GAO) 2004): Non-hospital-based providers were paid a base composite rate $BASEPAY_j$ for each full year of dialysis therapy rendered (pay-per-treatment), and separately reimbursed for the amount of EPO drug administered at a rate of $EPOPAY = \$10$ per 1,000 units (the chain's cost of acquisition was $EPOCOST = \$8.93$ per 1,000 units). The amount $BASEPAY_j$ includes a 20% copayment by patients and is meant to cover all treatment costs aside from certain drugs and laboratory tests. Of the separate drug payments, the EPO payments as described account for the overwhelming portion of provider drug revenues and profits (Thamer et al. 2007). Fees for the separately billable laboratory tests are paid directly to the laboratory. To simplify exposition we will not model the EPO drug component explicitly.[2] Under this pay-per-treatment structure providers were paid only when patients are alive and not in the hospital; thus, provider $j$'s 2003 compensation for treating patient $i$ was proportional to the raw downstream outcome:

$$CONTRACT_{ij}^{\text{current}} = BASEPAY_j \cdot DSOUT_{ij}. \quad (4)$$

$BASEPAY_j$ differed from facility to facility due to a labor wage index $WAGEIDX_j$:

$$BASEPAY_j = \$19,707 \cdot WAGEIDX_j. \quad (5)$$

For our chain cohort $BASEPAY_j$ ranged from \$20,168 to \$21,958 per year.[3] Before the enactment of the MMA (2003), Medicare lacked the authority to update the wage indices—in 2003 they were still based on a mix of the 1980 Bureau of Labor Statistics (BLS) index and the 1982 Health Care Financing Administration (HCFA) index for the *average* hospital wage for the provider's area (*Federal Register* 1986, 2005). As a result, $WAGEIDX_j$ reflected little of the actual variation in 2003 wages across providers. For example, a zipcode that was classified as rural under the 1980 census but has since become urbanized still received the rural wage adjustment in 2003. Indeed, part of the motivation behind the MMA was to grant Medicare the authority to update the wage index whenever necessary.

---

[2] The analysis of the full model is available from the authors' webpage.

[3] Providers received \$126.33 per treatment before wage adjustment, or $\$126.33 \times 3 \times 52 = \$19,707$ for one full year. In reality, the wage adjustment was not applied to the whole \$19,707, but the end result is equivalent to (5) after renormalization.

## 2.3. The New Payment System

The MMA also directed Medicare to risk adjust its payments to dialysis providers. Since 2005 limited adjustment based on patient age, body mass index, and body surface area has been applied to the composite rate (*Federal Register* 2004). The MIPPA (2008) also mandated that Medicare implement the ESRD QIP (2010) on January 1, 2012. This program aims to reward providers for compliance with quality targets for adequate dosage and controlling anemia, with plans to incorporate additional performance measures in future years. Specifically, the program is bound by section 1881(h)(3)(A) of the Social Security Act (added on by MIPPA) to develop a way to score providers on their dosage adequacy and anemia control performances. Payments are to be based on these intermediate outcome scores. In January 2011 the program finalized a scoring methodology (*Federal Register* 2011) that will score according to whether the provider's performance *in* 2010 exceeded either its own performance in 2007 or the 2008 national average performance. The provider will then receive a composite rate for each treatment rendered in 2012 that is adjusted by the intermediate outcome score based on 2010 performance. The specifics of this scoring methodology are too detailed to describe here, but a few points of concern are worth noting: (i) The scientific rationale behind the methodology is unclear (this is also true with regard to the way payments have been risk adjusted since 2005). (ii) If the 2012 payment rates are predetermined by past performance in 2010, there is no incentive for improving 2012 intermediate outcome performance. (iii) Retaining the pay-per-treatment structure of the composite rate effectively rewards providers for downstream outcomes. Together (ii) and (iii) mean that providers will actually be rewarded for downstream outcomes rather than intermediate performance, so the program's plan is not in fact pay for compliance—it is structurally identical to the 2003 system (4). In this paper we propose an alternate scoring and payment system that is an extension of (4). It is flexible enough to accommodate full risk adjustment and a pay-for-compliance system based on the two intermediate measures, but is still simple enough to permit analysis:

$$CONTRACT_{ij}^{\text{new}} = -\pi_{ij}^R + \pi_j^{int} f(DOSE_{ij}, ANEMIA_{ij})$$
$$+ \pi_j^{ds} DSOUT_{ij}, \quad (6)$$

where $f(DOSE_{ij}, ANEMIA_{ij})$ is the intermediate performance score for patient $i$'s $DOSE_{ij}$ and $ANEMIA_{ij}$, as sought by section 1881(h)(3)(A) of the Social Security Act; $\pi_{ij}^R$ is a statistical risk adjustment that is subtracted from provider $j$'s revenue from treating patient $i$; and $\pi_j^{int}$ and $\pi_j^{ds}$ are the piece rates paid

to provider $j$ for each unit of intermediate score and each unit of downstream outcome, respectively.

In §3 we will formally introduce $f(\cdot, \cdot)$ and show that it has the same units as $DSOUT$ ($f$ is *aligned* with $DSOUT$). Several instances of (6) are relevant to the questions posed in the introduction:

1. Setting $\pi_{ij}^R = \pi_j^{int} = 0$ and $\pi_j^{ds} = BASEPAY_j$ recovers the 2003 system (4) that rewards providers for raw downstream outcomes:

$$BASEPAY_j \cdot DSOUT_{ij}. \qquad (7)$$

2. Setting $\pi_j^{int} = 0$ yields a contract that rewards providers for downstream outcomes with risk adjustment:

$$\pi_j^{ds} DSOUT_{ij} - \pi_{ij}^R. \qquad (8)$$

For risk-neutral providers (Assumption 7 in §3.4) this is first-best.

3. Restricting $\pi_j^{ds} = 0$ results in a pay-for-compliance system that rewards providers for risk-adjusted performance on $DOSE$ and $ANEMIA$:

$$\pi_j^{int} f(DOSE_{ij}, ANEMIA_{ij}) - \pi_{ij}^R. \qquad (9)$$

This contract is only admissible if it can match first-best performance (8). We will show that if dosage adequacy and anemia control are the only treatment processes that influence downstream outcomes, then the contracts (8) and (9) are equivalent.

The difference in performance between (7) and (8) represents the welfare gain due to risk adjustment. A divergence in performance between (8) and (9) will indicate that the pay-for-compliance system under consideration cannot achieve first-best. To carry out these comparisons, the optimal performance of each case needs to be established. This requires choosing rates $\pi^R$, $\pi^{int}$, and $\pi^{ds}$, and the scoring for intermediate performance $f(\cdot, \cdot)$ to incentivize dialysis providers to improve downstream outcomes. In the next section we present a principal–agent model that can be used to guide these choices.

## 3. The model

For simplicity, we assume that the dialysis providers' actions are based on a single-period contract with Medicare, where the period runs from January 1, 2003, to December 31, 2003. At the beginning of the period, the manager of the chain's $j$th dialysis facility decides on how much resources and effort to devote to patient $i$'s care. This is described by the variables $z_{ij}^{int}$ and $z_{ij}^{ds}$, which are known to the provider but not observed by Medicare: $z_{ij}^{int}$ captures all treatment tasks that influence the intermediate measures, some of which were mentioned in §2; in turn, $z_{ij}^{ds}$ captures all additional treatment tasks that avert hospitalizations and/or deaths. The model will be developed in

three steps: First, we present the patient outcome production functions that show how treatment tasks are translated into patient outcomes. Then we describe the provider's problem, and finally conclude with Medicare's problem. All modeling assumptions are discussed in detail afterward.

### 3.1. Production Functions

For patient $i = 1, \ldots, N_j$ receiving treatment from provider $j$, we model his intermediate measures as

$$DOSE(z_{ij}^{int}, PAT_{ij}, \varepsilon_{ij}^{int}), \qquad (10)$$

$$ANEMIA(z_{ij}^{int}, PAT_{ij}, \varepsilon_{ij}^{int}), \qquad (11)$$

where $PAT_{ij}$ is a comprehensive set of demographic information and medical history that describes patient $i$'s underlying health (the variables comprising $PAT_{ij}$ are given in Table 2); and $\varepsilon_{ij}^{int}$ is a shock (independent over patients) that is realized after the period begins and is beyond of the control of the patient and provider: An example would be a defect in patient $i$'s dialysis filter that leads to decreased levels of $DOSE_{ij}$. We model the patient's downstream outcome as

$$DSOUT(z_{ij}^{int}, z_{ij}^{ds}, DOSE_{ij}, ANEMIA_{ij}, PAT_{ij}, \varepsilon_{ij}^{ds}), \qquad (12)$$

where $\varepsilon_{ij}^{ds}$ is a shock (also independent over patients) that is realized after the period begins and is beyond of the control of the patient and provider. An example of this shock would be an infection of the access point (from where blood is drawn into the dialysis machine) caused by accidentally coming into contact with unsanitary bath water. We model the functional form of (12) additively as

$$DSOUT_{ij} = e^{ds}(z_{ij}^{int}, z_{ij}^{ds}) + f(DOSE_{ij}, ANEMIA_{ij}) + h^{ds}(PAT_{ij}) + \varepsilon_{ij}^{ds}, \qquad (13)$$

where $e^{ds}(z_{ij}^{int}, z_{ij}^{ds})$ is the impact of $(z_{ij}^{int}, z_{ij}^{ds})$ on the downstream outcome. We assume that $e^{ds}(\cdot, \cdot)$ is nondecreasing in the first argument and increasing in the second. The dependence of the downstream outcome on the intermediate measures $(DOSE_{ij}, ANEMIA_{ij})$ is captured by the scalar function $f(\cdot, \cdot)$. It is clear from (13) that this is precisely the function of $DOSE$ and $ANEMIA$ that is most aligned with the downstream outcome. As discussed in Gibbons (2005), strong alignment between the principal's objective and the agent's incentives leads to more efficient contracts. Thus, Medicare should use $f(DOSE_{ij}, ANEMIA_{ij})$ as the unit for paying providers for their intermediate performance; i.e., $f(DOSE_{ij}, ANEMIA_{ij})$ is precisely the intermediate performance score sought by the Social Security Act. This was anticipated by the notation in the proposed payment scheme (6). Finally, $h^{ds}(PAT_{ij})$ is the

component of the residual that is due to the patient's underlying health $PAT_{ij}$.

In view of (10) and (11), $f(DOSE_{ij}, ANEMIA_{ij})$ is a function of $z_{ij}^{int}$, $PAT_{ij}$, and $\varepsilon_{ij}^{int}$. By also modeling this additively, we obtain

$$f(DOSE_{ij}, ANEMIA_{ij})$$
$$= e^{int}(z_{ij}^{int}) + h^{int}(PAT_{ij}) + \varepsilon_{ij}^{int}, \qquad (14)$$

where $e^{int}(\cdot)$ is assumed to be an increasing function, and $h^{int}(PAT_{ij})$ is the component of the score that is due to the patient's underlying health $PAT_{ij}$.

The monotonicity of $e^{int}(\cdot)$ and $e^{ds}(\cdot, \cdot)$ imply a bijection between $\{z_{ij}^{int}, z_{ij}^{ds}\}$ and $\{e_{ij}^{int}, e_{ij}^{ds}\} \equiv \{e^{int}(z_{ij}^{int}), e^{ds}(z_{ij}^{int}, z_{ij}^{ds})\}$, so by a change of basis we will refer to $\{e_{ij}^{int}, e_{ij}^{ds}\}$ as the new control variables. Equations (13) and (14) then describe the provider's production functions for downstream and intermediate outcomes, respectively:

$$DSOUT_{ij} = e_{ij}^{ds} + h^{ds}(PAT_{ij})$$
$$+ f(DOSE_{ij}, ANEMIA_{ij}) + \varepsilon_{ij}^{ds}, \qquad (15)$$
$$f(DOSE_{ij}, ANEMIA_{ij}) = e_{ij}^{int} + h^{int}(PAT_{ij}) + \varepsilon_{ij}^{int}.$$

## 3.2. Provider's Problem

The manager chooses $\{e_{ij}^{int}, e_{ij}^{ds}\}_{1 \le i \le N_j}$ at the beginning of the period, and the shocks $(\varepsilon_{ij}^{int}, \varepsilon_{ij}^{ds})_{1 \le i \le N_j}$ (hence patient outcomes) are realized afterward. Exerting effort to treat patients incurs a fixed cost to provider $j$, assumed to be quadratic in $e_{ij}^{int}$ and $e_{ij}^{ds}$: $c^0 + c^{int}e_{ij}^{int} + c^{ds}e_{ij}^{ds} + \frac{1}{2}[c^{int,int}(e_{ij}^{int})^2 + 2c^{int,ds}e_{ij}^{int}e_{ij}^{ds} + c^{ds,ds}(e_{ij}^{ds})^2]$. This captures the leading Taylor orders of the actual fixed cost. Under the payment scheme (6), provider $j$'s expected aggregate profit for treating $N_j$ patients is

$$\Pi_j = \mathbb{E}\sum_{i=1}^{N_j}\{-\pi_{ij}^R + \pi_j^{int}f(DOSE_{ij}, ANEMIA_{ij})$$
$$+ \pi_j^{ds}DSOUT_{ij} - c^0 - c^{int}e_{ij}^{int} - c^{ds}e_{ij}^{ds}$$
$$- \frac{1}{2}[c^{int,int}(e_{ij}^{int})^2 + 2c^{int,ds}e_{ij}^{int}e_{ij}^{ds} + c^{ds,ds}(e_{ij}^{ds})^2]\}$$
$$= \sum_{i=1}^{N_j}\{(-c^{int} + \pi_j^{int} + \pi_j^{ds})e_{ij}^{int} + (-c^{ds} + \pi_j^{ds})e_{ij}^{ds}$$
$$- \frac{1}{2}[c^{int,int}(e_{ij}^{int})^2 + 2c^{int,ds}e_{ij}^{int}e_{ij}^{ds} + c^{ds,ds}(e_{ij}^{ds})^2]$$
$$+ (\pi_j^{int} + \pi_j^{ds})h^{int}(PAT_{ij})$$
$$+ \pi_j^{ds}h^{ds}(PAT_{ij}) - \pi_{ij}^R - c^0\}. \qquad (16)$$

It follows from this that the profit-maximizing effort levels are fixed across patients within each provider:

$$e_{ij}^{int} = e_j^{int} = [c^{ds,ds}(-c^{int} + \pi_j^{int} + \pi_j^{ds}) - c^{int,ds}(-c^{ds} + \pi_j^{ds})]$$
$$/[c^{int,int}c^{ds,ds} - (c^{int,ds})^2]$$

$$e_{ij}^{ds} = e_j^{ds} = [c^{int,int}(-c^{ds} + \pi_j^{ds}) - c^{int,ds}(-c^{int} + \pi_j^{int} + \pi_j^{ds})]$$
$$/[c^{int,int}c^{ds,ds} - (c^{int,ds})^2]. \qquad (17)$$

## 3.3. Medicare's Problem

Medicare receives a reward $\nu$ for each unit of downstream outcome produced, but not directly for the intermediate measures. To implement the new payment scheme (6) for provider $j$, Medicare needs to choose the payment parameters $\pi_{ij}^R$, $\pi_j^{int}$, and $\pi_j^{ds}$. Being risk neutral, Medicare's objective is to maximize its expected gain

$$\Pi^{Medicare} = \mathbb{E}\sum_{i=1}^{N_j}\{\nu DSOUT_{ij} + \pi_{ij}^R$$
$$- \pi_j^{int}f(DOSE_{ij}, ANEMIA_{ij}) - \pi_j^{ds}DSOUT_{ij}\}$$
$$= \sum_{i=1}^{N_j}\{\pi_{ij}^R + (\nu - \pi_j^{int} - \pi_j^{ds})e_{ij}^{int} + (\nu - \pi_j^{ds})e_{ij}^{ds}$$
$$+ (\nu - \pi_j^{int} - \pi_j^{ds})h^{int}(PAT_{ij})$$
$$+ (\nu - \pi_j^{ds})h^{ds}(PAT_{ij})\}. \qquad (18)$$

Medicare's problem is to choose $\pi_{ij}^R$, $\pi_j^{int}$, and $\pi_j^{ds}$ to maximize $\Pi^{Medicare}$ subject to (i) the incentive compatibility constraint that $(e_{ij}^{int}, e_{ij}^{ds})$ maximizes $\Pi_j$, i.e., (17); and (ii) the individual rationality constraint that the provider's resultant profit $\Pi_j$ must be at least its current profit $\Pi_j^{current}$. It follows that the optimal statistical risk adjustments $\pi_{ij}^R$ depend on patient characteristics $PAT_{ij}$ through $h^{int}(PAT_{ij})$ and $h^{ds}(PAT_{ij})$, and the form of the dependence is automatically determined by the optimization.

Medicare does not need to know the constant term of the fixed cost $c^0$ to implement the proposed contract (6). However, it must know the functional forms of $f(\cdot, \cdot)$, $h^{int}(\cdot)$, $h^{ds}(\cdot)$, and the cost parameters $c^{int}$, $c^{ds}$, $c^{int,int}$, $c^{int,ds}$, and $c^{ds,ds}$. This raises the practical consideration of how Medicare can estimate these parameters. In §4 we present an estimation strategy for extracting these key parameters from data.

## 3.4. Model Assumptions

The model equations (15) and (17) are key to our empirical strategy, and the assumptions underlying their development are as follows:

1. The health shocks $(\varepsilon_{ij}^{int}, \varepsilon_{ij}^{ds})$ are assumed to be independent across all patients, and identically distributed across patients of the same facility.

2. Although by definition the intermediate effort $z_{ij}^{int}$ influences downstream outcomes through $DOSE_{ij}$ and $ANEMIA_{ij}$, we also allow for possible interactions between $z_{ij}^{int}$ and $z_{ij}^{ds}$ to influence $DSOUT_{ij}$ through $e^{ds}(z_{ij}^{int}, z_{ij}^{ds})$ in (13).

3. The mappings from provider efforts to their impact on outcomes $e^{int}(\cdot)$ and $e^{ds}(\cdot, \cdot)$ represent the providers' ability to translate effort into patient outcomes. Because the providers in our study all belong to the same chain in the same region, they share the same treatment technologies and practices. We therefore assume that all providers have the same ability. Note that this is compatible with the possibility that staffing quality could be heterogeneous across providers: The staffing quality mix at a facility is determined by its hiring decisions, which is part of $(z_{ij}^{int}, z_{ij}^{ds})$. Hence, a provider who chooses to hire more-experienced staff will have larger $z$-values than a provider who hires less-qualified staff, although both share the same $e^{int}(\cdot)$ and $e^{ds}(\cdot, \cdot)$.

4. The monotonicity conditions on $e^{int}(\cdot)$ and $e^{ds}(\cdot, \cdot)$ are the weakest possible that still guarantee a bijection between $\{z_{ij}^{int}, z_{ij}^{ds}\}$ and $\{e_{ij}^{int}, e_{ij}^{ds}\}$. Although both mappings are required to be strictly increasing in one argument, they are allowed to asymptote toward upper bounds since $e_j^{int} + e_j^{ds} \leq 1$.

5. The functions $h^{int}(\cdot)$ and $h^{ds}(\cdot)$ are the physiological impacts of patient attributes on outcomes. As such, these functional forms are fixed across patients and providers. Similarly, $f(\cdot, \cdot)$ is the physiological response function to intermediate measures, and hence is also fixed across patients and providers.

6. We specified the production functions of $DSOUT$ and $f(DOSE, ANEMIA)$ additively in the arguments $z$, $PAT$, and $\varepsilon$ in (13) and (14). These specifications represent the leading order terms of the ANOVA expansions of the production functions. Recall that the ANOVA expansion is of the form $g(x_1, \ldots, x_n) = \sum_i g_i(x_i) + \sum_{ij} g_{ij}(x_i, x_j) + \ldots$, so the production functions (15) do not contain interaction terms between provider effort and patient characteristics. As a consequence, the provider's optimized efforts (17) do not depend on patient characteristics; i.e., efforts are fixed across patients within each provider. Two sources of information suggest that this restriction provides a reasonable approximation to reality. First, the contrapositive of the consequence above is that if providers did customize efforts to individual patients, the production functions (15) should involve interactions between effort and patient characteristics. The optimized efforts (17) should then involve interactions between $PAT_{ij}$ and $BASEPAY_j$. We added these interaction terms to the specifications of the production estimating Equation (21) in §4 and tested for their significance. The nested ANOVA $F$-test did not reject the null hypothesis that the production functions do not contain these interactions ($p$-value $> 0.20$). Second, based on conversations with nephrologists, we understand that there is little room for treatment customization, because in the dialysis setting most of the treatment is in fixed setup:

A provider purchases equipment and chooses staffing levels based on the aggregate needs of the population because these resources are shared among all patients and cannot be customized to individuals. For example, once a facility installs a Web-based weight-monitoring system, all patients get access to it. Similarly, a provider cannot procure staff just to take care of one patient. Of course, sicker patients may get more staff attention, leaving less time for others, but these fluctuations within a provider are decidedly second order compared to the fixed setup.

7. In modeling the provider's utility, we implicitly assumed that the manager is risk neutral and seeks to maximize expected profit $\Pi_j$. This is a reasonable assumption given that the dialysis providers treat a large number of patients (on average, 70).

8. Because the provider cohort all belong to the same chain and are clustered in the same region, they are subject to the same salary pay scale as determined by the regional headquarters: Two nurses with identical qualifications but working for different providers within the cohort should receive the same salary. Furthermore, the providers share the same treatment technologies and supplies as negotiated through the chain. We therefore assume that the providers in our study cohort share the same cost parameters $c^{int}$, $c^{ds}$, $c^{int,int}$, $c^{int,ds}$, and $c^{ds,ds}$. Note that this does not contradict the fact that the composite rate $BASEPAY_j$ (5) is adjusted by geographic wage variations: The adjuster $WAGEIDX_j$ is based on the outdated *average* wage for provider $j$'s location and not on provider $j$'s *actual* pay scale. The existence of variation in the *average* wage across provider zipcodes does not preclude the existence of a subset of providers with a common pay scale. To see this, consider a hypothetical example of two locations $x$ and $y$ where a chain has one provider in each place, and each provider pays \$50,000 for a newly minted technician. Suppose one more dialysis provider exists in $x$ and pays an equally qualified technician \$51,000. Also, suppose one more provider exists in $y$ and pays \$49,000. Then variation exists in the average wage (\$50,500 for $x$ and \$49,500 for $y$) even though the chain providers pay identical salaries.

9. Following Fuloria and Zenios (2001), our model focuses on the main strategic players Medicare and the dialysis provider. Unlike a disease like diabetes, where the patient plays a major role in controlling the condition, in dialysis the main controls rest with the provider.

### 3.5. Necessary and Sufficient Condition for the Pay-for-Compliance Contract (9) to Match the First-Best Performance of (8)

Intuitively, if everything a provider can do to control downstream outcomes is through either managing

anemia or dialysis adequacy, then the risk-adjusted pay-for-compliance contract (9) based on $DOSE_{ij}$ and $ANEMIA_{ij}$ should do as well as paying for risk-adjusted downstream outcomes (8). In fact, under this scenario the two contracts are equivalent. To see this, note that $f(DOSE_{ij}, ANEMIA_{ij})$ and $DSOUT_{ij}$ differ by $e^{ds}(z_{ij}^{int}, z_{ij}^{ds}) + h^{ds}(PAT_{ij})$ in expectation. Applying risk adjustment to the outcomes eliminates the term $h^{ds}(PAT_{ij})$. The interaction term $e^{ds}(z_{ij}^{int}, z_{ij}^{ds})$, defined in the specification of $DSOUT_{ij}$ (13), exists only if $z_{ij}^{ds}$ does (Assumption 2). Therefore, if the additional treatment tasks $z_{ij}^{ds}$ do not exist, then the two payment metrics are equivalent and the pay-for-compliance contract is first-best. Unfortunately, the results from §5 suggest that $z_{ij}^{ds}$ exists so that paying for performance on $DOSE_{ij}$ and $ANEMIA_{ij}$ alone is suboptimal.

### 3.6. Benefits of Risk Adjustment
That risk adjustment is important because of the incentive for providers to screen *prospective* patients for selective admittance is obvious. It turns out that proper risk adjustment also strengthens provider incentive to deliver quality care to patients *already admitted*, as demonstrated by the following proposition. We simplify the setting to one provider (the subscript $j$ is dropped).

PROPOSITION 1. *Let $(\pi^{ds,0}, \Pi^{A,0}, \Pi^{P,0})$ be the optimal piece rate, provider (agent) earning, and Medicare's (principal's) reward under the contract that pays for raw downstream outcomes. Let $(\pi_i^R, \pi^{ds,1}, \Pi^{A,1}, \Pi^{P,1})$ denote the risk adjustments, piece rate, provider earning, and Medicare's reward under the contract that pays for downstream outcomes with risk adjustment. Then there exists a choice of $\pi_i^R$ and $\pi^{ds,1}$ such that (i) $\pi_i^R \geq 0$; (ii) $\pi^{ds,1} > \pi^{ds,0}$, thereby inducing higher provider effort (17) under the risk-adjusted contract; (iii) $\Pi^{A,1} = \Pi^{A,0}$ satisfying the individual rationality constraint; and (iv) $\Pi^{P,1} > \Pi^{P,0}$.*

The proof is trivial and is omitted. The intuition behind the result is as follows: The non-risk-adjusted contract (7) effectively forces itself to make a fixed payment of $[h^{int}(PAT_{ij}) + h^{ds}(PAT_{ij})]\pi^{ds,0}$ to the provider regardless of how it performs, and is therefore inefficient. The risk-adjusted contract converts a part of this deadweight ($\pi_i^R \geq 0$) into performance-contingent payment via an increase in the piece rate $\pi^{ds,1}$. This induces higher provider effort; thus, patients are hospitalized less (lower hospitalization expenses for Medicare) and receive more dialysis treatments from the provider (more revenue for provider to cover the reduction in payments due to risk-adjustment subtractions of $\pi_i^R$). Medicare's increase in reward is due to the savings from reduced hospitalizations being greater than the increase in Medicare's payment to the provider. Thus, risk adjustment is a Pareto improvement over no risk

adjustment for incumbent patients, Medicare, and the provider. In terms of magnitude, the numerical results from §5 suggest that an incumbent patient's hospital-free life can be lengthened by two weeks per year without increasing Medicare expenditures or reducing provider profits.
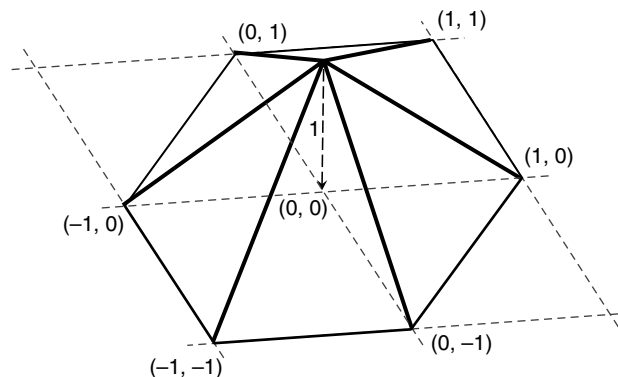
## 4. Estimation
The primitives we need to estimate are the functional forms $f$ and $h$ defined in the production functions (13) and (14) and the cost parameters $c$ in (16). Because the key inputs $(e_{ij}^{int}, e_{ij}^{ds})$ to the production functions are not observable to the econometrician, we use the provider's response (17) to Medicare's payment to infer them. Another primitive is Medicare's reward $\nu$ for each year of hospital-free life (downstream outcome). We will obtain the implied value for $\nu$ under the assumption that the current piece rates are optimal among the class of policies that reward raw downstream outcomes.

### 4.1. Specifications
We model the functional form of $f(\cdot, \cdot)$ using monoscaled finite elements, which provide a flexible mesh surface for modeling functions with domain $[0, 1]^2$. The mesh surface is constructed from linear combinations of nine centred-and-scaled copies of the Courant element $B(x, y)$, depicted in Figure 1, which is a type of box spline (Example 11 in de Boor et al. 1993). The arguments of all nine copies are scaled by a factor of 2, and each copy is centred at one of the coordinates $(0, 0)$, $(0, \frac{1}{2})$, $(0, 1)$, $(\frac{1}{2}, 0)$, $(\frac{1}{2}, \frac{1}{2})$, $(\frac{1}{2}, 1)$, $(1, 0)$, $(1, \frac{1}{2})$, and $(1, 1)$. Thus, the copies are $B(2(x - k/2), 2(y - l/2))$ for $k, l = 0, 1, 2$. The boundary conditions $f(0, \cdot) = f(\cdot, 0) = 0$ imply that the weights on the copies centred at $(0, 0)$, $(0, \frac{1}{2})$, $(0, 1)$, $(\frac{1}{2}, 0)$, and $(1, 0)$ are zero. Because $f(\cdot, \cdot)$ is also nondecreasing in the direction of the positive cone, additional constraints on the weights on the four remaining copies centred

**Figure 1　Graphical Depiction of the Courant Element** $B(x, y)$



*Note.* It is a piecewise-linear function with height 1 at the origin and 0 outside of the solid hexagonal base.

at $(\frac{1}{2}, \frac{1}{2})$, $(\frac{1}{2}, 1)$, $(1, \frac{1}{2})$, and $(1, 1)$ are needed during estimation. Some of these are binding in the estimation of the weights from data, which leads to the following parameterization of $f(\cdot, \cdot)$ by the scalar $\eta$:

$$f(DOSE, ANEMIA) = \eta F(DOSE, ANEMIA), \quad (19)$$

where $F(x, y) = B(2(x - \frac{1}{2}), 2(y - \frac{1}{2})) + B(2(x - \frac{1}{2}), 2(y - 1)) + 2B(2(x - 1), 2(y - \frac{1}{2})) + 2B(2(x - 1), 2(y - 1))$. To simplify exposition, we take the functional form (19) as given so the description of the estimation procedure will not involve constraints.

The functions $h^{int}(\cdot)$ and $h^{ds}(\cdot)$ are modeled as additively separable in each patient characteristic of the vector $PAT_{ij}$ listed in Table 2. Any redundancy (linear dependence) in the variables is removed by computing its principal components (Hastie et al. 2009) and retaining the nonsingular components, which we denote by the vector $CPAT_{ij}$. Thus, $h$ is linear in $CPAT_{ij}$ with parameters $\theta$:

$$
\begin{aligned}
h^{int}(PAT_{ij}) &= \theta^{int'} CPAT_{ij}, \\
h^{ds}(PAT_{ij}) &= \theta^{ds'} CPAT_{ij}.
\end{aligned}
\quad (20)
$$

The payment variation in the current system provides only four degrees of freedom for identifying the five cost parameters $c^{int}$, $c^{ds}$, $c^{int, int}$, $c^{int, ds}$, and $c^{ds, ds}$. To resolve this underidentification issue, we assume that $c^{int, ds} = 0$, which rules out cost interactions between the two types of efforts. This choice is justified by the fact that the symmetric matrix defined by $c^{int, int}$, $c^{int, ds}$, and $c^{ds, ds}$ is positive definite (otherwise the maximization of $\Pi_j$ will have no unique interior solution), so $c^{int, int} c^{ds, ds} > (c^{int, ds})^2$. Therefore, the diagonal matrix parameterization allows us to focus on identifying the dominant terms $c^{int, int}$ and $c^{ds, ds}$. This assumption implies that there is little overlap between the efforts $e^{int}(z_{ij}^{int})$ and $e^{ds}(z_{ij}^{int}, z_{ij}^{ds})$, or equivalently, that $e^{ds}(z_{ij}^{int}, z_{ij}^{ds})$ is driven primarily by $z_{ij}^{ds}$.

### 4.2. Estimating Equations
The estimation is based on treating the production functions (15) as a system of estimating equations. Because we observe neither $e_{ij}^{int}$ nor $e_{ij}^{ds}$, we substitute the provider's response (17) in terms of $c$ and $\pi$ into (15). Bearing in mind that the data were generated under the contract (4), we set $\pi_j^{int} = 0$ and $\pi_j^{ds} = BASEPAY_j$. Finally, we substitute the specifications (19) and (20) for $f(\cdot, \cdot)$, $h(\cdot)$, and $c^{int, ds} = 0$ into (15) as well, using the abbreviation $F_{ij} \equiv F(DOSE_{ij}, ANEMIA_{ij})$ to emphasize that $F(DOSE_{ij}, ANEMIA_{ij})$ is a data variable in its own right. Then we divide both sides of the equation for

the intermediate score by $\eta$ to obtain the system on which our estimation is based:

$$
\begin{aligned}
DSOUT_{ij} = {}&-c^{ds}/c^{ds, ds} + (c^{ds, ds})^{-1} BASEPAY_j + \eta F_{ij} \\
&+ \theta^{ds'} CPAT_{ij} + \varepsilon_{ij}^{ds} \\
F_{ij} = {}&-c^{int}/(\eta c^{int, int}) + (\eta c^{int, int})^{-1} BASEPAY_j \\
&+ (\theta^{int}/\eta)' CPAT_{ij} + \varepsilon_{ij}^{int}/\eta.
\end{aligned}
\quad (21)
$$

The downstream outcome and intermediate score are then functions of the observed variables $BASEPAY_j$, $F_{ij}$, and $CPAT_{ij}$, and of the structural parameters $c^{int}$, $c^{ds}$, $c^{int, int}$, $c^{ds, ds}$, $\eta$, $\theta^{int}$, and $\theta^{ds}$ that need to be estimated.

### 4.3. Estimation Procedure
We will assume for simplicity that $\varepsilon_{ij}^{int}$ and $\varepsilon_{ij}^{ds}$ are uncorrelated. A treatment of the correlated case yielded similar results and is omitted for brevity.[4] In the absence of correlation each of the equations in (21) may be estimated by OLS.

### 4.4. Identification
Several sources of variations and assumptions allow us to identify the model parameters.

First, $h^{int}(PAT_{ij}) = \theta^{int'} CPAT_{ij}$ and $h^{ds}(PAT_{ij}) = \theta^{ds'} CPAT_{ij}$ are both identified up to a constant from the estimation. The constants are determined by the nonnegativity conditions $h^{int}(PAT_{ij}) \geq 0$ and $h^{ds}(PAT_{ij}) \geq 0$ and are equal to the negative of the minimum values of $h^{int}(\cdot)$ and $h^{ds}(\cdot)$ over all physiologically feasible values of $PAT_{ij}$. In turn, the regression intercepts $-c^{int}/(\eta c^{int, int})$ and $-c^{ds}/c^{ds, ds}$ need to be offset by the same amount. We approximate the global minimums with the minimums over patients in this study.

Second, the cost parameters are identified from (i) the assumption that our chain of providers share a common cost structure (Assumption 8 in §3.4); (ii) the specification assumption $c^{int, ds} = 0$; and (iii) the geographic variation in the outdated composite rate $BASEPAY_j$ (5) received by each provider. Higher $BASEPAY_j$ rates translate to better $DOSE$, $ANEMIA$, and downstream outcomes, which traces out the reduced-form coefficients $1/(\eta c^{int, int})$ and $1/c^{ds, ds}$ common to all providers.

Third, we assume that there are no other relevant patient attributes beyond the expansive list $PAT_{ij}$. This rules out the possibility that patients have unobserved heterogeneities that correlate with their choice of provider. Some empirical support for this assumption is provided in Lee et al. (2010), which estimated the

---

[4] Treatment of the correlated case is available from the authors upon request.

2003 risk-adjusted hospitalization outcomes of dialysis providers by regressing patient hospital days on patient's choice of provider as well as $PAT_{ij}$. To deal with possible unobserved patient heterogeneities that correlate with provider choice, travel distances were used to instrument for patient provider choice. The IV estimates were then compared to the naive OLS estimates. The Hausman test did not reject the null hypothesis that the OLS estimates are consistent (*p*-value 0.10), suggesting that any bias arising from unobserved patient heterogeneities is weak.

Fourth, the residual in the second equation of (21), $\varepsilon_{ij}^{int}/\eta$, is excludable from the first equation. By the assumption that $\varepsilon_{ij}^{int}$ and $\varepsilon_{ij}^{ds}$ are uncorrelated, $\varepsilon_{ij}^{int}$ provides a source of exogenous variation in $F_{ij} \equiv F(DOSE_{ij}, ANEMIA_{ij})$ for identifying $\eta$ in the $DSOUT_{ij}$ equation.

## 5. Results

To ensure that we provide counterfactual results based on the most realistic modeling representation, the estimates presented herein are based on the full model that explicitly captures the EPO drug component.[5] Table 3 presents parameter estimates for the providers' cost structure and the intermediate score parameterization $\eta$. The estimates for $\theta^{int}$ and $\theta^{ds}$ are omitted, because each one is of very high dimension and the corresponding principal component patient attributes $CPAT_{ij}$ have little interpretability to offer much insight. Instead, knowledge of the provider average values $(1/N_j) \sum_i h^{int}(PAT_{ij})$ and $(1/N_j) \sum_i h^{ds}(PAT_{ij})$ alone suffice to risk-adjust aggregate patient outcomes for use in the optimal payment system. These values are given in Table 4. The estimated surface of $f(DOSE, ANEMIA)$ is plotted against $DOSE$ and $ANEMIA$ in Figure 2. We observe the following:

1. The linear cost component $c^{int}$ of the intermediate effort $e^{int}$ is twice as large as its downstream counterpart $c^{ds}$, whereas the cost curvatures $c^{int, int}$ and $c^{ds, ds}$ are about the same. This implies that motivating providers to exert intermediate effort is more expensive than motivating residual downstream effort.

2. Note that from Figure 2 there are diminishing returns to improving anemia control in patients. Specifically, increases in $ANEMIA$ beyond 50% will not further improve patient downstream outcomes and, hence, will not be rewarded by the optimal payment system.

---

[5] The full model operates under the additional assumption that the provider's response (17) also depends on the average EPO profit across the provider's patients, but not on the individual EPO earnings from each patient. Thus, treatment efforts are still fixed across patients within each provider. Support for this assumption comes from the data itself as well as discussions with nephrologists.

**Table 3    Parameters Estimates for Provider Cost and Intermediate Outcome Function**

| Parameter | Estimate (95% confidence interval) |
|---|---|
| $c^{int}$ | 20,300  (8,630, 23,800) |
| $c^{ds}$ | 10,400  (4,990, 14,100) |
| $c^{int, int}$ | 100,000  (74,000, 207,000) |
| $c^{ds, ds}$ | 92,900  (65,900, 134,000) |
| $\eta$ | 0.117  (0.0883, 0.146) |

**Table 4    Provider Average Values** $(1/N_j) \sum_i h^{int}(PAT_{ij})$ **and** $(1/N_j) \sum_i h^{ds}(PAT_{ij})$

| Provider | $(1/N_j) \sum_i h^{int}(PAT_{ij})$ | $(1/N_j) \sum_i h^{ds}(PAT_{ij})$ |
|---|---|---|
| 1 | 0.0896 years | 0.410 years |
| 2 | 0.0897 years | 0.481 years |
| 3 | 0.0963 years | 0.461 years |
| 4 | 0.0919 years | 0.489 years |
| 5 | 0.0997 years | 0.514 years |
| 6 | 0.0976 years | 0.437 years |
| 7 | 0.0894 years | 0.393 years |
| 8 | 0.1020 years | 0.516 years |
| 9 | 0.0892 years | 0.399 years |
| 10 | 0.0791 years | 0.427 years |
| 11 | 0.1040 years | 0.487 years |
| 12 | 0.0845 years | 0.397 years |

3. Note that $\eta > 0$ implies that $f(\cdot, \cdot) = \eta F(\cdot, \cdot)$ is not identically zero, so $DOSE$ and $ANEMIA$ do impact downstream outcomes. This finding is in agreement with Wolfe et al. (2005), which showed that dialysis dosage and anemia control were negatively associated with mortality rates. Similarly, the estimate for $c^{ds, ds}$ suggests that $e^{ds}$ varied across providers according to payments (see (17)). This points to the existence of treatment tasks $z_{ij}^{ds}$ that are under provider control. It follows from the discussion in §3 that rewarding for $DOSE$ and $ANEMIA$ alone is insufficient for the pay-for-compliance system to achieve first-best. The extent of the underperformance depends on the importance of the tasks $z_{ij}^{ds}$ that cannot be induced by

**Figure 2    Estimated Surface of Intermediate Performance Score** $f(DOSE, ANEMIA)$
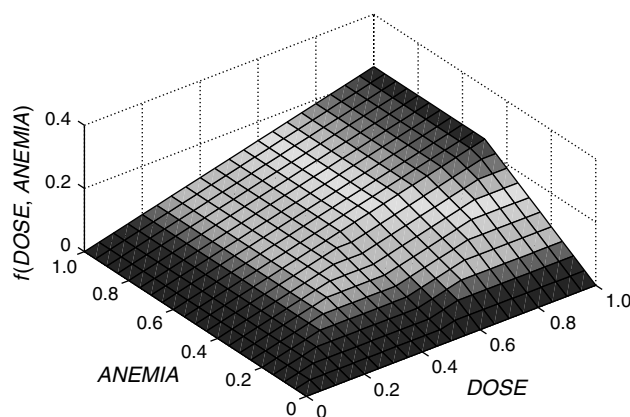
**Table 5**    Description and Summary of the Performances of Alternative Payment Systems

| $\nu$ implied by current system: $72,400 per year | (i) Reward raw downstream outcomes (2003 practice) (7) | (ii) Reward fully risk-adjusted downstream outcomes (8) | (iii) Reward fully risk-adjusted intermediate scores (9) |
|---|---|---|---|
| Average $\pi_{ij}^R$ | $0 | $700 | −$2,200 |
| Piece rate $\pi^{int}$ when $WAGEIDX = 1$ | $0 | $0 | $21,300 |
| Piece rate $\pi^{ds}$ when $WAGEIDX = 1$ | $19,700 | $23,000 | $0 |
| Chain's net earnings per patient | Baseline | +$0 | +$0 |
| Expected hospital-free life per patient | Baseline | +0.0439 yr. | −0.0199 yr. |
| Medicare savings per patient | Baseline | +$1,430 | −$732 |

rewarding for *DOSE* and *ANE6 MIA*. The simulations below quantify the efficiency loss.

4. The maximum value of $f(DOSE, ANEMIA)$ from Figure 2 is less than 0.25 years, so most of the patient downstream outcome cannot be explained by the two intermediate measures alone. The existence of a large residual component provides additional support for the existence of $z_{ij}^{ds}$.

### 5.1. Optimal Payment Systems for the ESRD Program

The remaining parameter required for simulating the performance of the proposed payment system (6) is $\nu$, Medicare's reward for each year of hospital-free life that its beneficiaries gain. Assuming that Medicare's current reimbursement rate is optimal among all policies that reward raw downstream outcomes, we arrive at an implied value of $\nu = \$72,400/\text{year}$. We use this value and the estimated parameters to compute optimal payments for (6) and to simulate performances for three special cases: (i) the 2003 contract that pays for raw downstream outcomes (7); (ii) the contract (8) that rewards providers for risk-adjusted downstream outcomes ($\pi^{int} = 0$); and (iii) the pay-for-compliance contract (9) that pays for risk-adjusted performance on *DOSE* and *ANEMIA* ($\pi^{ds} = 0$). The unit of analysis is the chain, and each provider is allowed to choose its own efforts according to the wage-adjusted rates it receives. For comparison to the 2003 contract, we will retain the wage adjustment indices $WAGEIDX_j$ (5) used in 2003. In addition, when setting its efforts, providers must obey a constraint not introduced in §3: the provider's new efforts cannot expand or contract by more than 20% of its current efforts. This constraint reflects the reality that provider performance cannot be arbitrarily extended or reduced, and also restricts efforts to the range for which we have adequate data to estimate the provider's cost function. The performance results for the three cases are presented in

Table 5. The first row entry is the average patient statistical risk adjustment $\sum_{ij} \pi_{ij}^R / N$. The second and third rows represent the optimal pre-wage-adjusted piece rates for the risk-adjusted intermediate score and downstream outcome, respectively. The last three rows represent the gains by the chain, patients, and Medicare, respectively relative to the contract used in 2003. A few remarks are in place:

1. As expected under provider risk neutrality, contract (ii) performed the best out of all three cases: each patient can expect to gain two weeks (0.0439 years) of hospital-free life per year, and at the same time Medicare stands to save $1,430 per patient. Compared to the 2003 contract (i), (ii) pays the provider $700 less up-front, and reallocates this amount to performance-contingent payment by increasing the piece rate $\pi_j^{ds}$. This induces higher provider effort; thus, patients are hospitalized less (lower hospitalization expenses for Medicare) and receive more dialysis treatments from the provider (more revenue for the provider to offset the $700 deduction). This was anticipated by Proposition 1.

2. Contract (iii) performs even worse than the 2003 contract—a patient's hospital-free life may be reduced by a week (0.0199 years) per year under (iii). The performance of (iii) would have been even worse had we not limited effort contraction to 20% of the 2003 levels. We discussed earlier in this section that this under-performance is due to the inability of the pay-for-compliance contract (iii) to induce provider effort in treatment tasks $z_{ij}^{ds}$. Until more intermediate measures are identified, the only way to induce effort in $z_{ij}^{ds}$ is to reward for downstream outcomes, as in contract (ii).

## 6. Concluding Remarks

Guided by the incentive dynamics between Medicare and the dialysis providers, we developed an empirical principal–agent model to estimate the providers' cost of effort and other parameters from patient-level and provider-level data. The estimates were used to address the following questions: Is the pay-for-compliance system envisioned by Medicare better than the current practice of paying for downstream outcomes? How can the two QIP process-compliance measures be integrated into a single intermediate outcome performance score that can be used to reward providers, as mandated by the Social Security Act? How should patient outcomes be risk adjusted, and what welfare gains can be achieved by doing so?

Our results suggest that although rewarding providers for dialysis dosage adequacy and anemia control can induce provider efforts in treatment tasks that enhance patient hospital-free life, it is still better to pay for downstream outcomes: There exist important treatment tasks that can only be induced by either

rewarding for the downstream outcome or for additional intermediate measures. For example, vascular access infection rates has been identified as a candidate intermediate measure for the QIP, but currently this data is not collected by the U.S. Renal Data System. If Medicare is insistent on paying for process compliance, then measures like this need to be empirically validated and included as performance metrics as well.

Relative to paying for raw downstream outcomes, our results also suggest that paying for fully risk-adjusted downstream outcomes can strengthen provider incentive to deliver quality care. For our study cohort the predicted gain in hospital-free life of dialysis patients is about two weeks (0.0439 years) per patient per year, whereas Medicare's predicted savings is over \$1,400 per patient per year. Extrapolating our finding to all Medicare beneficiaries receiving dialysis treatment from freestanding providers in the United States (265,958 patients), this would translate into a gain of 11,700 years in hospital-free life, along with Medicare savings of \$380 M. This is a significant potential improvement in patient outcomes without a net increase in Medicare expenditures. Note, however, that to implement this system nationwide would require separately estimating cost parameters for each group of homogeneous providers, possibly by location and chain.

Like any empirical analysis of a stylized model, our analysis and results have several important limitations. We will discuss these limitations and describe future research to address them.

First, to identify cost, we restricted our study to a group of providers that shared a homogeneous cost structure but received varying payment rates. Our study cohort consisted of similar-sized providers of the same chain and within close proximity to each other to ensure homogeneity in cost structure (Assumption 8). We relied upon Medicare's outdated labor wage adjustment to provide variation in payment rates across providers. For a general set of dialysis providers, we expect to find cost variations based on attributes such as type of ownership (for-profit, nonprofit, chain-affiliated versus independent). Thus, to implement this system across the country would require manually clustering dialysis providers into homogeneous groups and performing a separate estimation for each group. A more scalable approach, left for future research, is to investigate flexible cost function parameterizations that can vary with provider attributes, allowing for mixtures of heterogeneous agents.

Second, in this paper we focused on the benefit that risk adjustment brings to incumbent patients, but we did not explicitly model the selective admittance of prospective patients by providers. To empirically identify this selection mechanism requires data on the mix of prospective patients seeking dialysis therapy, as well as some variation in the part of the payment system pertaining to patient case-mix. Because risk adjustment was first applied to payments in April 2005 (*Federal Register* 2004), one line of future research would be to use data from that year to estimate the effect of risk adjustment on patient dialysis access.

Third, our analysis assumes that the providers' objective was to maximize profit alone. In fact, our framework also accommodates the relaxation where the provider's objective is additively separable in profit and other motivations. In this case, the latter is absorbed into the estimates for the cost function. To better understand the factors that motivate agents, another line of future work can be directed toward seeking variations in the data that would allow the nonmonetary component of provider objective to be separately identified.

Finally, the model used in this study is static, which precludes the possibility that providers may learn and game the reimbursement system over time. A dynamic model extension of our work here is particularly interesting because the data-driven reimbursement system depends on provider response in previous time periods.

Beyond healthcare, the problem of contracting on process compliance is important in situations where it is difficult to contract on the downstream outcome— the downstream outcome may either be costly to measure or can only be observed in the distant future. As long as a one-time effort can be made to collect sample observations of both intermediate and downstream outcomes, the method developed in this paper can be applied to these situations as well. With the abundance of data made available by modern information technology systems, we believe that the problem of developing evidenced-based incentive systems can be a very fruitful area of research for operations management scholars.

## Acknowledgments

## References

Bradley, E. H., J. Herrin, B. Elbel, R. L. McNamara, D. J. Magid, B. K. Nallamothu, Y. F. Wang, S. L. T. Normand, J. A. Spertus, H. M. Krumholz. 2006. Hospital quality for acute myocardial infarction: Correlation among process measures and relationship with short-term mortality. *J. Amer. Medical Assoc.* **296**(1) 72–78.

de Boor, C., K. Hollig, S. Riemenschneider. 1993. *Box Splines*. Springer, New York.

End-Stage Renal Disease Quality Incentive Program (QIP). 2010. Accessed December 15, 2010, https://www.cms.gov/ESRDQualityImproveInit/.

*Federal Register*. 1986. HCFA medicare program; ESRD program; composite rates and methodology for determining the rates; final notice. *Federal Register* **51** 29403–29419.

*Federal Register*. 2004. Payment for renal dialysis services final rule. *Federal Register* **69** 66319.

*Federal Register*. 2005. Medicare program; revisions to payment policies under the physician fee schedule for calendar year 2006; proposed rule. *Federal Register* **70** 45793.

*Federal Register*. 2011. End-stage renal disease quality incentive program; final rule. *Federal Register* **76** 628–646.

Fuloria, P. C., S. A. Zenios. 2001. Outcomes-adjusted reimbursement in a health-care delivery system. *Management Sci.* **47**(6) 735–751.

Gibbons, R. 2005. Incentives between firms (and within). *Management Sci.* **51**(1) 2–17.

Hastie, T., R. Tibshirani, J. Friedman. 2009. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer, New York.

Jha, A. K. 2006. Measuring hospital quality: What physicians do? How patients fare? Or both? *J. Amer. Medical Assoc.* **296**(1) 95–97.

Lee, D. K. K., G. M. Chertow, S. A. Zenios. 2010. Re-exploring differences among for-profit and non-profit dialysis providers. *Health Services Res.* **45**(3) 633–646.

Medicare Improvements for Patients and Providers Act (MIPPA). 2008. Accessed December 15, 2010, http://www.govtrack.us/congress/bill.xpd?bill=h110–6331.

Medicare Prescription Drug, Improvement, and Modernization Act (MMA). 2003. Accessed December 15, 2010, http://www.cms.hhs.gov/medicarereform/MMAactFullText.pdf.

Thamer, M., Y. Zhang, J. Kaufman, D. Cotter, F. Dong, M. A. Hernan. 2007. Dialysis facility ownership and epoetin dosing in patients receiving hemodialysis. *J. Amer. Medical Assoc.* **297**(15) 1667–1674.

U.S. Government Accounting Office (GAO). 2004. Medicare dialysis facilities: Beneficiary access stable and problems in payment system being addressed. Report to Congressional Committees (GAO-04-450), GAO, Washington, DC.

U.S. Renal Data System (USRDS). 2005. http://www.usrds.org/.

Wolfe, R. A., T. E. Hulbert-Shearson, V. B. Ashby, S. Mahadevan, F. K. Port. 2005. Improvements in dialysis patient mortality are associated with improvements in urea reduction ratio and hematocrit, 1999 to 2002. *Amer. J. Kidney Diseases* **45**(1) 127–135.